

AN INDUSTRIAL DESIGN ENGINEER is a FUTURE-TECH DESIGN ANALYST

In today's fast-moving digital landscape, technology leaders face a critical challenge: How to adopt emerging technologies with precision and purpose, while safeguarding trust, resilience, and long-term value. A focus on integrating industry-leading expertise with emerging technologies, delivering innovative, tech-enabled solutions that drive operational efficiency, enhance customer experiences, and support long-term business success. This includes harnessing data-driven strategies, advanced technology platforms, ecosystem partnerships, and scalable cloud solutions.

Design Trends 2025

In an age of rapid change, design not only helps us understand these shifts but also empowers us to actively shape them. This extends beyond simply integrating new innovations and technologies; alongside values and interdisciplinary collaboration, it encompasses the ability to envision and critically analyze future scenarios. But what are the most disruptive shifts reshaping the global design landscape in the near future?

Design Trends aids designers in their understanding of the transformative trends, connecting cross-disciplinary design trends with key technological and social trends. This interdisciplinary approach enables designers and industry experts to quickly transfer knowledge, offering a diverse and comprehensive perspective on the global design landscape. This year's Design Trends focuses on six major transformations in our society:

CONSCIOUS ECONOMY

A significant shift in corporate strategy is underway, placing purpose, service, and care at its core to generate societal impact. This evolution leverages new technologies and innovative service design to foster long-term loyalty among both customers and employees. Design is central to this shift, as it helps unify functional excellence with emotional connections.

HUMAN DIGITALITY

The more complex technologies become, the more the requirements for the interfaces between products and services also increase. While the aim for a long time was to make a product intuitive to use, this interaction has now become much more complex. A new culture of interaction and the AI boom require new design solutions, while playfulness offers a serious approach.

CO-SOCIETY

A multidimensional understanding of prosperity has become increasingly evident, taking immaterial components into consideration, such as individual quality of life, social participation, and ecological sustainability in equal parts. Products and services are imbued with experiences, feelings, and values beyond their practical use.

MINDSHIFT REVOLUTION

The topic of mental health has reached the center of society, as mental stress and loneliness are on the rise. A preventative approach for mental health is becoming the norm. Design drives inclusivity by focusing on function, accessibility, and emotional resonance to create diverse products and spaces.

ECO TRANSITION

Eco Transition describes the global transformation to a holistic, circular, and ecosystemic society. Thanks to green high-tech, the energy transition is in full swing. Design approaches such as Sustainable Sharing show how materials and artifacts can be used, regenerated, and reused in smart ways. By integrating biodiversity, design fosters environments that support both people and nature.

GLOBAL LOCALIZATION

Global Localization is creating a new balance between global connectedness and local stability that will make us more resilient on an economic, ecological, and social level. In a globalized, yet local world, cross-cultural learning fosters cultural exchange at an equal level and an understanding of other cultures, perspectives, and practices.

Use these 2025 strategic technology trends to shape the future with responsible innovation

In selecting this year's 10 top strategic technology trends, Gartner analysts organized them across three themes: AI imperatives and risks, new frontiers of computing, and human-machine synergy.

Theme No. 1: AI imperatives and risks drive organizations to protect themselves

Trend 1: [Agentic AI](#): Autonomous AI can plan and take action to achieve goals set by the user.

- **Business benefits:** A virtual workforce of agents to assist, offload, and augment the work of humans or traditional applications.
- **Challenges:** Requires robust guardrails to ensure alignment with providers' and users' intentions.

Trend 2: [AI governance platforms](#): Technology solutions enable organizations to manage the legal, ethical, and operational performance of their AI systems.

- **Business benefits:** Create, manage, and enforce policies that ensure responsible use of AI, explain how AI systems work, model lifecycle management, and provide transparency to build trust and accountability.
- **Challenges:** AI guidelines vary across regions and industries, making it difficult to establish consistent practices.

Trend 3: Disinformation security: An emerging technology category aimed at systematically discerning trust.

- **Business benefits:** Decreases fraud by strengthening controls for validating identity; prevents account takeover through continuous risk scoring, contextual awareness, and a continuous adaptive trust model; and protects brand reputation by identifying harmful narratives.
- **Challenges:** Requires a continuously updated, multilayered, adaptive learning, and a team approach.

Theme No. 2: New frontiers of computing prompt organizations to reconsider how they compute

Trend 4: Post-quantum cryptography (PQC): Data protection that is resistant to quantum computing (QC) decryption risks.

- **Business benefits:** Protects data from the security risks that will come with the advent of **quantum computing**.
- **Challenges:** PQC algorithms are not drop-in replacements for existing asymmetric algorithms. Current applications may have performance issues, will require testing, and may need to be rewritten.

Trend 5: Ambient invisible intelligence: Technology unobtrusively integrated into the environment to enable a more natural, intuitive experience

- **Business benefits:** Enables low-cost, real-time tracking and sensing of items, improving visibility and efficiency; potential for unforgeable provenance and new ways for objects to report identity, history and properties.

- **Challenges:** Providers will have to address privacy concerns and obtain consent for some types of data use. Users may opt to disable tags to preserve privacy.

Trend 6: Energy-efficient computing: An approach to increasing **sustainability** through more efficient architecture, code, and algorithms; hardware optimized for efficiency; and the use of renewable energy to run systems.

- **Business benefits:** Address legal, commercial, and social pressures to improve sustainability by reducing carbon footprint.
- **Challenges:** New hardware, cloud services, **skills**, tools, algorithms, and applications will be required; migrating to new computing platforms will be complex and expensive; energy prices may rise in the short term as green energy demand increases.

Trend 7: Hybrid computing: Combines different compute, storage and network mechanisms to solve computational problems.

- **Business benefits:** Highly efficient, high-speed, transformative innovation environments; AI that performs beyond current technological limits; autonomous businesses powered by higher levels of automation; augmented human capability allowing real-time personalization at scale and use of the human body as a computing platform.
- **Challenges:** Nascent, highly complex technologies require **specialized skills**; a system of autonomous modules introduces security risks; involves experimental technologies and high costs; and requires orchestration and integration.

Theme No. 3: Human-machine synergy brings together the physical and digital worlds

Trend 8: Spatial computing: Digitally enhances the physical world using technologies like augmented and virtual reality to offer immersive experiences.

- **Business benefits:** Addresses consumer demand for immersive and interactive experiences in gaming, education and e-commerce; satisfies demand for sophisticated visualization tools for decision making and efficiency in healthcare, retail and manufacturing.
- **Challenges:** Head-mounted displays are expensive and unwieldy, require frequent charging, isolate users and may increase the potential for accidents; user interfaces are complex; data privacy and security are major concerns.

Trend 9: Polyfunctional robots: Robots capable of performing multiple tasks and seamlessly switching between them as required.

- **Business benefits:** Improved efficiency; faster ROI; no need for architectural changes or bolt-down infrastructure means fast deployment, low risk and scalability; can substitute for or work with humans.
- **Challenges:** The industry hasn't yet standardized on price or minimum functionality required.

Trend 10: Neurological enhancement: Improving cognitive abilities with technologies that read and decode brain activity.

- **Business benefits:** Human **upskilling**, safety improvements, personalized education, allowing older people to work longer, next-generation marketing..

- **Challenges:** Initially expensive, limited battery and options for mobility and wireless connectivity; invasive and risky; UBMs and BBMs interface directly with the human brain, creating security challenges; ethical concerns (e.g., altering users' perception of reality).

2025 Top Technology Trends

What are the top technology trends for 2025?

The Top Technology Trends for 2025 are:

1. Agentic AI
2. Post-quantum Cryptography
3. Spatial Computing
4. AI Governance Platforms
5. Ambient Invisible Intelligence
6. Polyfunctional Robots
7. Disinformation Security
8. Energy-Efficient Computing
9. Neurological Enhancement
10. Hybrid Computing

The Future of Design: Tech Forecasting Essentials

Staying Ahead of the Curve with Technology Forecasting in Design and Technology, June 15, 2025

Understanding the Basics of Technology Forecasting

The rapidly evolving landscape of technology and design necessitates a forward-thinking approach to stay ahead of the curve. Technology forecasting, a systematic process of predicting future technological advancements and their potential impacts, has become an indispensable tool for designers, technologists, and business leaders alike. In this article, we'll delve into the essentials of technology forecasting, its significance in shaping the future of design and technology, and explore how to apply it in your work.

Definition and Principles of Technology Forecasting

Technology forecasting is a methodology used to predict the future characteristics of useful technological machines, procedures, or techniques. It involves analyzing current trends, available data, and potential future developments to forecast the trajectory of technological advancements. The core principles of technology forecasting include:

- Systematic analysis of historical data and current trends
- Identification of potential future developments and their implications
- Continuous monitoring and updating of forecasts

Types of Technology Forecasting: Exploratory and Normative

There are two primary types of technology forecasting: exploratory and normative.

- **Exploratory forecasting** involves analyzing current trends and available data to predict potential future developments. This approach is useful for identifying emerging technologies and understanding their potential impact.
- **Normative forecasting**, on the other hand, starts with a desired future outcome and works backward to identify the necessary steps to achieve it. This approach is useful for setting strategic goals and developing roadmaps to achieve them.

Importance of Technology Forecasting in Design and Technology

Technology forecasting plays a vital role in design and technology by enabling organizations to:

- Anticipate and prepare for future technological advancements
- Identify potential opportunities and challenges
- Develop strategic plans to stay ahead of the competition
- Make informed decisions about investments and resource allocation

The importance of technology forecasting is highlighted by a study published in the Journal of Forecasting, which found that organizations that use technology forecasting are more likely to achieve their strategic goals [¹](<https://onlinelibrary.wiley.com/journal/1099131x>).

Advanced Techniques for Technology Forecasting

As technology continues to evolve at an unprecedented rate, advanced techniques for technology forecasting have become increasingly important. Some of these techniques include:

Overview of Advanced Forecasting Techniques: Scenario Planning and Delphi Method

- **Scenario planning** involves developing multiple plausible scenarios of potential future developments. This approach helps organizations prepare for different possible outcomes and develop strategies to address them.
- The **Delphi method** is a structured communication technique that involves gathering expert opinions to forecast future developments. This approach is useful for gathering insights from a diverse group of experts and achieving a consensus on potential future outcomes.

Introduction to AI and Machine Learning in Technology Forecasting

Artificial intelligence (AI) and machine learning (ML) are increasingly being used in technology forecasting to analyze large datasets, identify patterns, and make predictions. These technologies have the potential to significantly improve the accuracy and efficiency of technology forecasting.

Case Studies of Successful Technology Forecasting Using Advanced Techniques

Several organizations have successfully used advanced techniques for technology forecasting to drive innovation and stay ahead of the competition. For example:

- A study published in the Journal of Technology Forecasting and Social Change found that using scenario planning helped a leading technology company anticipate and prepare for future developments in the field of artificial intelligence ^[^3](<https://www.journals.elsevier.com/technological-forecasting-and-social-change/>).
- Another study published in the Journal of Forecasting found that using the Delphi method helped a group of experts predict the future development of renewable energy technologies ^[^4](<https://onlinelibrary.wiley.com/journal/1099131x>).

Implementing Technology Forecasting in Your Organization

Implementing technology forecasting in your organization requires a systematic approach. Here are some steps to follow:

Steps to Implement Technology Forecasting in Your Organization

1. **Define the scope and goals** of your technology forecasting effort
2. **Gather relevant data** and information
3. **Select a suitable forecasting technique** (e.g., exploratory, normative, scenario planning, Delphi method)
4. **Analyze the data and develop forecasts**
5. **Communicate the results** to stakeholders and decision-makers
6. **Continuously monitor and update** your forecasts

Best Practices for Integrating Technology Forecasting into Your Workflow

To get the most out of technology forecasting, it's essential to integrate it into your workflow. Here are some best practices to follow:

- **Make technology forecasting a regular activity**, rather than a one-time exercise
- **Involve a diverse group of stakeholders** in the forecasting process
- **Use a combination of forecasting techniques** to achieve a more comprehensive understanding of potential future developments
- **Continuously monitor and update** your forecasts to reflect changing circumstances

Overcoming Common Challenges and Obstacles in Technology Forecasting

Despite its many benefits, technology forecasting can be challenging. Here are some common obstacles and how to overcome them:

- **Limited data and information:** Use alternative data sources, such as expert opinions or proxy data, to supplement limited data.
- **Uncertainty and ambiguity:** Use techniques like scenario planning to develop multiple plausible scenarios and prepare for different possible outcomes.
- **Resistance to change:** Communicate the benefits and value of technology forecasting to stakeholders and decision-makers.

Technology forecasting is a powerful tool for anticipating and preparing for future technological advancements. By understanding the basics of technology forecasting, using advanced techniques, and implementing them in your organization, you can stay ahead of the curve and drive innovation.

FAQ

What is technology forecasting?

Technology forecasting is a systematic process of predicting future technological advancements and their potential impacts.

Why is technology forecasting important?

Technology forecasting is important because it enables organizations to anticipate and prepare for future technological advancements, identify potential opportunities and challenges, and develop strategic plans to stay ahead of the competition.

What are the different types of technology forecasting?

The two primary types of technology forecasting are exploratory and normative. Exploratory forecasting involves analyzing current trends and available data to predict potential future developments, while normative

forecasting starts with a desired future outcome and works backward to identify the necessary steps to achieve it.

How can I implement technology forecasting in my organization?

To implement technology forecasting in your organization, follow these steps: define the scope and goals of your technology forecasting effort, gather relevant data and information, select a suitable forecasting technique, analyze the data and develop forecasts, communicate the results to stakeholders and decision-makers, and continuously monitor and update your forecasts.

Executive Summary: What are the Top 10 Technology Trends to Watch in 2026 & Beyond?

In 2026 and beyond, emerging technology trends will present unparalleled opportunities for growth, efficiency, and transformation. This data-driven report highlights the top technology trends set to shape the future. We categorized the tech trends into 10 clusters, and each category features specific technologies:

1. [Artificial Intelligence & Automation](#)
 - [Agentic AI](#)
 - [Generative AI](#)
 - [Hyperautomation](#)
 - [Predictive Maintenance](#)
2. [Advanced Connectivity](#)
 - [Internet of Things \(IoT\)](#)
 - [5G & 6G](#)
 - [Zero-Trust Security](#)
3. [Sustainability](#)
 - [Renewables](#)
 - [Climate Tech](#)
4. [Biotech & Healthtech](#)

- [Preventive Medicine](#)
- [Gene Editing](#)
- [Synthetic Biology](#)
- [Longevity Tech](#)
- 5. [Advanced Engineering & Robotics](#)
 - [Advanced Robots](#)
 - [Autonomous Vehicles](#)
- 6. [Advanced Computing](#)
 - [Edge Computing & Distributed Cloud](#)
 - [Quantum Computing](#)
 - [Neuromorphic Computing](#)
- 7. [Decentralized & Digital Trust Technologies](#)
 - [Blockchain](#)
 - [Decentralized Finance \(DeFi\)](#)
- 8. [Material Science](#)
 - [Nanotechnology](#)
 - [Biomaterials](#)
 - [Metamaterials](#)
- 9. [Immersive Technologies & Human-Machine Interaction \(HMI\)](#)
 - [Extended Reality](#)
 - [Voice User Interfaces \(VUI\)](#)
 - [Brain-Computer Interfaces](#)
- 10. [Space Tech](#)
 - [Space Exploration](#)
 - [Space Tourism](#)

The Top 10 **Digital Transformation Technology Trends** are:

- [Additive Manufacturing](#)

- [Advanced Computing](#)
- [Artificial Intelligence & Machine Learning](#)
- [Big Data & Analytics](#)
- [Blockchain](#)
- [5G Connectivity](#)
- [Cybersecurity Evolution](#)
- [Digital Twins](#)
- [Extended Reality](#)
- [Internet of Things](#)

The Road Ahead

As we navigate through 2025 and beyond, AI has transitioned from a novel concept to a fundamental component of business strategy.

Looking ahead, several trends are poised to shape the future of AI in business:

- **Generative AI:** Enables businesses to create content, design products, and develop solutions at unprecedented speeds.
- **Agentic AI:** The rise of autonomous AI agents capable of making decisions and performing tasks without human intervention is set to accelerate workflows.
- **AI-Driven Personalization:** AI's ability to analyze vast datasets allows for hyper-personalized customer experiences that foster deeper engagement and loyalty. With thousands of emerging AI technologies and providers, navigating the right investment and partnership opportunities is challenging. Stay relevant, resilient, and ready for what is next.

Future Technology Trends: 10 Innovations to Watch in 2025 & Beyond

Adarsh R., Last updated: May 27, 2025

This guide unpacks 10 future technology trends that will redefine industries over the next decade. From AI and quantum computing to nuclear energy and spatial computing, these trends are fueled by breakthrough innovations, record investments, and enterprise-scale adoption. Explore real-world applications, market signals, and strategic insights for tech leaders to stay ahead.

The future technology trends landscape is rapidly changing, and by 2027, global spending on digital transformation is expected to reach [USD 3.9 trillion](#). As enterprises quickly adopt AI, green energy, robotics, quantum computing, and more, investor confidence is at an all-time high.

This guide will enable you to explore the most important future tech trends, as well as their market momentum, business value drivers, and the innovative startups that are leading the way in the next wave of global innovation.

Executive Snapshot

10 Future Technology Trends To Watch in 2025 and Beyond

1. [Agentic AI Systems and Cognitive Automation](#)
2. [AI-Optimized Neuromorphic Hardware](#)
3. [Quantum Hardware Advances & Applications](#)
4. [Cloud-Edge Convergence & 6G](#)
5. [Industrial Metaverse & Spatial Computing](#)
6. [Clean Energy Expansion](#)
7. [Synthetic Biology](#)
8. [Humanoid & Collaborative Robots](#)
9. [Cyber-Resilient Data-Centric Security](#)
10. [Space-Based Manufacturing & Energy](#)

Top FAQs by Leaders

1. What are the most critical future technology trends to track?

- AI-driven Transformation
- Quantum Computing

- Cybersecurity

2. How to strategically adopt these future technologies?

- Pilot new technologies with controlled experiments before scaling.
- Adopt modular, cloud-native architectures for flexibility and rapid adaptation.
- Invest systematically in talent upskilling across technologies.

3. How to integrate these future technology trends without disrupting existing systems?

- Employ APIs and modular architecture designs for smooth integration.
- Conduct phased rollouts and compatibility assessments through pilot programs.
- Establish dedicated teams for lifecycle management and support during transitions.

10 Future Technology Trends To Watch [2025 – 2030]

1. Agentic AI Systems and Cognitive Automation

Agentic and autonomous AI integrate perception, reasoning, and continuous learning to optimize workflows across various industries.

They autonomously perceive environments, make informed decisions, and refine their actions through ongoing learning loops.

2025 is widely recognized as the “year of the agent,” with [Google](#), [OpenAI](#), and [Microsoft](#) launching agentic AI-powered assistants.

Market Momentum

The global agentic AI market is projected to reach [USD 47 billion](#) by 2030 at a 44% CAGR. The autonomous AI agents market size is forecast to grow from [USD 9.93 billion](#) in 2025 to [USD 44.76 billion](#) by 2029 (45.7% CAGR).

Business Value Drivers

Cost Avoidance

Autonomous shipping solutions reduce annual fuel costs. For instance, [Orca AI](#)'s alert system saved [USD 100 000](#) fuel per vessel per year. [Orca AI](#) achieved a [54%](#) reduction in close encounter events that saved approximately [195 000](#) tons of CO₂ emissions in 2024.

Also, HR AI agents automate candidate screening and save recruiters [25 hours](#) per week.

New Revenue Streams

Financial AI agents (e.g., [Unique](#)) offer productivity gains of **2 hours** per employee per week. Startups like [Affiniti](#) are deploying AI agents as CFOs to manage financial tasks and aid in strategic financial decisions for small and medium-sized businesses (SMBs).

Risk Reduction

AI agents monitor transactions and ensure compliance with regulatory standards to mitigate the risk of financial penalties and reputational damage. Similarly, in supply chain management, AI agents detect disruptions, assess supplier risks, and provide real-time recommendations to prevent costly delays and inventory shortages.

Breakthroughs & Milestones

Funding Surges

- [Adopt AI](#) raised **USD 6M** for no-code Agentic AI enterprise platforms.
- [Orca AI](#) secured **USD 72.5M** for autonomous maritime navigation.
- [Auditoria.AI](#) garnered **USD 38M** for AI-driven financial automation.

Technological Breakthroughs

- [IBM 5-Minute Agent Builder](#): Enables rapid creation of custom AI agents in less than five minutes on multiple LLM platforms.
- [OpenAI Codex Cloud Agent](#): It specializes in coding tasks and improving software development efficiency.
- [CrewAI](#) with [Llama 3](#): Open-source orchestration for collaborative agents that democratize AI teamwork for startups.
- [BMW × Figure Humanoids](#): Deployment of humanoid robots at BMW's Spartanburg plant for precision metal panel loading tasks, validating industrial-scale applications.

Application Areas

Software Development and DevOps

AI coding assistants like [OpenAI Codex Cloud Agent](#) and orchestration frameworks like [CrewAI](#) with [Llama 3](#) enable collaborative agent teams to write, review, and deploy code autonomously.

Companies, including [Block](#), integrate these tools to democratize software creation across non-technical teams.

Manufacturing and Industrial Automation

Agentic AI agents are now operating in physical environments.

[BMW](#) has partnered with [Figure AI](#) to deploy humanoid robot agents at its Spartanburg plant. It automates the loading of complex metal panels.

Spotlighting an Innovator: Synergetics.ai

[Synergetics.ai](#) is a US-based company that offers an agent-to-agent communication protocol. It enables autonomous AI agents to securely interact, transact, and collaborate across digital ecosystems.

The platform uses advanced machine learning and automation to enable agents to interpret data, make decisions, and execute tasks with minimal human intervention.

Its *AgentWorks* suite includes tools like *AgentTalk* that secure inter-agent communication, *AgentWallet* for digital asset management, and *AgentMarket* for deploying AI agents.

Synergetics.ai simplifies the development and deployment of AI solutions with a no-code, drag-and-drop interface.

2. AI-Optimized Neuromorphic Hardware

Inspired by the human brain's neural architecture, AI-optimized neuromorphic hardware performs complex computation with [100 times](#) less energy consumption.

For example, [Intel](#)'s Hala Point features [1.15 billion](#) artificial neurons. It contains more than [10 times](#) more neuron capacity and [12 times](#) higher performance to tackle real-world challenges.

Market Momentum

The neuromorphic computing market is projected to be worth [USD 1.32 billion](#) by 2030 with a CAGR of [89.7%](#).

As per Intel's report, there's a rising demand across the consumer, automotive, and defense industries. [55%](#) of technology companies are integrating neuromorphic processors into their AI and machine learning systems.

Around [61%](#) of research institutes explore neuromorphic systems for developing energy-efficient computational models.

The demand is especially high in the autonomous vehicle industry, with [42%](#) of companies using it for real-time data processing. By application, the image processing segment is expected to [lead the market](#).

Business Value Drivers

Energy Efficiency

Neuromorphic chips consume less power than GPUs and CPUs. This makes them ideal for edge devices and battery-powered applications. For example, [IBM's TrueNorth chip](#) processes complex AI tasks such as image recognition using a fraction of the energy.

Enhanced Cognitive and Perceptual Abilities

Neuromorphic hardware excels at tasks that involve pattern recognition, sensory data processing, and cognitive functions. This enables more natural human-computer interaction. [Qualcomm's Zeroth processor](#) powers advanced gesture and voice recognition in smartphones.

Radical Energy & Cost Efficiency

Neuromorphic computing is the future of technology as it offloads compute tasks and reduces data center workloads by up to [90%](#). This directly shrinks electricity and cooling spending. Moreover, these neural processors process sensor data with significantly less energy than digital microcontroller units (MCUs).

Breakthroughs and Innovations

Funding Surges

- [Innatera](#) raised [EUR 15 million](#) to scale spiking-sensor systems on chips (SoCs) for radar & audio.
- [Rain AI](#) raised [USD 8.1 million](#) to prep the first analog in-memory SNN chip.
- [EU EBRAINS 2.0](#) expanded Europe's neuromorphic research infra by raising EUR 38 million.

Technological Breakthroughs

- [IBM NorthPole's On-chip](#) SRAM/in-memory computer eliminates memory shuttling by improving 13 TB/s on-chip bandwidth and [46.9x](#) GPU latency.

Credit: [IBM Research](#)

- [Innatera Spiking Neural Processor](#)'s mixed-signal analog SNN cores consume [500x](#) less energy and are [100x](#) faster than MCUs.
- SpiNNaker 2's [asynchronous mesh](#) accelerates distributed graphics processing units to enhance computations for neuromorphic, hybrid, and mainstream AI models.

Application Areas

Edge AI and IoT Devices

Neuromorphic hardware is ideal for edge applications as it enables ultra-low-power, real-time processing. For instance, Intel's [Loihi chip](#) has demonstrated significant power savings in real-time facial expression recognition tasks.

Aerospace and Satellite Communications

Neuromorphic hardware offers efficient on-board data processing capabilities for space applications. Studies have shown that [spiking neural networks](#) implemented on neuromorphic chips perform tasks like interference detection and beamforming with lower energy consumption.

Healthcare and Biomedical Devices

Neuromorphic processors are being explored for use in medical devices that require real-time signal processing, like prosthetics and neural interfaces. They mimic neural processing for more natural and responsive interactions between devices and biological systems.

Spotlighting an Innovator: Twistient

[Twistient](#) is a US-based startup that develops neuromorphic chips for edge AI applications. It uses a novel transistor architecture to create compute-in-memory systems that emulate the brain's synaptic behavior at room temperature.

By integrating memory and processing within the same physical layer, these chips eliminate the von Neumann bottleneck and enable real-time, event-driven computation with significantly reduced energy consumption.

Twistient's technology leverages two-dimensional materials and nanoscale fabrication techniques to achieve scalable, biorealistic architectures capable of adaptive learning.

3. Quantum Hardware Advances & Applications

Quantum hardware is advancing from the laboratory to the real world. Users have experienced error-corrected superconducting, trapped-ion, photonic, topological, annealing, and cat-qubit processors that either break the 100-qubit "utility" barrier or chart clear roadmaps to multi-thousand-qubit machines.

These advances are already powering early commercial pilots in chemistry, finance, logistics, and mobility.

Market Momentum

The global quantum computing market is projected to reach [USD 6.8 billion by 2030](#) with a 30.7% CAGR.

Credit: [Grand View Research](#)

McKinsey estimates that quantum technologies could gain up to [USD 2 trillion](#) in new value within ten years. To the government capital, Australia committed [AUD 940 million](#) to build [PsiQuantum](#)'s fault-tolerant photonic computer in Brisbane.

Further, BMW reports multiple in-house proofs of concept and lists [quantum optimization](#) as a 2025 production target.

Business Value Drivers

Accelerate R&D and Time-to-Market

Quantum simulators model atomic interactions with far greater fidelity to reduce lead-optimization cycles and de-risk expensive lab work in drug discovery. These processors also let manufacturers iterate new catalysts and solid-state battery chemistries without in-house quantum labs.

Operations and Energy Efficiency

D-Wave's 4400-qubit Advantage2 prototype solves materials-science lattice problems **25 000 times** faster than its predecessor. Neutral-atom machines operate at room temperature and reduce cryogenic OPEX and CO2 footprint compared to superconducting stacks.

Strengthen Cybersecurity and Digital Trust

Quantum-generated entropy and post-quantum-ready key material delivered through cloud APIs provide immediate, differentiable security upgrades for SaaS, IoT, and payments platforms.

Breakthroughs and Innovations

Funding Surges

- **PsiQuantum** raised **USD 620 million** to build the world's first utility-scale photonic QC.
- **QuEra**, led by **Google Quantum AI** and **SoftBank**, raised **USD 230 million** to commercialize 1000-qubit neutral-atom arrays.
- Quantum Machines targets **USD 100 million** to control electronics that link heterogeneous quantum chips.

Technological Breakthroughs

- **IBM** "Flamingo/Kookaburra" **modular 1 386-qubit chips** networked into 4158 qubits that pave the path to quantum-centric supercomputers.
- **Atom Computing's 1225-qubit neutral-atom processor** is the first universal gate-based system beyond 1000 qubits.
- **IonQ Forte** enterprise achieved **AQ 35-36 metric** that delivers commercially usable circuit depth for chemistry and ML workloads.
- **Microsoft's Majorana 1**, powered by topological qubits, utilizes a novel topological material to enable scalable, fault-tolerant quantum computing.

Application Areas

Life Sciences

Quantum hardware enables the simulation of complex molecular interactions. Its algorithm accelerates the development of treatments for neurodegenerative disease.

Roche's partnership with **Pasqal** utilizes **neutral atom quantum processors** to model complex protein-ligand interactions more accurately.

Mobility & Manufacturing

Quantum computing optimizes manufacturing processes in supply chains. Its solutions reduce CO2 emissions in supply chains by replacing them with electricity from renewable energy.

[Airbus](#) and [BMW](#)'s 2024 challenge winners used quantum solvers to [trim supply-chain CO₂](#) and paint-shop cycle times.

Finance

Quantum algorithms are being explored to enhance complex calculations such as options pricing. For instance, [Goldman Sachs](#) and [Quantum Motion](#) are prototyping [option-pricing algorithms](#) that promise real-time risk recalibration.

Logistics

The USDOT 2024 workshop cites [quantum-optimized routing](#) for EV-charger placement and freight scheduling. Also, quantum algorithms are being developed to optimize the placement of EV charging stations by considering factors like driving range and demand flow to support the growing adoption of electric vehicles.

Spotlighting an Innovator: Diraq

[Diraq](#), an Australian company, develops silicon-based quantum computing technology that integrates billions of qubits onto a single chip using standard CMOS fabrication processes.

The company employs modified MOSFET transistors, scaled down to control individual electron or hole spins within quantum dots, to enable quantum logic operations.

The company achieves high-fidelity control, with single-qubit gate operations reaching two-qubit gate fidelities. By leveraging the compactness and energy efficiency of silicon spin qubits, Diraq delivers scalable, fault-tolerant quantum processors for commercial applications.

4. Cloud Edge Convergence & 6G

The combination of cloud and edge computing enables businesses to distribute workloads while leveraging real-time analytics, low latency, and better privacy.. This means putting computing closer to users and devices while keeping the cloud's scalability.

This technology is the foundation for 6G, which is expected to have data rates of terabits per second, latency of less than a millisecond, and built-in support for AI-driven network orchestration.

The AI edge computing market size is projected to reach [USD 59 million](#) by 2030 with a CAGR of 21.2%. The 6G solutions market is projected to hit [USD 5.23 billion](#) by 2031 with a 32% CAGR.

Market Momentum

According to [ABI Research](#), North America and Asia-Pacific are [leading](#) early investments, with China accounting for over [40%](#) of global 6G patent filings and the US Department of Defense toward next-gen wireless research.

Meanwhile, Europe is advancing through programs like [Hexa-X](#), targeting 6G trials by 2028. Together, cloud-edge convergence and 6G are expected to support immersive experiences, autonomous systems, and ultra-reliable industrial automation.

Business Value Drivers

Cost and Energy Optimization

Processing data locally reduces backhaul traffic, lowers bandwidth usage, and lowers power consumption. Firms report direct OPEX reductions that drive sustainable IT operations and lower total cost of ownership (TCO).

Millisecond-Level Latency for Critical Operations

Deploying computers inside 5G cells, like [AWS Wavelength](#), eliminates [100+ ms](#) internet hops. This unlocks real-time applications like robotic automation, remote surgery, and vehicle-to-everything (V2X) safety that centralized cloud models cannot support.

Real-Time AI & Immersive Interactions

Edge nodes reduce latency by up to [50%](#) and run local AI inference that supports predictive maintenance, AR-based service manuals, and more.

Breakthroughs and Innovations

Funding Surges

- South Korea's Ministry of Science and ICT invested [USD 481.7 billion](#) in R&D on core 6G technologies.
- The Smart Networks and Services Joint Undertaking (SNS JU) receives [EUR 130 million](#) in EU funding to build first-class 6G technology capacities across Europe.

Technological Breakthroughs

- Future 6G networks will leverage [sub-THz](#) and THz frequencies for holographic communication, wireless backhaul, and high-precision sensing.
- 6G will embed AI-native network architectures to enable self-optimizing, predictive, and adaptive systems. [Nvidia](#) collaborates with telecom giants on AI-radio access networks (RAN) to boost spectral efficiency and enable real-time network slicing.

Credit: [Cadence](#)

Application Areas

Telemedicine & Remote Monitoring

Nokia's [sub-THz sensing networks](#) are investing in key components of 6G networks, such as [joint communications and sensing \(JCAS\)](#), to allow systems to sense their surroundings. This will enable no-contact heart-rate monitoring.

Autonomous Mobility

Ultra-reliable, high-bandwidth links coordinate fleets of self-driving cars, trucks, and drones. [IIT-Hyderabad and Japan's SSIC](#) developed 6 G/B5 G SDRs supporting V2X navigation at [Mobile World Congress 2025](#).

Credit: [Qualcomm](#)

Immersive Extended Reality (XR)

Real-time volumetric calls stream a caller's face and torso as a life-size hologram for enterprise collaboration, tele-learning, and customer service.

[Telefónica](#), [Ericsson](#), and [MATSUKO](#) integrated a holographic dialer into [Samsung](#) Galaxy phones in the EU-funded [6G-XR](#) project.

Spotlighting an Innovator: Celtonn

[Celtonn](#) is an Irish startup that develops high-frequency millimeter-wave semiconductors and systems. It enables ultra-fast 6G data transmission across W-band (75–110 GHz) and D-band (110–170 GHz) spectrums.

The startup's technology utilizes proprietary amplifier modules and broadband components like power amplifiers, low-noise amplifiers, upconverters, and downconverters to deliver high gain, low noise, and precise frequency control.

Celtonn's advanced mmWave solutions address the growing demand for high-bandwidth, low-latency connectivity.

5. Industrial Metaverse & Spatial Computing

Advances in simulation engines, spatial computing hardware, and edge-cloud connectivity support the industrial metaverse.

The market value of the industrial metaverse is projected to grow with a [37%](#) CAGR to 2030. Enterprises are leveraging photorealistic digital twins to plan, simulate, and optimize operations in real time. It improves efficiency, reduces emissions, and shortens time-to-market.

Spatial computing further enables these environments to function as fully interactive, AI-integrated workspaces that transform how industrial systems are designed, monitored, and scaled.

Market Momentum

The global industrial metaverse market size is projected to reach [USD 183.7 billion](#) by 2032 with a CAGR of 30.1%.

Credit: [Grand View Research](#)

The global spatial computing market is expected to reach [USD 469 million](#) by 2030. According to [Siemens](#) and [S7P Global Market Intelligence](#), [62%](#) of companies increase investment in industrial metaverse technologies.

The US leads in investment in the industrial metaverse, with [38.4%](#) of companies actively using these technologies and [35.1%](#) in the testing phase.

A study by [Deloitte](#) confirms [92%](#) of manufacturers are piloting more than 1 metaverse use case and expect between [12%](#) and [14%](#) boosts in throughput & quality.

Business Value Drivers

Productivity and Quality Gains

GenAI copilots embedded in industrial design tools reduce programming efforts. For instance, [Deloitte](#) finds executives expect [14%](#) improvements in sales, throughput, and first-time quality as metaverse pilots scale.

Accelerates Time-to-Market for New Products

By simulating entire factories and production flows in advance, companies finalize layouts, test scenarios, and identify bottlenecks before physical construction begins. [BMW](#) utilizes [NVIDIA Omniverse](#) to simulate its Neue Klasse production lines and reduce planning time by [30%](#) before laying a single brick.

Enables Scalable, Safe, and Faster Workforce Training

Immersive simulations reduce onboarding time and increase retention by allowing people to learn complex procedures without physical risk.

[Apple Vision Pro](#) is being used by [energy](#), [aviation](#), and [construction](#) firms for interactive in-situ training. [Deloitte](#) reports that over [80%](#) of manufacturers plan to use metaverse tools to achieve a [12%](#) labor productivity gain.

Breakthroughs and Innovations

Funding Surges

- [Cintoo](#) raised [37 million](#) to turn massive laser-scan files into lightweight meshes for as-built industrial metaverse twins.
- [Hololight](#) raised [USD 12 million](#) for heavy-CAD models to stream securely to head-mounted displays (HMDs) on shop floors.

Technology Breakthroughs

- NVIDIA [Omniverse Cloud API](#): It streams photorealistic, RTX-rendered digital twins directly into their simulation and automation platforms. This integration enables real-time collaboration across different tools on high-fidelity industrial models and eliminates the need for file conversions and design iteration among globally distributed teams.
- [TeamViewer Frontline Spatial](#): [GE Aerospace](#) leverages TeamViewer's augmented reality-based 3D training platform to reduce global technician onboarding time while addressing global skills shortages.

Application Areas

Workforce Training & Onboarding

Spatial computing platforms are transforming employee onboarding and upskilling into interactive, high-fidelity experiences.

[Hololight Hub](#) enables heavy CAD models to be streamed to lightweight XR devices. It provides trainees access to life-size simulations without high-end hardware.

Logistics & Warehouse Digital Twins

By simulating material flow, robot movement, and operational layouts, logistics operators optimize systems before physical rollout.

[Dematic](#)'s AI Control Tower, revealed at [NVIDIA GTC 2025](#), models AMR fleet behavior and facility layouts to streamline intralogistics ahead of deployment.

Cross-Enterprise Design Collaboration & Review

Photorealistic digital twins enable globally distributed teams to collaborate on complex product designs and system models in real time.

Spotlighting an Innovator: UNWARE

German-based startup [UNWARE](#) develops AI-driven tracking software that powers the *UNWARE Holodeck*, a portable cave automatic virtual environment (CAVE). It offers immersive virtual reality experiences without headsets or controllers.

The startup employs camera-based motion capture and deep learning algorithms to track full-body movements in real time. The software synchronizes user actions with the virtual environment.

Its intuitive gesture control supports collaborative, multi-user interactions without physical wearables. The Holodeck's modular design and portability facilitate industrial design, defense training, real estate visualization, and retail engagement.

6. Clean Energy Expansion

The global shift toward sustainable energy is gaining momentum due to green hydrogen, fusion power, and solid-state batteries. These innovations are central to decarbonizing key sectors.

These are gaining popularity due to growing demands for safer, high-density storage in mobility, electronics, and grid systems.

Market Momentum

The green hydrogen market is set to grow to [USD 199.22 billion](#) by 2034, expanding at a CAGR of [41.46%](#), as industries adopt zero-emission fuel alternatives for transport and manufacturing.

Credit: [Precedence Research](#)

Fusion power, on the other hand, offers abundant, carbon-free baseload energy that could surpass [USD 350 billion](#) by 2050.

Meanwhile, solid-state batteries have market projections rising from USD 1.63 billion in 2025 to [USD 19.14 billion](#) by 2033 at an impressive [36.03%](#) CAGR.

The Asia-Pacific region leads the green hydrogen market with a [47.22%](#) share. The market is fueled by India's National Green Hydrogen, which is targeting [5 million tonnes/year](#) by 2030.

Credit: [Precedence Research](#)

France invested [EUR 7.2 billion](#) in public investment to develop a decarbonized hydrogen industry. Also, the US Department of Energy (DOE)'s [Hydrogen Shot](#) initiative targets [USD 1/kg](#) production costs by 2031.

The fusion energy sector is expected to exceed [USD 350 billion](#) by 2050 if technological milestones are achieved. Private fusion investment surpassed [USD 7 billion](#) by 2025, led by Commonwealth Fusion Systems (CFS).

Further, the UK's [STEP prototype](#) targets grid-connected fusion by 2035.

Solid-state batteries are projected to make a major impact in the EV battery market, as supported by [Toyota's](#) 1000 km-range prototypes and [QuantumScape's](#) [15-minute](#) charging cells.

Business Value Drivers

Green Hydrogen

Decarbonization

Green hydrogen offers a zero-emission substitute for fossil fuels in hard-to-abate industries like steelmaking and power generation. Sweden's [HYBRIT](#) project cut CO2 emissions by at least [10%](#) using hydrogen-based iron reduction, with SSAB targeting fossil-free steel production by 2030.

Balancing Renewable Supply with Hydrogen Storage

By converting surplus solar and wind energy into storable hydrogen, firms stabilize grid operations. Luxembourg's [Energy Dome](#) facility stores [200 MWh](#) of solar energy with 75 %+ round-trip efficiency.

Fusion Power

Abundant, Low-Carbon Baseload Energy

Fusion power promises a virtually inexhaustible source of clean electricity with negligible emissions.

Inherent Operational Safety

Fusion systems, unlike fission reactors, avoid meltdown risks and do not produce long-lived radioactive waste. [General Fusion](#)'s magnetized target fusion approach is an example of this safety profile.

Solid-State Batteries

High-Energy Storage for Advanced Mobility

Solid-state batteries generate twice the energy density of lithium-ion cells, which accelerates electric aviation and long-range EVs. [NASA](#)'s SABERS project reached [500 Wh/kg](#) with a solid-state lithium-sulfur battery.

Improved Safety and Durability

With solid electrolytes replacing flammable liquids, these batteries drastically reduce the risks of thermal runaway, leaks, or combustion.

Affordable, Long-Duration Grid Storage

Form Energy's iron-air solid-state batteries offer [100-hour](#) storage at just USD [20/kWh](#). This approach provides a cost-effective solution to intermittency challenges in wind and solar integration.

Breakthroughs and Innovations

Funding Surges

Green Hydrogen: India approved an [INR 19 744 crore](#) national incentive plan to scale green hydrogen production.

Fusion Power: Over [USD 7.1 billion](#) invested in fusion energy, with government funding up 57% to [USD 426 million](#) in a single year.

Technology Breakthroughs

Green Hydrogen

- **PEM Dominance:** Proton Exchange Membrane (PEM) electrolyzers forecast a [USD 2.4 billion](#) market share.
- **Manufacturing Expansion:** [Industrie De Nora](#) began building Italy's [largest](#) electrolyzer manufacturing facility, with demand already matching its annual output capacity.

Fusion Power

- **Helion Energy** received [USD 425 million](#) and announced plans to construct a [500 MW](#) fusion plant in Everett, Washington. It targets operational readiness by 2030.
- **High-Temperature Superconductors:** [Commonwealth Fusion Systems](#)' SPARC tokamak uses $\text{YBa}_2\text{Cu}_3\text{O}_7$ magnets to achieve [20 Tesla](#) confinement and double conventional magnetic field strengths.

Solid-State Batteries

- [NASA Advancements](#): NASA is engineering ultralight, high-density [solid-state batteries](#) for next-gen electric aircraft propulsion.

Application Areas

Green Hydrogen

Heavy Industries

Green hydrogen is being deployed in the transportation, aviation, shipping, and industrial sectors to replace fossil fuels and cut emissions.

[Maersk](#)'s methanol-powered vessels reduce absolute scope 1 and 2 GHG emissions by [96%](#) by 2040. [ZeroAvia's ZA600](#) engines, which are powered by [600 kW](#) hydrogen fuel cells and support [300 NM](#) commercial flights in 10 to 20-seat aircraft by 2026.

In heavy industry, green hydrogen is increasingly used in steelmaking and chemical manufacturing to displace carbon-intensive processes.

Fusion Power

Industrial Heat & Materials Manufacturing

Heavy industries like steel, cement, and chemicals require temperatures that renewables can't reliably deliver. Fusion's high-grade heat (above 1000 degrees Celsius) could power direct electrification of kilns, forges, and reactors.

Solid-State Batteries

Consumer Electronics and Wearables

In smartphones, laptops, and wearables, solid-state batteries enable longer battery life, slimmer form factors, and greater safety due to their non-flammable solid electrolytes.

Companies like [Samsung](#) and [Apple](#) are exploring the [integration of these batteries](#) to power devices that support high-refresh displays, AI processing, and always-on connectivity without sacrificing battery endurance.

Spotlighting an Innovator: HYDGEN

[HYDGEN](#) is a Singaporean company that builds advanced green hydrogen electrolyzers. The startup utilizes proprietary catalysts, membranes, and stack designs to achieve high energy efficiency and cost-effectiveness.

It's anion exchange membrane (AEM) technology operates without precious metals like iridium or platinum. The electrolyzers deliver hydrogen purity that is suitable for semiconductors, pharmaceuticals, and steel manufacturing.

With a modular design, the systems also offer operational flexibility and scalability, catering to small and mid-sized industries. HYDGEN's on-site hydrogen generation further minimizes reliance on external supply chains and contributes to industrial decarbonization.

7. Synthetic Biology

Metabolomics therapeutics and synthetic biology are two separate but related fields.

Synthetic biology is the study of how to design and build new biological parts and systems. Metabolomics, on the other hand, looks at all the metabolites in a sample to learn more about how diseases work and how to make treatments that work better.

Together, these two fields are changing healthcare, agriculture, and manufacturing.

The synthetic biology market is expected to grow from [USD 24.58 billion](#) in 2025 to [USD 192.95 billion](#) by 2034, with a CAGR of [28.63%](#).

Credit: [Precedence Research](#)

At the same time, the metabolomics therapeutics market will grow from [USD 3.64 billion](#) to [USD 11.32 billion](#) over the same time, with a CAGR of [13.43%](#).

Credit: [Biospace](#)

Market Momentum

The synthetic biology market received [USD 12.2 billion](#) in venture capital funding in 2024. North America leads with a [39.6%](#) market share, driven by initiatives like the [Seattle Hub for Synthetic Biology](#). It secured [USD 10 million](#) this year to advance mammalian cell engineering and support startup formation.

The US market is expected to reach [USD 41.26 billion](#) by 2033. The EU's bioeconomy strategy has mobilized [EUR 4 billion](#) for synthetic biology applications in green chemistry.

Credit: [Precedence Research](#)

The metabolomics services market is forecasted to reach [USD 8.6 billion](#) by 2034, driven by demand for cancer diagnostics and pharmaceutical R&D. North America leads this market as well, with Metabolon's [USD 60 million](#) credit facility and NIH-backed [initiatives](#).

Credit: [Insight Ace Analytic](#)

Asia-Pacific is the fastest-growing region as India and China deploy NMR and mass spectrometry technologies for national-scale metabolic profiling and precision medicine efforts.

Business Value Drivers

Synthetic Biology

Bio-Manufacturing Revolution

Engineered microbial systems produce [60%](#) of industrial enzymes and reduce chemical synthesis costs.

For example, [Ginkgo Bioworks](#) has accelerated production cycles by engineering yeast strains to synthesize flavor compounds. It reduced manufacturing timelines from 18 months to [6 months](#).

Improved Yield and Sustainability

[Pivot Bio](#)'s nitrogen-fixing microbes increase crop yields while significantly lowering synthetic fertilizer use. This positions synthetic biology as a sustainable solution to a [USD 200 billion](#) global fertilizer market.

Programmable Therapeutics

Synthetic biology powers CRISPR-based gene circuits for programmable cell therapies. Many clinical trials include synthetic biology components, especially for cancer immunotherapy, that show growing adoption in advanced medical pipelines.

Metabolomics Therapeutics

Personalized Medicine

By decoding individual metabolic profiles, metabolomics enables tailored treatment plans that reflect a patient's unique biochemical state.

[Bruker](#)'s 15T NMR platforms identified numerous metabolites in cerebrospinal fluid from Alzheimer's patients. The platform spotlights dysregulated purine pathways as new therapeutic targets.

Early and Accurate Disease Detection

Metabolomics improves diagnostic accuracy by identifying disease-specific metabolic biomarkers.

Breakthroughs and Innovations

Funding Surges

Synthetic Biology: Venture capital investment in synthetic biology reached [USD 12.2 billion](#) year-to-date. This increase is primarily driven by renewed interest in biofoundries, precision fermentation, and programmable therapeutics.

Metabolomics: [Partnerships and investments](#) are speeding up advances in metabolomics. For instance, [Metabolon](#) secured a [USD 60 million](#) credit facility to scale research, expand analytics capabilities, and grow its footprint in precision diagnostics and pharmaceutical R&D.

Technology Breakthroughs

Synthetic Biology

- **SynTCE:** Researchers developed the [Synthetic Translational Coupling Element](#) to enable complex synthetic pathways like artemisinin biosynthesis in *E. coli*.
- **Rapid DNA and mRNA Synthesis:** [Twist Bioscience](#)'s silicon-based gene printer achieves [1.6 million](#) base pairs/day at [USD 0.07 per base](#). This reduces turnaround times and costs.

Metabolomics Therapeutics

- **Advanced Analytical Tools:** The integration of high-resolution [mass spectrometry with nuclear magnetic resonance spectroscopy](#) has significantly expanded metabolite detection. It enables broader metabolic coverage in disease profiling.

Application Areas

Synthetic Biology

Agriculture

[Pivot Bio](#) develops [microbial nitrogen](#) fertilizers that enhance yields while reducing synthetic fertilizer use.

[Indigo Ag](#) improves biofertilizer effectiveness using metabolomics-guided microbial inoculants.

Food

In the food industry, [Ginkgo Bioworks](#) engineers yeast strains for custom production of flavors and fragrances that streamline fermentation-based manufacturing.

For industrial bioproduction, [LanzaTech](#)'s CO₂-to-ethanol pathway is optimized with machine learning to achieve carbon conversion efficiency and scales to [100 million gallons/year](#).

Metabolomics Therapeutics

Diagnostics

In oncology, [Metabolon](#) partners with pharmaceutical companies to identify disease-specific metabolic biomarkers.

Drug Development

Metabolomics strengthens drug development pipelines. For instance, [Moderna](#)'s AI model was trained on numerous metabolic profiles to optimize mRNA vaccine stability and increase durability.

Hepatocyte-based metabolite screening also achieves [89%](#) specificity in detecting drug-induced liver injury, saving clinical trial attrition costs.

Spotlighting an Innovator: Metastate Bio

[Metastate Bio](#) is a US-based company that develops an in-silico simulation and optimization platform for bioprocess engineering. It enables researchers to model and refine microbial systems such as E. coli, yeast, and CHO cells without extensive laboratory experimentation.

The platform integrates genome-scale metabolic modeling, flux balance analysis (FBA), and genetic algorithms to simulate the effects of environmental conditions and genetic configurations on cellular performance.

This assists companies in identifying metabolic bottlenecks, optimizing production conditions, and minimizing byproducts. This way, Metastate Bio streamlines bioprocesses and provides a customizable toolbox for in-silico experiments.

8. Humanoid & Collaborative Robots

Humanoids replicate human motion and cognition to handle tasks ranging from assembly line support to patient care. Meanwhile, collaborative robots are enabling agile workflows in factories, hospitals, retail, and education.

Market Momentum

The global humanoid robot market is projected to rise from [USD 2.92 billion](#) in 2025 to [USD 15.26 billion](#) by 2030 with a CAGR of [39.2%](#).

Credit: [Precedence Research](#)

Simultaneously, the cobot market is set to expand from USD 5.58 billion to [USD 71.26 billion](#) by 2034, growing at a [32.7%](#) CAGR.

The US humanoid robot market is set to grow to [USD 3.83 billion](#) by 2029, at a CAGR of [45.7%](#), driven by investments from [OpenAI](#), [NVIDIA](#), and [Microsoft](#).

In healthcare, cobots improve rehabilitation and assistive care, while the construction industry reduces worksite injuries and accelerates project timelines with cobots. Europe is the global leader in cobot adoption.

Business Value Drivers

Labor Shortage Mitigation

Humanoid robots and cobots directly address global labor shortages by automating repetitive and physically strenuous tasks. Cobots are expected to fill critical gaps, with over [64 000](#) deployed per year in North America by 2030.

Cost Efficiency and Rapid ROI

Cobots reduce production costs in sectors like automotive assembly. For instance, the payback period for the UR10 robot installed at Orkla is just [195 days](#).

AI-Driven Operational Optimization

Machine learning enhances cobot performance by dynamically allocating tasks based on real-time production needs. This intelligent coordination improves assembly line throughput and supports agile manufacturing and scalable operations.

Breakthroughs and Innovations

Funding Surges

Humanoid Robots: Robotics-focused startups secured investments of approximately [USD 9.7 billion](#) in 7 months, with a large share directed toward humanoid systems. For instance, [Ray Kurzweil's humanoid robot](#) startup is in talks for a [USD 100 million](#) investment.

[Physical Intelligence](#) raised [USD 400 million](#) at a [USD 2 billion](#) valuation to accelerate the development of general-purpose humanoid robots.

Collaborative Robotics: [Collaborative Robotics](#). closed a [USD 100 million](#) Series B round to scale cobots.

Technology Breakthroughs

Humanoid Robots:

- [Tesla Optimus Gen 2](#): Demonstrated a [20 kg](#) payload capacity and 5-hour operational runtime.
- [Boston Dynamics' Atlas](#): Transitioned to a fully electric system with [360-degree](#) joint rotation. It enables complex manipulation tasks tailored to construction environments.
- [Berkeley Humanoid Lite](#): Introduced a fully open-source design that uses 3D-printed cycloidal gears to reduce hardware costs and can be assembled under [USD 5000](#).

Collaborative Robotics

- [Universal Robots' PolyScope X](#): Released a no-code interface that simplifies cobot programming and reduces deployment timelines.
- [NVIDIA's Project GR00T](#): Developed a generative AI foundation model that enables humanoids and cobots to learn new tasks from a single demonstration.

Application Areas

Manufacturing

[Agility Robotics'](#) *Digit* is deployed in [Amazon](#) warehouses to retrieve and carry bins. This reduced item retrieval time by [25%](#) through advanced bipedal mobility.

Healthcare

Cobots in healthcare support medication dispensing and improve both care delivery and staff efficiency. For instance, [Mentee Robotics'](#) *Stevie* robot assists elderly patients with companionship and fall detection while reducing caregiver workload.

Retail and Hospitality

Humanoid and cobot systems elevate customer interactions in service environments. [SoftBank's Pepper](#) is deployed in many retail stores for personalized recommendations. It increased sales conversions by [13%](#) in revenue and a [6-fold](#) increase in sales.

Spotlighting an Innovator: Robots For Humanity

[Robots For Humanity](#) is an Argentine company that offers AI-driven humanoid robotics integrated with digital twin technology to automate high-risk and physically demanding tasks.

The startup's robots undergo training and optimization within hyper-realistic virtual simulations powered by platforms like NVIDIA Omniverse. This enables precise task execution before real-world deployment.

Consequently, this approach enhances safety, reduces operational costs, and liberates human workers from hazardous and repetitive duties.

9. Cyber-Resilient Data-Centric Security

Cyber-resilient, data-centric security frameworks protect the data regardless of where it resides or how secure the surrounding infrastructure may be. It embeds encryption, access controls, and policy enforcement directly into data assets to ensure continuous protection across cloud, edge, and on-premise environments.

Data-centric models offer resilience against ransomware, AI-generated phishing campaigns, and expanding hybrid architectures. They align with zero-trust principles and support regulatory compliance for handling sensitive personal, financial, and operational data.

Market Momentum

The global data-centric security market is projected to grow at a CAGR of [24.2%](#), reaching [USD 24.01 billion](#) by 2030.

Credit: [Mordor Intelligence](#)

In parallel, the broader cybersecurity market will expand to [USD 424.14 billion](#) by 2030.

North America holds a [40%](#) market share as of 2024 due to the strict enforcement of CCPA and HIPAA.

The data-centric security market is set to grow at a CAGR of [21.4%](#) through 2030.

Europe is witnessing an increase in compliance-driven security spending, with GDPR fines totaling [EUR 1.2 billion](#) in 2024.

[SAP's Data Custodian platform](#) reported GDPR compliance adherence across EU enterprises. This reflects increased adoption of data masking and classification tools.

Business Value Drivers

Enhanced Data Protection at the Core

Focusing security controls directly on the data ensures that sensitive assets remain encrypted and access-controlled, even if networks are compromised.

This strategy reduces breach risks and safeguards intellectual property. [LevelBlue](#) highlights this strategy as a critical defense pillar for modern enterprises.

Zero Trust Integration for Granular Control

Data-centric security aligns seamlessly with zero trust principles that leverage continuous authentication, least-privilege access, and microsegmentation.

For example, [Palo Alto Networks' Prisma Cloud](#) applies data classification tags to SaaS transactions that cut lateral movement risks and strengthen end-to-end visibility.

Cost Optimization via Exposure Reduction

Organizations deploying data-centric security report a [USD 4.88 million](#) reduction in breach remediation costs. This cost efficiency stems from minimized data exposure and faster incident containment.

Breakthroughs and Innovations

Funding Surges

- **Venture Capital Growth:** Cybersecurity startups raised [USD 2.7 billion](#) in Q1 2025. This marks a [29%](#) quarter-over-quarter increase. Data-centric security platforms such as [Torus](#) secured [USD 67 million](#) in venture financing to advance secure data collaboration tools.
- **Israeli Cyber Surge:** Israeli cybersecurity firms attracted [USD 4 billion](#) in funding.

Technology Breakthroughs

- [Seclore](#) was named [“Data Security Solution of the Year”](#) in 2024 for its innovative approach that embeds protection within the data itself, rather than relying on securing the surrounding infrastructure.
- [Trend Micro](#) released an AI-powered security suite that autonomously predicts, identifies, and neutralizes threats.

Application Areas

Government

Governments are adopting post-quantum cryptography and zero-trust frameworks to protect classified datasets and mission-critical systems. [NIST-approved CRYSTALS-Kyber algorithms](#) provide resilience against future quantum decryption threats.

Smart Cities

Barcelona invested [USD 150 million](#) in a resilient city initiative that secures traffic, water, and energy systems by encrypting IoT data lakes. This smart infrastructure ensures both service continuity and real-time anomaly detection in distributed sensor networks.

Spotlighting an Innovator: Praeryx

[Praeryx](#) is a Japanese company that offers a cyber threat intelligence solution. Its proprietary agentic intelligence, *Fabric*, integrates artificial intelligence and customized large language models (LLMs) to analyze and provide contextualized and timely threat insights.

This system deciphers complex cyber threats by eliminating legacy biases and offering localized intelligence in diverse geographies. Additionally, Praeryx’s Cyber Risk Quantification framework translates abstract cyber risks into measurable metrics. This enables enterprises to anticipate threats and strengthen cyber defenses

10. Space-Based Manufacturing & Energy

Space-based manufacturing and energy leverage microgravity, vacuum purity, and continuous solar exposure to redefine industrial processes.

Microgravity enables the fabrication of ultra-pure materials, like optical fibers and semiconductors, while space-based solar power (SBSP) offers a resilient energy supply unconstrained by weather or day-night cycles.

The convergence of these technologies is driving innovation across pharmaceuticals, advanced electronics, and clean energy infrastructure.

Market Momentum

The in-space manufacturing market size is expected to reach [22.47%](#) CAGR through 2030. Japan deployed gallium arsenide (GaAs) solar cells that demonstrated [32%](#) energy conversion efficiency in prototype satellites this year.

Business Value Drivers

Material Science Revolution

Semiconductor wafers are manufactured in orbit with precision using microgravity casting. Such precision is critical for scaling photonic computing and quantum devices, where surface integrity and alignment are paramount.

Breakthroughs in Optical Fiber Technology

[Flawless Photonics](#) has demonstrated that ZBLAN optical fibers fabricated in space deliver signal losses as low as [0.001 dB/km](#). This advancement could transform global communications by enabling ultra-low-latency, high-bandwidth networks.

Cost-Effective Lunar Industrialization

Innovations in lunar infrastructure dramatically lower space exploration costs. Lunar-sourced liquid oxygen (LOX), for instance, lowers the cost and launch burden of Mars missions through in-situ resource utilization (ISRU). Meanwhile, KUKA's MMIC-I robotic swarms autonomously build modular lunar bases that reduce construction costs.

Breakthroughs and Innovations

Funding Surges

Private Sector Momentum: 5 companies focused on space infrastructure and defense technologies collectively raised [USD 513 million](#).

[Eascra Biotech](#) has been awarded a US National Science Foundation Small Business Innovation Research grant. The grant is valued at [USD 274 million](#) to fund targeted microgravity drug production and next-gen GaN-based high-voltage power systems.

Technology Breakthroughs

- [Eascra Biotech](#) leverages [microgravity](#) to enhance targeted drug delivery via Janus Base Nanoparticles (JBNs). This will improve treatments for cancer and osteoarthritis.
- [Space Solar](#), backed by over [GBP 10 million](#) in government funding, is developing orbital solar power stations assembled by autonomous robots. The planned prototype

will transmit energy to Earth using radio waves, which will be generated by large-scale solar arrays assembled by robots.

Application Areas

Semiconductor Production

Microgravity is now being explored as a platform for precision semiconductor fabrication. [Sierra Space collaborated with Astral Materials and Space Forge](#) to develop orbital chip manufacturing systems. It utilizes gravity-free conditions for improved crystal quality and defect reduction.

Renewable Energy: Space-Based Solar Power and Orbital Sustainability

[Space Solar](#)'s orbital energy platforms deliver continuous, clean power to Earth. With robotic assembly of large-scale solar arrays in orbit, this initiative supports global decarbonization.

Spotlighting an Innovator: AndroMach

[AndroMach](#) is a French company that provides reusable spaceplanes for suborbital and orbital missions. Its suborbital vehicle, *ENVOL*, employs turbojet engines for atmospheric flight and a pressure-fed LOX/biopropane rocket engine for space ascent.

The orbital shuttle, *ETOILE*, launches atop small partner rockets that support missions lasting from days to months. This way, AndroMach makes microgravity applications more accessible and advances space-based industrial capabilities.

Explore Disruptive Future Technology Trends & Solutions

With thousands of emerging technologies and startups, navigating the right investment and partnership opportunities that bring returns quickly is challenging.

With access to over 7 million emerging companies and 20K+ technologies & trends globally, our AI and Big Data-powered [Discovery Platform](#) equips you with the actionable insights you need to stay ahead of the curve in your market.

Leverage this powerful tool to spot the next big thing before it goes mainstream. Stay relevant, resilient, and ready for what is next.

Data-Driven Design Analyst: Elevate Product Innovation

Data-Driven Design Analyst: Elevate Product Innovation

In the rapidly evolving landscape of business, the integration of business intelligence and data analytics with design services is no longer just an advantage—it is a necessity. For a Design Analyst, mastering the art and science of product design and development means leveraging data to craft innovative solutions that meet market needs. This article explores the pivotal role of Design Analysts, the methodologies they use, and how platforms like [DataCalculus](#) empower professionals to transform raw data into actionable insights, ultimately driving product innovation and delivering exceptional user experiences.

Introduction: Bridging Design and Data

In today's competitive market, product design and development are intertwined with data analytics. Organizations modernize their design processes by combining creative strategy with meticulous data mining, enabling them to create products that are not only aesthetically appealing but also functionally efficient and market responsive. Business intelligence tools and data-driven approaches help design teams forecast trends, analyze customer behavior, and fine-tune their offerings.

At the heart of this innovative transformation is the Design Analyst—a professional who uses a blend of creative acumen and analytical prowess to optimize design strategies and product development cycles. This role is more than just a designer's job: it catalyzes the connection between market insights and design execution, setting the stage for revolutionary product experiences. By systematically integrating data into creative processes, Design Analysts ensure that every design decision is supported by quantifiable insights.

The digital age has brought forward platforms that centralize these critical functions. For instance, [Data Dictionary](#) functionalities enable a deeper understanding of datasets, streamlining the identification of vital metrics in the design process. With such tools, design decisions are no longer based solely on intuition but on solid, data-backed evidence.

The Role of a Design Analyst in Product Design and Development

Design Analysts play a significant role in product design and development within organizations. They combine expertise in user experience (UX), graphic design, and data analysis to create products that are not only visually attractive but also aligned with market expectations. They are tasked with scrutinizing every element of design—from typography and layout to functionality and usability—to ensure that products resonate with target audiences.

A quintessential Design Analyst uses several data-centric techniques to inform design decisions:

- **Market Analysis:** Understanding customer preferences and emerging trends can help anticipate shifts in market demand. With data analytics, even subtle trends can be harnessed into strong design strategies.
- **Usability Analytics:** By studying user interactions and engagement metrics, these professionals identify areas where improvements are needed. Tools such as the [Overall AI Report](#) provide insights that help gauge user sentiment regarding design elements.
- **Feedback Incorporation:** Gathering and analyzing user feedback ensures that design iterations are continuously improved.

When DataCalculus enters the picture, the ability to integrate vast datasets with real-time feedback becomes even more streamlined. The platform's robust capabilities—ranging from data cleaning to predictive analytics—empower Design Analysts to pinpoint opportunities for improvement and innovation effortlessly. This integration facilitates a design process that is agile, iterative, and highly responsive to market dynamics.

Integrating Business Intelligence into Design Services

Business intelligence (BI) is the backbone of informed decision-making in organizations of all sizes. In design services, BI methodologies serve to uncover trends that can make or break a product launch. The integration of BI into design services entails the collaboration of interdisciplinary teams—the creatives and the analysts—who work in tandem to transform raw data into creative outputs.

Some of the ways BI and data analytics improve design services include:

- **Identifying Core Trends:** Using BI, teams can explore hidden trends in market data. By deploying a [Pattern Report](#), companies can unearth subtle correlations between design elements and user engagement patterns that might otherwise go unnoticed.
- **Strategic Insights:** Design Analysts employ data to analyze competitor strategies and industry benchmarks, ensuring that their design choices are competitive and unique. This involves a detailed review of user engagement metrics, conversion rates, and overall market performance.
- **Data-Driven Refinements:** Refining products through iterative design cycles driven by real-time analytics equips organizations with the ability to quickly pivot and respond to market feedback. A focused review using tools like the [Classification Report](#) can isolate factors that influence user satisfaction and product performance.

Moreover, tools such as the [Clustering Report](#) can segment audiences and design solutions to cater to specific user needs. These applications of BI make design services not only reactive but also anticipatory, paving the way for the creation of breakthrough products.

Data Analytics: The Catalyst for Innovative Product Development

Data analytics is at the core of modern design strategies, transforming traditional decision-making into a competitive advantage. In the context of product design and development, data analytics helps pivot the focus from aesthetics alone to functionality coupled with user satisfaction. The process begins with data collection and culminates in actionable insights that influence the design lifecycle.

Key factors that shape how data analytics underpins product development include:

- **User Experience Metrics:** Analyzing metrics such as click-through rates, time-on-page, and engagement levels helps identify which design elements are successful. A deep dive using a [Data Scientist AI](#) approach can yield nuanced insights into consumer behavior.
- **Prototyping and Iterative Testing:** Data from A/B testing and usability studies allow teams to iterate on design concepts rapidly. As a result, prototypes evolve into product-ready designs that accurately reflect user expectations.

- **Predictive Analytics:** Forecasting future trends based on historical data is an essential part of design strategy. Platforms like DataCalculus offer predictive functionalities that guide the creative process and help design professionals stay ahead of the curve.

By continuously incorporating feedback loops, design teams remain agile in the face of changing market dynamics. This process can be further enriched by leveraging the company's [Support AI](#) capabilities to provide personalized assistance and troubleshoot design issues as they arise.

Understanding the Data-Driven Design Process

The modern design process involves several iterative stages, each enriched by data analytics and business intelligence. These stages include research, conceptualization, prototyping, testing, and final product development. Within each stage, data serves multiple purposes: inspiration, validation, and optimization.

One of the key stages is the ideation phase. Design Analysts work to meld creative ideas with user data. This ensures that initial design sketches are informed by trends spotted via deep data dives. An efficient approach is to consider the [Dataset Operations](#) available, allowing teams to manage and manipulate vast amounts of data seamlessly. Once design ideas are ready for prototyping, the process moves into rapid iterations where data plays an essential role in fine-tuning each element of the product.

As development advances, systematic testing using methods such as user surveys, focus groups, and digital analytics becomes indispensable. Results from testing sessions are analyzed using [Bulk Operations](#) techniques, enabling aggregated insights that drive refinements. Ultimately, the finished product embodies both creative excellence and data-validated functionality, underscoring the dual importance of design and analytics.

Integrating DataCalculus in the Design Workflow

A critical consideration for any design team today is the selection and integration of advanced analytics platforms. Among the few that support this integration, DataCalculus stands out by providing a comprehensive suite designed for data-driven product innovation. This platform simplifies the transition from raw data to insightful reports, ensuring that each stage of the design process is backed by robust analytics.

For example, when faced with complex datasets, the [Data Dictionary](#) feature of DataCalculus enables teams to quickly explore dataset structures, column types, and associated statistics. This functionality is pivotal in establishing a baseline understanding before venturing into deeper analytics.

Furthermore, DataCalculus offers a suite of reporting tools that makes it simple for design teams to visualize trends. The [Report Assembly](#) function, for instance, pulls together comprehensive reports which become key decision-making documents. With such a strong analytics backbone, teams can confidently navigate the product design cycle from ideation through to launch.

Best Practices for Data-Driven Design in Product Development

The journey toward becoming an effective data-driven Design Analyst involves adopting a series of best practices that center around continuous learning, iterative improvement, and seamless collaboration. Below are some practices that can help integrate data analytics into the design process:

1. **Establish Clear Objectives:** Define what success looks like for each design project. Use quantitative targets, such as conversion rates and engagement levels, combined with qualitative insights to set clear objectives.
2. **Leverage Comprehensive Data Sources:** Pull together customer insights, market trends, and usability metrics to create a holistic view. Access aggregated data through robust platforms such as [DataCalculus](#) to optimize the decision-making process.
3. **Implement Iterative Design Cycles:** Adopt a methodology where design is an evolving process. Regular iterations grounded in data feedback ensure that the design keeps pace with changing market demands.
4. **Foster Cross-Department Collaboration:** Encourage collaboration between designers, data analysts, and marketing teams. Collaborative tools like [Team Chat](#) facilitate efficient communication and data sharing across different units.
5. **Utilize Visual Reporting:** Translate complex analytics into visual dashboards that underscore design insights. Regularly review these dashboards to align design decisions across the organization.
6. **Monitor and Adapt:** Stay agile by constantly monitoring design performance and adapting strategies based on real-time data. Use

diagnostic analytics from the [Overall AI Report](#) to assess and refine your approach continuously.

In addition to the above practices, strategic decisions should integrate deeper analysis tools. For instance, a [Classification Report](#) can be used to identify key factors that are influencing both the design process and user engagement. This granular understanding helps in tailoring design solutions that cater specifically to distinct audience segments.

Overcoming Challenges in Data-Driven Design

While the benefits of integrating data analytics into design are profound, organizations must be proactive in addressing the inherent challenges. Some common obstacles include:

- **Data Quality and Consistency:** Ensuring that the data collected is clean, consistent, and relevant can be challenging. Regular audits and validation checks are necessary to maintain data integrity.
- **Cross-Functional Communication:** Bridging the gap between creative teams and technical data experts requires a culture of seamless collaboration, which can sometimes be difficult to achieve.
- **Complexity of Data Analysis:** The rapid growth of data, along with its inherent complexity, can overwhelm design teams. This is where automated solutions by systems like DataCalculus become critical, streamlining data operations and reducing manual effort.
- **Adoption of New Technologies:** Resistance to change and the learning curve associated with new data analytics platforms can slow down innovation. However, progressive training programs and pilot projects can encourage smoother transitions.

Addressing these challenges involves a combination of technology, training, and cultural shift. Embedding a data-centric mindset within the design team not only mitigates these challenges but also turns them into opportunities for growth. The implementation of capabilities such as the [Bulk Operations](#) can simplify large-scale data handling, ensuring that even extensive datasets are manageable and insightful.

Future Trends: The Evolution of the Design Analyst Role

The role of a Design Analyst is set to evolve even further as new technologies and data methodologies continue to emerge. Future trends in product design and development powered by business intelligence include:

- **Advanced Predictive Analytics:** The use of artificial intelligence to predict design trends before they emerge, allowing companies to stay ahead of their competitors.
- **Greater Personalization:** With more detailed data segmentation available from tools like a [Clustering Report](#), products can be further customized to meet the nuanced demands of target markets.
- **Seamless Integration of AI:** The incorporation of AI-driven tools, such as [Data Scientist AI](#) and [Support AI](#), into everyday design tasks will reduce manual effort and enhance creative outputs.
- **Collaborative Ecosystems:** Future design platforms will likely offer even more integrated environments where cross-functional teams use shared dashboards, insights, and real-time communications. The use of [Admin Tools](#) along with [Team Chat](#) functionalities will support stronger collaborative efforts.

As data analytics becomes even more embedded in product design, Design Analysts will find themselves less focused on traditional tasks and more on strategic innovation. Bidirectional feedback loops between creative design and data analysis are expected to fuel breakthroughs in both user experience and product functionality. The integration of a robust system like DataCalculus not only empowers designers with the right insights but also facilitates experimentation and risk-taking, ultimately carving out new frontiers in design innovation.

Case Studies and Real-World Applications

Understanding the transformative power of a data-driven approach is enhanced by examining real-world applications. Numerous organizations have harnessed the capabilities of business intelligence and data analytics to revolutionize their product development processes. Consider the following illustrative case studies:

Case Study 1: Streamlining Product Prototyping

A leading consumer electronics firm decided to overhaul its design process by integrating comprehensive data analysis during the prototyping phase. The Design Analysts collected real-time user engagement data and used iterative

testing protocols to refine the product prototype. The company leveraged its [Report Assembly](#) module to gather insights across different demographic segments. These insights allowed the team to adjust design elements such as interface layout and interactive features, leading to a significant decrease in user friction and a noticeable uptick in product adoption rates once launched.

Case Study 2: Personalizing User Experience in Digital Platforms

Another success story comes from a tech startup that focused on delivering a highly personalized user experience for its digital platform. The Design Analysts on the team used clustering algorithms provided by the platform's [Clustering Report](#) tool to segment users based on behavior and preferences. This granularity in understanding user groups allowed them to design targeted interfaces, personalize content, and enhance the overall user journey. Additionally, the startup utilized an integrated approach, incorporating real-time feedback through [Team Chat](#) systems that connected various stakeholders from creative and technical backgrounds.

Case Study 3: Enhancing Visual Design through Data Trends

A major design agency embraced data trends in visual design by tapping into proprietary analytics for understanding customer trends. By using tools for generating a [Pattern Report](#), the team was able to detect subtle shifts in color schemes and layout preferences across various markets. This allowed them to proactively adjust their design templates, resulting in higher client satisfaction and repeat business. Such proactive measures underscore the importance of incorporating robust data measures in the creative process.

Implementing a Data-Driven Culture in Your Organization

Achieving success through data-driven design requires not just the right tools, but also a cultural shift within organizations. Leaders must encourage their design teams to embrace data analytics as an integral component of the creative process. Here are some practical steps to foster a data-driven culture:

- **Invest in Training:** Equip design teams with the knowledge and skills to interpret data and use analytics platforms effectively. Regular workshops and hands-on sessions can make a significant difference.

- **Encourage Experimentation:** Create an environment where testing and experimentation are welcomed. Allow teams the flexibility to pilot new ideas, measure outcomes, and iterate rapidly.
- **Establish Clear Communication Channels:** Use collaborative tools such as [Team Chat](#) to ensure that insights and data findings are shared across departments in real time.
- **Integrate Analytics into Everyday Workflows:** Regularly use data insights from platforms like [DataCalculus](#) to guide strategic initiatives, whether during brainstorming, design refinement, or user testing phases.
- **Monitor Performance Metrics:** Leverage visual dashboards and reporting tools to continuously track the performance of design projects. The [Overall AI Report](#) is an excellent example of how visual data can be instrumental in performance tracking.

By embedding these principles into everyday operations, organizations can not only enhance their product design and development but also foster innovation that is responsive to market dynamics. As businesses continue to evolve, the interplay between creative design and data analytics will be a primary driver of competitive advantage.

Conclusion: The Future of Design Analysis

The future of design services is bright and data-driven. Design Analysts are uniquely positioned to propel organizations into a new era of product development where decisions are reinforced by data insights and strategic foresight. By leveraging platforms like DataCalculus and integrating advanced analytics into every phase of design, companies can achieve a level of innovation and precision that was previously unimaginable.

As we have seen, the landscape of design is being reshaped by the fusion of creative ingenuity and quantitative analysis. The role of a Design Analyst now extends beyond traditional boundaries, requiring a mastery of both aesthetics and analytics. Whether through iterative testing, advanced segmentation, or predictive forecasting, the journey toward product innovation is marked by a steadfast commitment to excellence and continual learning.

Organizations that embrace this integrated approach stand to benefit not only from enhanced product outcomes but also from a culture of innovation that permeates every department. By adopting best practices in data-driven design and leveraging the full spectrum of business intelligence tools—

including [Admin Tools](#) for effective team management—companies can ensure that their offerings continually set new benchmarks in quality and performance.

Ultimately, the integration of data analytics into product design and development is more than a trend—it is a fundamental shift that promises to redefine the competitive landscape of modern business. With the right tools and a proactive mindset, Design Analysts can turn data into a creative powerhouse, unlocking new opportunities and paving the way for a future rich with innovation.

For more information on transforming your raw data into insightful reports, explore the capabilities of [DataCalculus](#). Continue your journey by visiting specialized sections such as the [Data Dictionary](#), checking out the benefits of the [Overall AI Report](#) for best actionable insights, or exploring even deeper trends with the [Classification Report](#). Each of these offerings is designed to help you navigate the evolving landscape of design analytics, ensuring that your strategies are both innovative and firmly grounded in data.

By harnessing the power of data, Design Analysts are empowered to not only create outstanding products but also to lead their organizations into a future where data drives discovery, innovation, and success. As companies continue to evolve in this exciting era, embracing a data-driven design philosophy will be the crucial differentiator between those that thrive and those that merely survive.

To learn more about managing your datasets and optimizing your design strategies at scale, visit our [Dataset Operations](#) section. For comprehensive support on your data journey, our [Support AI](#) is always ready to provide expert guidance and personalized assistance.

This synergy between analytical precision and creative excellence is redefining the future of design. Whether it's through advanced analytics, iterative design methods, or the seamless integration of team communication via [Team Chat](#), the impact of data-driven design is unmistakable and profound. Embrace the revolution in product design and development by anchoring your strategies in data—a move that promises robust innovation and unprecedented success in the market.

Explore the world of data-driven design and join the forefront of innovation. With a commitment to continuous improvement and a strategic emphasis on integrating analytics into every facet of design, the future is bright for

companies ready to leverage the full capabilities of platforms like DataCalculus. Revolutionize your product development process, craft innovative solutions, and set new industry standards by making data your most valuable design tool.

Begin your journey today by exploring more on our website and tapping into the transformative power of data analytics. The era of data-driven design is here. Embrace it, and be prepared to lead, innovate, and inspire.

Design Analyst: Brand Identity Design Insights Leveraging Data Analytics for Effective Brand Identity Design

In today's hyper-competitive market, the role of a **Design Analyst** has evolved significantly. With a focus on transforming raw data into actionable insights, design analysts are at the forefront of combining creativity with analytical rigor. This article explores how integrating **Data Analytics** and **Business Intelligence** can elevate **Brand Identity Design** strategies while ensuring that every design decision is backed by critical data. As companies continue to rely on metrics and performance indicators, the fusion of creativity and data leads to enhanced **Design Services** outcomes.

Introduction to Brand Identity Design and Data Analytics

Brand identity is not merely about logos, color palettes, or typography; it is the very essence of a company's presence in the marketplace. In our data-driven era, a robust brand identity is built on insights gleaned from extensive research, user behavior analytics, and market trends. For a **Design Analyst**, the process involves critical examinations of data to guide creative decisions. This leads to designs that are both aesthetically compelling and strategically sound.

The integration of **Business Intelligence** into design practices means that decisions are made based on comprehensive research and data models. These models help identify the most effective strategies to cultivate a brand's presence. By employing a systematic approach, a design analyst can leverage actionable insights to drive innovative design solutions in **Brand Identity Design**.

The Evolving Role of the Design Analyst

The modern **Design Analyst** wears many hats. They are intermediaries between creativity and analytics, responsible for transforming qualitative design concepts into quantifiable outcomes.

The traditional notion of design as a purely aesthetic domain is slowly giving way to a field where analytics and creative processes merge. Today’s design professionals analyze customer data, market trends, and competitive positioning to form comprehensive strategies that underpin every aspect of design work.

Being adept at **Data Analytics**, design analysts utilize a range of methodologies—from data mining to predictive analytics—to chart the course for successful brand identities. By understanding user interactions and measuring performance metrics, they can diagnose the effectiveness of a design before its full-scale launch. For example, utilizing platforms like [DataCalculus](#) can help convert raw data into clear, insightful reports, providing a robust basis for design recommendations.

Integrating Business Intelligence into Brand Identity

Building a powerful brand identity design starts with a clear understanding of the target audience and the competitive landscape. **Business Intelligence** provides a rich context in which design decisions are carefully aligned with business objectives. In this context, data is not just numbers—it reveals trends, user preferences, and evolving market narratives.

One of the first steps for a **Design Analyst** is to conduct audience research, including psychodemographic segmentation and consumer behavior analysis. These insights help determine which visual elements resonate best with the audience. For instance, a detailed exploration of the dataset can be achieved via the [Data Dictionary](#) tool, which provides clarity on each column’s type and statistics. By thoroughly exploring internal data, design analysts can build a detailed understanding of user engagement and brand perception.

The Process of Brand Identity Design

A comprehensive brand identity design process often involves several key phases:

- **Research and Discovery:** Gathering data on market trends, consumer behavior, and competitor analysis. During this phase, evaluative methods such as customer surveys and digital analytics are vital.
- **Concept Development:** Using the insights gathered to formulate design concepts that accurately reflect the brand’s ethos. Here, creative ideation meets data-driven insights.
- **Design Execution:** Translating creative concepts into tangible design elements with careful consideration of typography, color schemes, and visual hierarchy.
- **Testing and Feedback:** Implementing prototypes and gathering performance data through feedback loops to iterate and refine the design.
- **Implementation and Launch:** Finalizing the design for its rollout, ensuring consistency across all channels and that every design element reinforces the brand identity.

Throughout these stages, the iterative process is deepened by a focus on measurable outcomes. Tools such as the [Overall AI Report](#) and [Pattern Report](#) assist in identifying key performance

indicators during testing phases. These insights help design analysts refine their strategies to optimize overall brand impact.

Data-Driven Design: Methods That Work

The approach of **Data-driven Design** reflects an evolution from intuition-based methods to those that rely on comprehensive analytics. In order to make informed decisions, a design analyst must be comfortable with dissecting large volumes of data. The systematic analysis of digital footprint, social media trends, and website analytics is central to this process.

Consider the role of visual analytics within the realm of **Design Services**. By harnessing the power of interrelated data sets, design analysts can pinpoint which design elements evoke the most positive responses from users. This enables analysts to fine-tune visual components that contribute to stronger brand recall and digital engagement.

For example, by leveraging insights from the [Classification Report](#) tool, one can identify key factors influencing user perceptions. Similarly, the [Clustering Report](#) groups similar data points to reveal hidden trends in user behavior. These analytical approaches ensure that design decisions are not made in isolation but are supported by comprehensive data models.

Best Practices for Design Analysts in Brand Identity

Success in **Brand Identity Design** for a Design Analyst depends on a combination of creative acumen and data analytics expertise. Below are some best practices recommended for professionals in this domain:

1. **Embrace a Structured Approach:** Begin with thorough data collection through surveys, social media analytics, and customer feedback. Use structured methods to segment your audience, ensuring that each design decision is targeted and effective.
2. **Leverage Business Intelligence:** Integrate **Business Intelligence** tools to gather cross-functional insights. In doing so, you can understand market trends and predict consumer behavior. Generating automated reports via [DataCalculus](#) can streamline this process.
3. **Maintain Consistent Brand Messaging:** Ensure that data-driven insights are always aligned with the broader brand strategy. Consistency is key in reinforcing brand identity both internally and externally.
4. **Iterate Continuously:** The design process should be iterative, with continuous refinements based on ongoing analysis. Tools such as the [Data Scientist AI](#) enable you to ask targeted questions regarding dataset trends and anomalies.
5. **Monitor and Evaluate:** Post-launch, it's crucial to track performance metrics and evaluate the impact of design elements. Platforms like [Support AI](#) provide personalized assistance, ensuring that any emerging issues are addressed swiftly.

Data and Analytics: Transforming Creative Processes

One of the most significant advantages of modern **Data Analytics** is its ability to transform creative processes. By applying analytical models, a Design Analyst can quantify design performance in ways that were once deemed intangible. This represents a major shift in the field of **Design Services**, where creativity and analytics now operate in tandem.

For instance, a comprehensive review of customer interactions on digital platforms can reveal which aspects of a brand's identity resonate most effectively. Utilizing tools like [Dataset Operations](#), analysts can efficiently manage datasets to extract meaningful patterns and trends. This integration of data science into design thinking has opened new avenues for innovation, allowing brands to remain agile in a dynamic market environment.

Implementing a Data-Driven Culture Within Design Teams

The shift towards a **data-driven design** culture within creative teams is essential for fostering innovative and resilient brand identities. For a Design Analyst, it means creating an environment where data insights inform every step of the design process.

Collaboration is key. Modern design teams benefit from robust communication platforms such as the [Team Chat](#) system, which facilitates real-time discussions and decision-making. Furthermore, administrative tools like [Admin Tools](#) help organize team roles, manage responsibilities, and streamline workflow management, ensuring that every member is aligned with the overarching strategy informed by business intelligence.

An added advantage of this approach is the ability to conduct **Bulk Operations** on large sets of data. When design changes need to be implemented at scale, using [Bulk Operations](#) can save time and ensure consistency across multiple projects.

Real-World Applications and Success Stories

The application of **Data Analytics** within **Brand Identity Design** has already transformed businesses across various industries. Case studies have shown that when analytical insights are effectively integrated, companies enjoy improved customer engagement, increased brand loyalty, and overall higher ROI.

Consider a scenario where an organization is rebranding. A design analyst might begin by utilizing tools such as [Report Assembly](#) to gather comprehensive insights on existing brand perceptions. Based on the data, refined design strategies are then tested through iterative prototypes. The resulting changes do not merely alter the aesthetic appeal but resonate deeply with the target audience, ultimately achieving the desired brand positioning.

Moreover, organizations that have embraced advanced analytical methods report significant improvements in operational efficiency. The ability to predict market trends and consumer behavior allows these companies to stay ahead of competitors. This is especially important in the fast-paced sectors of **Design Services** and marketing, where trends can shift rapidly. By

continuously refining **Brand Identity Design** strategies, companies optimize both creative output and financial performance.

Overcoming Challenges in Data-Driven Design

No process is without its challenges. For a Design Analyst, balancing creativity and data analytics is an ongoing balancing act. Some common challenges include:

- **Data Overload:** With abundant information available, filtering out irrelevant data to focus on actionable insights is crucial. This is where the importance of utilizing specific, targeted tools like [DataCalculus](#) becomes evident.
- **Interpretation Complexity:** Raw data alone cannot paint the complete picture. The challenge lies in interpreting analytics in a way that meaningfully informs design decisions. Consultants must transform numbers into narratives that are relevant to the brand's creative vision.
- **Integration with Creative Workflows:** Merging analytical processes with creative brainstorming sessions can sometimes create friction. Establishing clear protocols and communication channels helps mitigate these issues.
- **Keeping Up with Rapid Trends:** Data models must be updated frequently to reflect market dynamics. Continuous learning and system upgrades are necessary to stay agile. Tools like the [Data Scientist AI](#) can assist in asking the right questions in real time.

By proactively addressing these challenges, organizations not only refine the design process but also create a more resilient strategy for adapting to market changes.

How to Measure the Impact of Brand Identity Design

Having executed a data-driven strategy, measuring the impact of **Brand Identity Design** is the final, yet critical step. Metrics such as customer engagement, social media interactions, conversion rates, and overall brand sentiment provide clear indicators of success. With detailed reports from platforms like [Overall AI Report](#) and [Pattern Report](#), businesses can draw correlations between design changes and market performance.

Furthermore, continuous monitoring helps in making iterative adjustments. An effective **Design Analyst** must be open to evolving strategies based on the data collected, ensuring that the design not only meets aesthetic criteria but also drives business growth. By measuring **Brand Optimization** metrics, companies can fine-tune every aspect of their visual identity.

The Future of Design Analysis in Brand Identity

Looking ahead, the role of the **Design Analyst** is poised to become even more integral to the success of **Design Services** in the era of digital transformation. As artificial intelligence and machine learning continue to evolve, these technologies will further democratize the access to complex analytics, making it even easier to translate data into creativity.

Innovative solutions, such as those offered through [Data Dictionary](#) and [Dataset Operations](#), help reveal deeper insights at a granular level, pushing the envelope of what is possible in **Analytical Design**. Companies are increasingly looking towards integrated solutions that allow for rapid decision-making and agile response strategies to dynamic market conditions.

This convergence of technology and creativity will further blur the lines between the analytic and the artistic. In such a future, the **Design Analyst** will not only be a gatekeeper of visual trends but also a strategic advisor who drives the enterprise's **Brand Strategy** and overall market positioning.

Conclusion: Data-Driven Insights Propel Brand Innovation

In summary, the modern world of **Brand Identity Design** relies on much more than just creative flair. By embracing **Business Intelligence** and **Data Analytics**, a **Design Analyst** transforms the way brands communicate with their audiences. From initial research and iterative design processes to the final analysis of performance metrics, every step is enriched by data-driven insights.

The integration of insights from platforms such as [DataCalculus](#), [Classification Report](#), and [Clustering Report](#) not only facilitates smarter decision-making but also fosters innovation in design execution. As you drive forward in the world of **Design Services**, remember that effective brand identity is built on the foundations of **Creative Analysis** and analytical integrity.

Furthermore, the continuous evolution of data-driven methodologies promises to keep pushing the boundaries of what design can achieve. The challenges associated with **Data-driven Design** are surmountable by adopting best practices and leveraging the right analytics tools. By doing so, businesses can harness the full potential of **Business Intelligence** to not only anticipate customer needs but also to create lasting impressions through innovative brand experiences.

For those ready to explore the full capabilities of data-informed design workflows, consider extending your analysis with additional resources such as the [Bulk Operations](#) module to manage projects at scale, and the [Report Assembly](#) tool to compile insights into digestible, actionable reports.

Ultimately, the future of **Brand Identity Design** lies in the seamless integration of creative energy with detailed analytics—a future where every design decision is as informed as it is inspired. Adopting these practices will help your organization uncover hidden trends, optimize your strategies, and forge lasting connections with your audience.

As companies continue to navigate the intersection of design and data, the role of the Design Analyst becomes indispensable. Embrace the journey of learning and innovation, and let data lead you to new heights of creativity and market success. With the right balance between art and science, your brand can stand out in even the most saturated markets.

Additional Resources and Internal Links

For further exploration of analytical approaches and cutting-edge insights that complement your design strategy, consider reviewing the following resources:

- [DataCalculus](#) – Transform your raw data into insightful reports with just one click.
- [Data Dictionary](#) – Explore your dataset and each column's type and statistics effortlessly.
- [Overall AI Report](#) – Take your next best actions based on the top discovered insights confidently.
- [Pattern Report](#) – Uncover hidden trends and patterns in your data for smarter decision-making.
- [Classification Report](#) – Identify key factors influencing your future outcomes and optimize strategies.
- [Clustering Report](#) – Group similar data points to reveal insights and enhance operational efficiency.
- [Data Scientist AI](#) – Ask questions to an AI agent that is specially trained for each of your datasets.
- [Support AI](#) – Get help from an AI that knows you and your company to receive personalized support.
- [Dataset Operations](#) – Upload, export, share, unfollow, transfer, or delete your datasets smoothly.
- [Bulk Operations](#) – Perform all the above actions at scale on multiple datasets at once.
- [Report Assembly](#) – Pull together insightful reports and share them with your team while controlling the access.
- [Team Chat](#) – Collaborate on datasets, reports, decisions, and strategies by chatting with your team on any topic.
- [Admin Tools](#) – Add or remove team members, assign roles, and manage payments and plans via a dashboard.

Each of these resources can provide additional insights that support the transformative power of data within your brand identity initiatives. By integrating these tools and practices, you position your organization to remain agile, responsive, and innovative in the face of evolving market dynamics.

Final Thoughts

The journey of merging creative artistry with rigorous analytics is an evolving narrative. As a **Design Analyst**, your ability to harness the power of data and to transform it into actionable insights drives not only the visual evolution of a brand but also its strategic success. Through continuous learning, collaboration, and the strategic use of **Business Intelligence** tools, you pave the way for designs that are both meaningful and impactful.

In conclusion, the dynamic field of **Brand Identity Design** is being reshaped by the convergence of creative vision and data-driven analysis. By adopting best practices, leveraging state-of-the-art methodologies, and continuously iterating on design strategies, you can build brands that not only look compelling but also perform remarkably well in their respective markets. The future of design is here, and it is powered by data!

Embrace the balance between aesthetics and analytics. Let your next design project be a testament to the transformative impact of a well-informed strategy, and watch as your brand story unfolds with a newfound clarity and purpose.

Welcome to the new era of design—where every pixel, line, and curve is strategically aligned with your data-driven vision.

Forecasting: theory and practice

<https://doi.org/10.1016/j.ijforecast.2021.11.001> Get rights and content

Abstract

Forecasting has always been at the forefront of decision making and planning. The uncertainty that surrounds the future is both exciting and challenging, with individuals and organisations seeking to minimise risks and maximise utilities. The large number of forecasting applications calls for a diverse set of forecasting methods to tackle real-life challenges. This article provides a non-systematic review of the theory and the practice of forecasting. We provide an overview of a wide range of theoretical, state-of-the-art models, methods, principles, and approaches to prepare, produce, organise, and evaluate forecasts. We then demonstrate how such theoretical concepts are applied in a variety of real-life contexts.

We do not claim that this review is an exhaustive list of methods and applications. However, we wish that our encyclopedic presentation will offer a point of reference for the rich work that has been undertaken over the last decades, with some key insights for the future of forecasting theory and practice. Given its encyclopedic nature, the intended mode of reading is non-linear. We offer cross-references to allow the readers to navigate through the various topics. We complement the theoretical concepts and applications covered by large lists of free or open-source software implementations and publicly-available databases.

Prediction

“In theory, there is no difference between theory and practice. But, in practice, there is.”

Benjamin Brewster (1882)

1. Introduction¹

Forecasting has come a long way since early humans looked at the sky to see if the weather would be suitable for hunting, and even since hunters could get a forecast such as “a high of 40 with a chance of rain”. Now a hunter can look at a smartphone to instantly get hour-by-hour forecasts of temperatures and probabilities of rain at multiple locations as well as videos of maps showing forecasted weather patterns over

the coming hours. Tailored forecasts of increasing sophistication can be generated to inform important decisions of many different types by managers, public officials, investors, and other decision makers.

In the 15 years since the excellent review paper by [De Gooijer and Hyndman \(2006\)](#), the field of forecasting has seen amazing growth in both theory and practice. Thus, this review is both timely and broad, ranging from the highly theoretical to the very applied. Rapid advances in computing have enabled the analysis of larger and more complex data sets and stimulated interest in analytics and data science. As a result, the forecaster's toolbox of methods has grown in size and sophistication. Computer science has led the way with methods such as neural networks and other types of machine learning, which are getting a great deal of attention from forecasters and decision makers. Other methods, including statistical methods such as Bayesian forecasting and complex regression models, have also benefited from advances in computing. And improvements have not been limited to those based on computing advances. For example, the literature on judgmental forecasting has expanded considerably, driven largely by the "wisdom of crowds" notion.

The combining, or aggregation, of forecasts, which is not a new idea, has received increased attention in the forecasting community recently and has been shown to perform well. For example, the top-performing entries in the M4 Competition run by Spyros Makridakis combined forecasts from multiple methods. Many models have been developed to forecast the number of deaths that will be caused by COVID-19, and combining the forecasts makes sense because it is hard to know which one will be the most accurate. It is consistent with Bayesian ideas since it can be viewed as updating, with each individual forecast added to the combined forecast (also called an ensemble) contributing some new information.

Despite the excitement surrounding these new developments, older methods such as ARIMA and exponential smoothing are still valuable too. Exponential smoothing, along with other simple approaches, are quite robust and not as prone to overfitting as more complex methods. In that sense, they are useful not only on their own merits, but as part of an ensemble that also includes more sophisticated methods. Combined forecasts are more valuable if the forecasts come from methods that are diverse so that their forecast errors are not highly correlated.

The conditions leading to larger, more sophisticated toolboxes for forecasters have also led to larger data sets with denser grids and improved models in areas of application. This has happened with models of the atmosphere, which are important in formulating improved weather forecasts. More detailed information about customers and their preferences allows the development of improved models of customer behaviour for managers. In turn, forecasting methods that can handle all of that information quickly are valuable for decision-making purposes. This process has spurred an explosion in trying to gather information on the internet.

Risk is an important consideration in decision making, and probability forecasts can quantify such risks. Theoretical work in probability forecasting has been active for some time, and decision makers in many areas of practice have embraced the use of probability forecasts. In the Bayesian approach, inferences and forecasts are probabilistic in nature, and probability forecasts can be generated in many other ways too.

The U.S. National Weather Service began issuing probabilities of precipitation to the public in the 1960s. Yet extensive widespread use and dissemination of probabilities has only developed since the turn of the century. Now probability forecasts are increasingly communicated to the public and used as inputs in decision making. Nate Silver's [FiveThirtyEight.com](https://www.fivethirtyeight.com) report gives probability forecasts for elections, medicine and science, sporting events, economic measures, and many other areas, often looking at multiple forecasting models individually and also combining them.

It is natural for people to desire certainty. When probability forecasts of precipitation were first disseminated widely, many were very sceptical about them, with some accusing the forecasters of hedging and saying "Don't give me a probability. I want to know if it's going to rain or not". Of course, point forecasts often are given along with probability forecasts. The current frequent exposure to probabilities helps the general public better understand, appreciate, and feel more comfortable with them. And the current situation in the world with COVID-19, increases in huge fires, big storms, political polarisation, international conflicts, etc., should help them realise that we are living in an age with huge uncertainties, and forecasts that quantify these uncertainties can be important. Where possible, visualisation can help, as indicated by the saying that a picture is worth a thousand words. Examples are the cones of uncertainty on maps in

forecasts of the speed, severity, and future path of hurricanes, and the time line of the probability of a team winning a game, updated quickly after each play.

Put simply, this is an exciting time for the field of forecasting with all of the new theoretical developments and forecasting applications in practice. Forecasting is so ubiquitous that it's not possible to cover all of these developments in a single article. This article manages to cover quite a few, and a good variety. Using short presentations for each one from an expert "close to the ground" on that theoretical topic or field of practice works well to provide a picture of the current state of the art in forecasting theory and practice.

2. Theory

2.1. Introduction to forecasting theory²

The theory of forecasting is based on the premise that current and past knowledge can be used to make predictions about the future. In particular for time series, there is the belief that it is possible to identify patterns in the historical values and successfully implement them in the process of predicting future values. However, the exact prediction of futures values is not expected. Instead, among the many options for a forecast of a single time series at a future time period are an expected value (known as a point forecast), a prediction interval, a percentile and an entire prediction distribution. This set of results collectively could be considered to be "the forecast". There are numerous other potential outcomes of a forecasting process. The objective may be to forecast an event, such as equipment failure, and time series may play only a small role in the forecasting process. Forecasting procedures are best when they relate to a problem to be solved in practice. The theory can then be developed by understanding the essential features of the problem. In turn, the theoretical results can lead to improved practice.

In this introduction, it is assumed that forecasting theories are developed as forecasting methods and models. A forecasting method is defined here to be a predetermined sequence of steps that produces forecasts at future time periods. Many forecasting methods, but definitely not all, have corresponding stochastic models that produce the same point forecasts. A stochastic model provides a data generating process that can be used to produce prediction intervals and entire prediction distributions in addition to

point forecasts. Every stochastic model makes assumptions about the process and the associated probability distributions. Even when a forecasting method has an underlying stochastic model, the model is not necessarily unique. For example, the simple exponential smoothing method has multiple stochastic models, including state space models that may or may not be homoscedastic (i.e., possess constant variance). The combining of forecasts from different methods has been shown to be a very successful forecasting method. The combination of the corresponding stochastic models, if they exist, is itself a model. Forecasts can be produced by a process that incorporates new and/or existing forecasting methods/models. Of course, these more complex processes would also be forecasting methods/models.

Consideration of the nature of the variables and their involvement in the forecasting process is essential. In univariate forecasting, the forecasts are developed for a single time series by using the information from the historical values of the time series itself. While in multivariate forecasting, other time series variables are involved in producing the forecasts, as in time series regression. Both univariate and multivariate forecasting may allow for interventions (e.g., special promotions, extreme weather). Relationships among variables and other types of input could be linear or involve nonlinear structures (e.g., market penetration of a new technology). When an explicit functional form is not available, methodologies such as simulation or artificial neural networks might be employed. Theories from fields, such as economics, epidemiology, and meteorology, can be an important part of developing these relationships. Multivariate forecasting could also mean forecasting multiple variables simultaneously (e.g., econometric models). The data or observed values for time series come in many different forms that may limit or determine the choice of a forecasting method. In fact, there may be no historical observations at all for the item of interest, when judgmental methods must be used (e.g., time taken to complete construction of a new airport). The nature of the data may well require the development of a new forecasting method. The frequency of observations can include all sorts of variations, such as every minute, hourly, weekly, monthly, and yearly (e.g., the electricity industry needs to forecast demand loads at hourly intervals as well as long term demand for ten or more years ahead). The data could be composed of everything from a single important time series to billions of time series. Economic analysis often includes multiple variables, many of which affect one another. Time series for businesses are likely to be important at many different levels

(e.g., stock keeping unit, common ingredients, or common size container) and, consequently, form a hierarchy of time series. Some or many of the values might be zero; making the time series intermittent. The list of forms for data is almost endless. Prior to applying a forecasting method, the data may require pre-processing. There are basic details, such as checking for accuracy and missing values. Other matters might precede the application of the forecasting method or be incorporated into the methods/models themselves. The treatment of seasonality is such a case. Some forecasting method/models require de-seasonalised time series, while others address seasonality within the methods/models. Making it less clear when seasonality is considered relative to a forecasting method/model, some governmental statistical agencies produce forecasts to extend time series into the future in the midst of estimating seasonal factors (i.e., X-12 ARIMA).

Finally, it is extremely important to evaluate the effectiveness of a forecasting method. The ultimate application of the forecasts provides guidance in how to measure their accuracy. The focus is frequently on the difference between the actual value and a point forecast for the value. Many loss functions have been proposed to capture the “average” of these differences. Prediction intervals and percentiles can be used to judge the value of a point forecast as part of the forecast. On the other hand, the quality of prediction intervals and prediction distributions can themselves be evaluated by procedures and formulas that have been developed (e.g., ones based on scoring rules). Another assessment tool is judging the forecasts by metrics relevant to their usage (e.g., total costs or service levels).

In the remaining subsections of Section 2, forecasting theory encompasses both stochastic modelling and forecasting methods along with related aspects.

2.2. Pre-processing data

2.2.1. Box–Cox transformations³

A common practice in forecasting models is to transform the variable of interest using the transformation initially proposed by [Box and Cox \(1964\)](#) as

The range of the transformation will be restricted in a way that depends on the sign of , therefore [Bickel and Doksum \(1981\)](#) propose the following modification which has a

range from $\lambda = -1$ for any value of λ . For a recent review of the Box–Cox (and other similar) transformations see [Atkinson, Riani, and Corbellini \(2021\)](#).

The initial motivation for the Box–Cox transformation was to ensure data conformed to assumptions of normality and constant error variance that are required for inference in many statistical models. The transformation nests the log transformation when $\lambda = 1$ and the case of no transformation (up to an additive constant) when $\lambda = 0$. Additive models for $\lambda < 0$ correspond to multiplicative models on the original scale of Y . Choices of λ between -1 and 1 therefore provide a natural continuum between multiplicative and additive models. For examples of forecasting models that use either a log or Box–Cox transformation see Sections [2.3.5 Forecasting for multiple seasonal cycles](#), [2.3.6 State-space models](#) and for applications see Sections [3.2.5 Promotional forecasting](#), [3.6.2 Epidemics and pandemics](#), [3.8.4 Call arrival forecasting](#).

The literature on choosing λ is extensive and dates back to the original [Box and Cox \(1964\)](#) paper - for a review see [Sakia \(1992\)](#). In a forecasting context, a popular method for finding λ is given by [Guerrero \(1993\)](#). The method splits the data into blocks, computes the coefficient of variation within each block and then computes the coefficient of variation again between these blocks. The λ that minimises this quantity is chosen. Since the transformations considered here are monotonic, the forecast quantiles of the transformed data will, when back-transformed, result in the correct forecast quantiles in terms of the original data. As a result finding prediction intervals in terms of the original data only requires inverting the transformation. It should be noted though, that prediction intervals that are symmetric in terms of the transformed data will not be symmetric in terms of the original data. In a similar vein, back-transformation of the forecast median of the transformed data returns the forecast median in terms of the original data. For more on using the median forecast see Section [2.12.2](#) and references therein.

The convenient properties that apply to forecast quantiles, do not apply to the forecast mean, something recognised at least since the work of [Granger and Newbold \(1976\)](#). Back-transformation of the forecast mean of the transformed data does not yield the forecast mean of the original data, due to the non-linearity of the transformation. Consequently forecasts on the original scale of the data will be biased unless a correction is used. For some examples of bias correction methods see [Granger and Newbold](#),

1976, Guerrero, 1993, Pankratz and Dudley, 1987 and Taylor (1986b) and references therein.

The issues of choosing and bias correcting are accounted for in popular forecasting software packages. Notably, the method of Guerrero (1993) both for finding and bias correcting is implemented in the R packages *forecast* and *fable* (see Appendix B).

2.2.2. Time series decomposition⁴

Time series decomposition is an important building block for various forecasting approaches (see, for example, Sections 2.3.3 [Theta method and models](#), 2.7.6 [Bagging for time series forecasting](#), 3.8.3 [Traffic flow forecasting](#)) and a crucial tools for statistical agencies. Seasonal decomposition is a way to present a time series as a function of other time series, called components. Commonly used decompositions are additive and multiplicative, where such functions are summation and multiplication correspondingly. If logs can be applied to time series, any additive decomposition method can serve as multiplicative after applying log transformation to the data.

The simplest additive decomposition of a time series with single seasonality comprises three components: trend, seasonal component, and the “remainder”. It is assumed that the seasonal component has a repeating pattern (thus sub-series corresponding to every season are smooth or even constant), the trend component describes the smooth underlying mean and the remainder component is small and contains noise.

The first attempt to decompose time series into trend and seasonality is dated to 1847 when Buys-Ballot (1847) performed decomposition between trend and seasonality, modelling the trend by a polynomial and the seasonality by dummy variables. Then, in 1884 Poynting (1884) proposed price averaging as a tool for eliminating trend and seasonal fluctuations. Later, his approach was extended by Anderson and Nochmals, 1914, Hooker, 1901 and Spencer (1904). Copeland (1915) was the first who attempted to extract the seasonal component, and Macaulay (1931) proposed a method which is currently considered “classical”.

The main idea of this method comes from the observation that averaging a time series with window size of the time series seasonal period leaves the trend almost intact, while effectively removes seasonal and random components. At the next step, subtracting the estimated trend from the data and averaging the result for every season gives the seasonal component. The rest becomes the remainder.

Classical decomposition led to a series of more complex decomposition methods such as X-11 (Shishkin, Young, & Musgrave, 1967), X-11-ARIMA (Dagum, 1988, Ladiray and Quenneville, 2001), X-12-ARIMA (Findley, Monsell, Bell, Otto, & Chen, 1998), and X-13-ARIMA-SEATS (Findley, 2005); see also Section 2.3.4.

Seasonal trend decomposition using Loess (STL: Cleveland, Cleveland, McRae, & Terpenning, 1990) takes iterative approach and uses smoothing to obtain a better estimate of the trend and seasonal component at every iteration. Thus, starting with an estimate of the trend component, the trend component is subtracted from the data, the result is smoothed along sub-series corresponding to every season to obtain a “rough” estimate of the seasonal component. Since it might contain some trend, it is averaged to extract this remaining trend, which is then subtracted to get a detrended seasonal component. This detrended seasonal component is subtracted from the data and the result is smoothed again to obtain a better estimate of the trend. This cycle repeats a certain number of times.

Another big set of methods use a single underlining statistical model to perform decomposition. The model allows computation of confidence and prediction intervals naturally, which is not common for iterative and methods involving multiple models. The list of such methods includes TRAMO/SEATS procedure (Monsell, Aston, & Koopman, 2003), the BATS and TBATS models (De Livera, Hyndman, & Snyder, 2011), various structural time series model approaches (Commandeur et al., 2011, Harvey, 1990), and the recently developed seasonal-trend decomposition based on regression (STR: Dokumentov, 2017, Dokumentov and Hyndman, 2018); see also Section 2.3.2. The last mentioned is one of the most generic decomposition methods allowing presence of missing values and outliers, multiple seasonal and cyclic components, exogenous variables with constant, varying, seasonal or cyclic influences, arbitrary complex seasonal schedules. By extending time series with a sequence of missing values the method allows forecasting.

2.2.3. Anomaly detection and time series forecasting⁵

Temporal data are often subject to uncontrolled, unexpected interventions, from which various types of anomalous observations are produced. Owing to the complex nature of domain specific problems, it is difficult to find a unified definition for an anomaly and mostly application-specific (Unwin, 2019). In time series and forecasting literature, an anomaly is mostly defined with respect to a specific context or its relation to past

behaviours. The idea of a context is induced by the structure of the input data and the problem formulation (Chandola et al., 2007, Chandola et al., 2009, Hand, 2009). Further, anomaly detection in forecasting literature has two main focuses, which are conflicting in nature: one demands special attention be paid to anomalies as they can be the main carriers of significant and often critical information such as fraud activities, disease outbreak, natural disasters, while the other down-grades the value of anomalies as it reflects data quality issues such as missing values, corrupted data, data entry errors, extremes, duplicates and unreliable values (Talagala, Hyndman, & Smith-Miles, 2020).

In the time series forecasting context, anomaly detection problems can be identified under three major umbrella themes: detection of (i) contextual anomalies (point anomalies, additive anomalies) within a given series, (ii) anomalous sub-sequences within a given series, and (iii) anomalous series within a collection of series (Gupta et al., 2013, Talagala, Hyndman, Smith-Miles and Sevvandi et al., 2020). According to previous studies forecast intervals are quite sensitive to contextual anomalies and the greatest impact on forecast are from anomalies occurring at the forecast origin (Chen & Liu, 1993a).

The anomaly detection methods in forecasting applications can be categorised into two groups: (i) model-based approaches and (ii) feature-based approaches. Model-based approaches compare the predicted values with the original data. If the deviations are beyond a certain threshold, the corresponding observations are treated as anomalies (Luo et al., 2018a, Luo, Hong, Yue, 2018, Sobhani et al., 2020). Contextual anomalies and anomalous sub-sequences are vastly covered by model-based approaches. Limitations in the detectability of anomalous events depend on the input effects of external time series. Examples of such effects are included in SARIMAX models for polynomial approaches (see also Section 2.3.4). In nonlinear contexts an example is the generalised Bass model (Bass, Krishnan, & Jain, 1994) for special life cycle time series with external control processes (see Section 2.3.18). SARMAX with nonlinear perturbed mean trajectory as input variable may help separating the mean process under control effects from anomalies in the residual process. Feature-based approaches, on the other hand, do not rely on predictive models. Instead, they are based on the time series features measured using different statistical operations (see Section 2.7.4) that differentiate anomalous instances from typical behaviours (Fulcher &

Jones, 2014). Feature-based approaches are commonly used for detecting anomalous time series within a large collection of time series. Under this approach, it first forecasts an anomalous threshold for the systems typical behaviour and new observations are identified as anomalies when they fall outside the bounds of the established anomalous threshold (Talagala et al., 2019, Talagala, Hyndman, Smith-Miles and Sevvandi et al., 2020). Most of the existing algorithms involve a manual anomalous threshold. In contrast, Burridge and Robert Taylor (2006) and Talagala, Hyndman, Smith-Miles and Sevvandi et al. (2020) use extreme value theory based data-driven anomalous thresholds. Approaches to the problem of anomaly detection for temporal data can also be divided into two main scenarios: (i) batch processing and (ii) data streams. The data stream scenario poses many additional challenges, due to nonstationarity, large volume, high velocity, noisy signals, incomplete events and online support (Luo, Hong, Yue, 2018, Talagala, Hyndman, Smith-Miles and Sevvandi et al., 2020).

The performance evaluation of the anomaly detection frameworks is typically done using confusion matrices (Luo, Hong, Yue, 2018, Sobhani et al., 2020). However, these measures are not enough to evaluate the performance of the classifiers in the presence of imbalanced data (Hossin & Sulaiman, 2015). Following Ranawana and Palade (2006) and Talagala et al. (2019), Leigh et al. (2019) have used some additional measures such as negative predictive value, positive predictive value and optimised precision to evaluate the performance of their detection algorithms.

2.2.4. Robust handling of outliers in time series forecasting⁶

Estimators of time series processes can be dramatically affected by the presence of few aberrant observations which are called differently in the time series literature: outliers, spikes, jumps, extreme observations (see Section 2.2.3). If their presence is neglected, coefficients could be biasedly estimated. Biased estimates of ARIMA processes will decrease the efficiency of predictions (Bianco, García Ben, Martínez, & Yohai, 2001). Moreover, as the optimal predictor of ARIMA models (see Section 2.3.4) is a linear combination of observed units, the largest coefficients correspond to observations near the forecast origin and the presence of outliers among these units can severely affect the forecasts. Proper preliminary analysis of possible extreme observations is an unavoidable step, which should be carried out before any time series modelling and forecasting exercise (see Section 2.3.9). The issue was first raised in the seminal paper by Fox (1972), who suggests a classification of outliers in time series, separating additive

outliers (AO) from innovation outliers (IO). The influence of different types of outliers on the prediction errors in conditional mean models (ARIMA models) is studied by [Chen and Liu, 1993a](#), [Chen and Liu, 1993b](#) and [Ledolter, 1989](#), [Ledolter, 1991](#), while the GARCH context (see also Section 2.3.11) is explored by [Catalán and Trávez \(2007\)](#) and [Franses and Ghijssels \(1999\)](#). [Abraham and Box \(1979\)](#) propose a Bayesian model which reflects the presence of outliers in time series and allows to mitigate their effects on estimated parameters and, consequently, improve the prediction ability. The main idea is to use a probabilistic framework allowing for the presence of a small group of discrepant units.

A procedure for the correct specification of models, accounting for the presence of outliers, is introduced by [Tsay \(1986\)](#) relying on iterative identification-detection-removal of cycles in the observed time series contaminated by outliers. The same issue is tackled by [Abraham and Chuang \(1989\)](#): in this work non-influential outliers are separated from influential outliers which are observations with high residuals affecting parameter estimation. Tsay's procedure has been later modified ([Balke, 1993](#)) to effectively detect time series level shifts. The impulse- and step-indicator saturation approach is used by [Marczak and Proietti \(2016\)](#) for detecting additive outliers and level shifts estimating structural models in the framework of nonstationary seasonal series. They find that timely detection of level shifts located towards the end of the series can improve the prediction accuracy.

All these works are important because outlier and influential observations detection is crucial for improving the forecasting performance of models. The robust estimation of model parameters is another way to improve predictive accuracy without correcting or removing outliers (see Section 3.4.2, for the application on energy data). [Sakata and White \(1998\)](#) introduce a new two-stage estimation strategy for the conditional variance based on Hampel estimators and S-estimators. [Park \(2002\)](#) proposes a robust GARCH model, called RGARCH exploiting the idea of least absolute deviation estimation. The robust approach is also followed for conditional mean models by [Gelper, Fried, and Croux \(2009\)](#) who introduce a robust version of the exponential and Holt-Winters smoothing technique for prediction purposes and by [Cheng and Yang \(2015\)](#) who propose an outlier resistant algorithm developed starting from a new synthetic loss function. Very recently, [Beyaztas and Shang \(2019\)](#) have introduced a robust forecasting

procedure based on weighted likelihood estimators to improve point and interval forecasts in functional time series contaminated by the presence of outliers.

2.2.5. Exogenous variables and feature engineering⁷

Exogenous variables are those included in a forecasting system because they add value but are not being predicted themselves, and are sometimes called ‘features’ (see Section 2.7.4). For example, a forecast of county’s energy demand may be based on the recent history of demand (an *endogenous* variable), but also weather forecasts, which are exogenous variables. Many time series methods have extensions that facilitate exogenous variables, such as autoregression with exogenous variables (ARX). However, it is often necessary to prepare exogenous data before use, for example so that it matches the temporal resolution of the variable being forecast (hourly, daily, and so on). Exogenous variables may be numeric or categorical, and may be numerous. Different types of predictor present different issues depending on the predictive model being used. For instance, models based on the variable’s absolute value can be sensitive to extreme values or skewness, whereas models based on the variable value’s rank, such as tree-based models, are not. Exogenous variables that are correlated with one another also poses a challenge for some models, and techniques such as regularisation and partial leased squares have been developed to mitigate this.

Interactions between exogenous variables may also be important when making predictions. For example, crop yields depend on both rainfall and sunlight: one without the other or both in excess will result in low yields, but the right combination will result in high yields. Interactions may be included in linear models by including product of the two interacting exogenous as a feature in the model. This is an example of feature engineering, the process of creating new features based on domain knowledge or exploratory analysis of available data. In machine learning (see Section 2.7.10), many features may be created by combining exogenous variables speculatively and passed to a selection algorithm to identify those with predictive power. Combinations are not limited to products, or only two interacting variables, and where many exogenous variables are available, could include summary statistics (mean, standard deviation, range, quantiles...) of groups of variables.

Where exogenous variables are numerous dimension reduction may be applied to reduce the number of features in a forecasting model (see also Section 2.5.3). Dimension reduction transforms multivariate data into a lower dimensional representation while

retaining meaningful information about the original data. Principal component analysis (PCA) is a widely used method for linear dimension reduction, and non-linear alternatives are also available. PCA is useful when the number of candidate predictors is greater than the number of time series observations, as is often the case in macroeconomic forecasting ([Stock & Watson, 2002](#)). It is routinely applied in applications from weather to sales forecasting. In retail forecasting, for example, past sales of thousands of products may be recorded but including them all as exogenous variables in the forecasting model for an individual product may be impractical. Dimension reduction offers an alternative to only using a subset of the available features.

Preparation of data for forecasting tasks is increasingly important as the volume of available data is increasing in many application areas. Further details and practical examples can be found in [Albon \(2018\)](#) and [Kuhn and Johnson \(2019\)](#) among other texts in this area. For deeper technical discussion of a range of non-linear dimension reduction algorithms, see [Hastie, Tibshirani, and Friedman \(2009\)](#).

2.3. Statistical and econometric models

2.3.1. Exponential smoothing models⁸

Exponential smoothing is one of the workhorses of business forecasting. Despite the many advances in the field, it is always a tough benchmark to bear in mind. The development of exponential smoothing dates back to 1944, where Robert G. Brown through a mechanical computing device estimated key variables for fire-control on the location of submarines ([Gardner, 1985](#)). More details about the state of the art of exponential smoothing can be found in [Gardner \(2006\)](#).

The idea behind exponential smoothing relies on the weighted average of past observations, where that weight decreases exponentially as one moves away from the present observations. The appropriate exponential smoothing method depends on the components that appear in the time series. For instance, in case that no clear trend or seasonal pattern is present, the simplest form of exponential smoothing methods known as Simple (or Single) Exponential Smoothing (SES) is adequate, such as:

In some references, is also known as Exponentially Weighted Moving Average ([Harvey, 1990](#)). The formula for SES can be obtained from minimising the discounted least

squares error function and expressing the resulting equation in a recursive form (Harvey, 1990). If observations do not have the same weight, the ordinary least squares cannot be applied. On the other hand, the recursive form is very well-suited for saving data storage.

In order to use SES, we need to estimate the initial forecast (\hat{y}_1) and the exponential smoothing parameter (α). Traditionally, the initialisation was done by using either ad hoc values or a heuristic scheme (Hyndman, Koehler, Ord, & Snyder, 2008), however nowadays it is standard to estimate both the initial forecast and the optimal smoothing parameter by minimising the sum of squares of the one-step ahead forecast errors. The estimation of the smoothing parameter usually is restricted to values between 0 and 1. Once SES is defined, the method only provides point forecasts, i.e., forecasts of the mean. Nonetheless, it is of vital importance for many applications to provide density (probabilistic) forecasts. To that end, Hyndman, Koehler, Snyder, and Grose (2002) extended exponential smoothing methods under State Space models using a single source of error (see Section 2.3.6) to equip them with a statistical framework capable of providing future probability distributions. For example, SES can be expressed in the State Space as a local level model: where the state is the level (μ_t) and ϵ_t is the Gaussian noise. Note the difference between traditional exponential smoothing methods and exponential smoothing models (under the state space approach). The former only provide point forecasts, meanwhile the latter also offers probabilistic forecasts, which obviously includes prediction intervals. In addition, some exponential smoothing models can be expressed as ARIMA models (see also Section 2.3.4).

So far, we have introduced the main exponential smoothing using SES, however real time series can include other components as trends, seasonal patterns, cycles, and the irregular (error) component. In this sense, the exponential smoothing version capable of handling local trends is commonly known as Holt's method (Holt, 2004 originally published in 1957) and, if it also models a seasonality component, which can be incorporated in an additive or multiplicative fashion, it is called Holt-Winters method (Winters, 1960). Exponential smoothing models have been also extended to handle multiple seasonal cycles; see Section 2.3.5.

Fortunately, for various combinations of time series patterns (level only, trended, seasonal, trended and seasonal) a particular exponential smoothing can be chosen. Pegels (1969) proposed a first classification of exponential smoothing methods,

later extended by [Gardner \(1985\)](#) and [Taylor \(2003a\)](#). The state space framework mentioned above, developed by [Hyndman et al. \(2002\)](#), allowed to compute the likelihood for each exponential smoothing model and, thus, model selection criteria such as AIC could be used to automatically identify the appropriate exponential smoothing model. Note that the equivalent state space formulation was derived by using a single source of error instead of a multiple source of error ([Harvey, 1990](#)). [Hyndman et al. \(2008\)](#) utilised the notation (E,T,S) to classify the exponential smoothing models, where those letters refer to the following components: Error, Trend, and Seasonality. This notation has gained popularity because the widely-used *forecast* package ([Hyndman et al., 2020](#)), recently updated to the *fable* package, for R statistical software, and nowadays exponential smoothing is frequently called ETS.

2.3.2. Time-series regression models⁹

The key idea of linear regression models is that a target (or dependent, forecast, explained, regress) variable, y , i.e., a time series of interest, can be forecast through other regressor (or independent, predictor, explanatory) variables, x , i.e., time series or features (see Section 2.2.5), assuming that a linear relationship exists between them, as follows where ϵ is the residual error of the model at time t , β_0 is a constant, and coefficient β_1 is the effect of regressor x after taking into account the effects of all regressors involved in the model. For example, daily product sales may be forecast using information related with past sales, prices, advertising, promotions, special days, and holidays (see also Section 3.2.4).

In order to estimate the model, forecasters typically minimise the sum of the squared errors (ordinary least squares estimation, OLS), using the observations available for fitting the model to the data ([Ord, Fildes, & Kourentzes, 2017](#)) and setting the gradient equal to zero. If the model is simple, consisting of a single regressor, then two coefficients are computed, which are the slope (coefficient of the regressor) and the intercept (constant). When more regressor variables are considered, the model is characterised as a multiple regression one and additional coefficients are estimated. A common way to evaluate how well a linear regression model fits the target series, reporting an average value of R^2 , is through the coefficient of determination, R^2 , indicating the proportion of variation in the dependent variable explained by the model. Values close to one indicate sufficient goodness-of-fit, while values close to zero insufficient fitting. However, goodness-of-fit should not be confused with forecastability ([Harrell,](#)

2015). When the complexity of the model is increased, i.e., more regressors are considered, the value of the coefficient will also rise, even if such additions lead to overfitting (see Section 2.5.2). Thus, regression models should be evaluated using cross-validation approaches (see Section 2.5.5), approximating the post-sample accuracy of the model, or measures that account for model complexity, such as information criteria (e.g., AIC, AICc, and BIC) and the adjusted coefficient of determination, (James, Witten, Hastie, & Tibshirani, 2013). Other diagnostics are the standard deviation of the residuals and the t-values of the regressors. Residual standard error, σ , summarises the average error produced by the model given the number of regressors used, thus accounting for overfitting. The t-values measure the impact of excluding regressors from the model in terms of error, given the variation in the data, thus highlighting the importance of the regressors.

To make sure that the produced forecasts are reliable, the correlation between the residuals and the observations of the regressors must be zero, with the former displaying also insignificant autocorrelation. Other assumptions suggest that the residuals should be normally distributed with an average value of zero and that their variability should be equal across time (no heteroscedasticity present). Nevertheless, in practice, it is rarely necessary for residuals to be normally distributed in order for the model to produce accurate results, while the homoscedasticity assumption becomes relevant mostly when computing prediction intervals. If these assumptions are violated, that may mean that part of the variance of the target variable has not been explained by the model and, therefore, that other or more regressors are needed. In case of non-linear dependencies between the target and the regressor variables, data power transformations (see Section 2.2.1) or machine learning approaches can be considered (see Section 2.7.10).

Apart from time series regressors, regression models can also exploit categorical (dummy or indicator) variables (Hyndman & Athanasopoulos, 2018) which may e.g., inform the model about promotions, special events, and holidays (binary variables), the day of the week or month of the year (seasonal dummy variables provided as one-hot encoded vectors), trends and structural changes, and the number of trading/working days included in the examined period. In cases where the target series is long and displays complex seasonal patterns, additional regressors such as Fourier series and lagged values of both the target and the regressor variables may become

useful. Moreover, when the number of the potential regressor variables is significant compared to the observations available for estimating the respective coefficients (see Section 2.7.1), step-wise regression (James et al., 2013) or dimension reduction and shrinkage estimation methods (see Section 2.5.3) can be considered to facilitate training and avoid overfitting. Finally, mixed data sampling (MIDAS) regression models are a way of allowing different degrees of temporal aggregation for the regressors and predictand (see also Section 2.10.2 for further discussions on forecasting with temporal aggregation).

2.3.3. Theta method and models¹⁰

In the age of vast computing power and computational intelligence, the contribution of simple forecasting methods is possibly not *en vogue*; the implementation of complicated forecasting systems becomes not only expedient but possibly desirable. Nevertheless forecasting, being a tricky business, does not always favour the complicated or the computationally intensive. Enter the theta method. From its beginnings 20 years back in Assimakopoulos and Nikolopoulos (2000) to recent advances in the monograph of Nikolopoulos and Thomakos (2019), to other work in-between and recently too, the theta method has emerged as not only a powerfully simple but also enduring method in modern time series forecasting. The reader will benefit by reviewing Sections 2.3.1 Exponential smoothing models, 2.3.4 Autoregressive integrated moving average (ARIMA) models, 2.3.9 Forecasting with many variables for useful background information.

The original idea has been now fully explained and understood and, as Nikolopoulos and Thomakos (2019) have shown, even the revered AR(1) model forecast is indeed a theta forecast – and it has already been shown by Hyndman and Billah (2003) that the theta method can represent SES (with a drift) forecasts as well. In its simplest form the method generates a forecast from a linear combination of the last observation and some form of “trend” function, be that a constant, a linear trend, a non-parametric trend or a non-linear trend. In summary, and under the conditions outlined extensively in Nikolopoulos and Thomakos (2019), the theta forecasts can be expressed as functions of the “theta line”: where \hat{y}_t is the trend function, variously defined depending on the modelling approach and type of trend one is considering in applications. It can be shown that the, univariate, theta forecasts can given either as when the trend function is defined as $\hat{y}_t = \mu$ and as when the trend function is left otherwise unspecified. The choice of the

weight parameter on the linear combination of the theta line, the choice and number of trend functions and their nature and other aspects on expanding the method have been recently researched extensively.

The main literature has two strands. The first one details the probabilistic background of the method and derives certain theoretical properties, as in [Hyndman and Billah \(2003\)](#), [Thomakos and Nikolopoulos, 2012](#), [Thomakos and Nikolopoulos, 2015](#) and a number of new theoretical results in [Nikolopoulos and Thomakos \(2019\)](#). The work of Thomakos and Nikolopoulos provided a complete analysis of the theta method under the unit root data generating process, explained its success in the M3 competition ([Makridakis & Hibon, 2000](#)), introduced the multivariate theta method and related it to cointegration and provided a number of other analytical results for different trend functions and multivariate forecasting. The second strand of the literature expands and details various implementation (including hybrid approaches) of the method, as in the theta approach in supply chain planning of [Nikolopoulos, Assimakopoulos, Bougioukos, Litsa, and Petropoulos \(2012\)](#), the optimised theta models and their relationship with state space models in [Fioruci, Pellegrini, Louzada, and Petropoulos \(2015\)](#) and [Fiorucci, Pellegrini, Louzada, Petropoulos, and Koehler \(2016\)](#), hybrid approaches as in [Theodosiou \(2011\)](#) and [Spiliotis, Assimakopoulos and Nikolopoulos \(2019\)](#), to the very latest generalised theta method of [Spiliotis, Assimakopoulos and Makridakis \(2020\)](#). These are major methodological references in the field, in addition to many others of pure application of the method.

The theta method is also part of the family of adaptive models/methods, and a simple example illustrates the point: the AR(1) forecast or the SES forecast are both theta forecasts but they are also both adaptive learning forecasts, as in the definitions of the recent work by [Kyriazi, Thomakos, and Guerard \(2019\)](#). As such, the theta forecasts contain the basic building blocks of successful forecasts: simplicity, theoretical foundations, adaptability and performance enhancements. Further research on the usage of the theta method within the context of adaptive learning appears to be a natural next step. In the context of this section, see also Section [2.3.16](#) on equilibrium correcting models and forecasts.

Given the simplicity of its application, the freely available libraries of its computation, its scalability and performance, the theta method should be considered as a critical

benchmark henceforth in the literature – no amount of complexity is worth its weight if it cannot beat a single Greek letter!

2.3.4. Autoregressive integrated moving average (ARIMA) models¹¹

Time series models that are often used for forecasting are of the autoregressive integrated moving average class (ARIMA: [Box, George, Jenkins, & Gwilym, 1976](#)). The notation of an ARIMA(p, d, q) model for a time series y_t is where the lag operator B is defined by $B^k y_t = y_{t-k}$. The ϵ_t is a zero-mean uncorrelated process with common variance σ^2 . Some exponential smoothing models (see Section 2.3.1) can also be written in ARIMA format, where some ETS models assume that $d=1$ or $d=2$. For example, SES is equivalent to ARIMA(0,1,1) when $\alpha=1$. The parameters in the ARIMA model can be estimated using Maximum Likelihood, whereas for the ARIMA($p, d, 0$) Ordinary Least Squares can be used. The iterative model-building process ([Franses, van Dijk, & Opschoor, 2014](#)) requires the determination of the values of p , d , and q . Data features as the empirical autocorrelation function and the empirical partial autocorrelation function can be used to identify the values of p and q , in case of low values of p and q . Otherwise, in practice one relies on the well-known information criteria like AIC and BIC (see Section 2.5.4). The function `auto.arima` of the *forecast* package ([Hyndman et al., 2020](#)) for R statistical software compares models using information criteria, and has been found to be very effective and increasingly being used in ARIMA modelling.

Forecasts from ARIMA models are easy to make. And, at the same time, prediction intervals can be easily computed. Take for example the ARIMA(1,0,1) model: The one-step-ahead forecast from forecast origin t is $\hat{y}_{t+1|t}$ as the expected value $E(y_{t+1}|y_t)$. The forecast error is $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$ and, hence, the forecast error variance is σ^2 . The two-steps-ahead forecast from t is $\hat{y}_{t+2|t}$ with the forecast error equal to $e_{t+2|t} = y_{t+2} - \hat{y}_{t+2|t}$ and the forecast error variance $2\sigma^2$. These expressions show that the creation of forecasts and forecast errors straightforwardly follow from the model expressions, and hence can be automated if necessary.

An important decision when using an ARIMA model is the choice for the value of d . When $d=0$, the model is created for the levels of the time series, that is, y_t . When $d=1$, there is a model for Δy_t , and the data need to be differenced prior to fitting an ARMA model. In some specific but rare cases, $d=2$. The decision on the value of d is usually based on so-called tests for unit roots ([Dickey and Fuller, 1979](#), [Dickey and Pantula, 1987](#)). Under the null hypothesis that y_t is non-stationary, and the test involves non-standard statistical theory ([Phillips, 1987](#)). One can also choose to make Δy_t as the null

hypothesis (Hobijn et al., 2004, Kwiatkowski et al., 1992). The power of unit root tests is not large, and in practice one often finds signals to consider (Nelson & Plosser, 1982). For seasonal data, like quarterly and monthly time series, the ARIMA model can be extended to Seasonal ARIMA (SARIMA) models represented by $ARIMA(p, d, q)(P, D, Q)_m$, where p , d , and q are the seasonal parameters and the m is the periodicity. When $m=12$, the data are transformed as $y_t = y_{12t}$. It can also be that p and q , and then one can replace y_t by $y_t + \delta_t$ where the δ_t with δ_t are seasonal dummies. The choice of m is based on tests for so-called seasonal unit roots (Franses, 1991, Ghysels et al., 1994, Hylleberg et al., 1990).

Another popular extension to ARIMA models is called ARIMAX, implemented by incorporating additional exogenous variables (regressors) that are external to and different from the forecast variable. An alternative to ARIMAX is the use of regression models (see Section 2.3.2) with ARMA errors.

2.3.5. Forecasting for multiple seasonal cycles¹²

With the advances in digital data technologies, data is recorded more frequently in many sectors such as energy (Wang, Liu, & Hong, 2016 and Section 3.4), healthcare (Whitt & Zhang, 2019 and 3.6.1), transportation (Gould, Koehler, Ord, Snyder, Hyndman, & Vahid-Araghi, 2008), and telecommunication (Meade & Islam, 2015a). This often results in time series that exhibit multiple seasonal cycles (MSC) of different lengths. Forecasting problems involving such series have been increasingly drawing the attention of both researchers and practitioners leading to the development of several approaches. Multiple Linear Regression (MLR) is a common approach to model series with MSC (Kamisan et al., 2018, Rostami-Tabar and Ziel, 2020); for an introduction on time-series regression models, see Section 2.3.2. While MLR is fast, flexible, and uses exogenous regressors, it does not allow to decompose components and change them over time. Building on the foundation of the regression, Facebook introduced Prophet (Taylor & Letham, 2018), an automated approach that utilises the Generalised Additive Model (Hastie & Tibshirani, 1990). Although the implementation of Prophet may be less flexible, it is easy to use, robust to missing values and structural changes, and can handles outliers well.

Some studies have extended the classical ARIMA (see Section 2.3.4) and Exponential Smoothing (ETS; see Section 2.3.1) methods to accommodate MSC.

Multiple/multiplicative Seasonal ARIMA (MSARIMA) model is an extension of ARIMA for the case of MSC (Taylor, 2003b). MSARIMA allows for exogenous regressors and

terms can evolve over time, however, it is not flexible, and the computational time is high. [Svetunkov and Boylan \(2020\)](#) introduced the Several Seasonalities ARIMA (SSARIMA) model which constructs ARIMA in a state-space form with several seasonalities. While SSARIMA is flexible and allows for exogenous regressors, it is computationally expensive, especially for high frequency series.

[Taylor \(2003b\)](#) introduced Double Seasonal Holt-Winters (DSHW) to extend ETS for modelling daily and weekly seasonal cycles. Following that, [Taylor \(2010\)](#) proposed a triple seasonal model to consider the intraday, intraweek and intrayear seasonalities. [Gould et al. \(2008\)](#) and [Taylor and Snyder \(2012\)](#) instead proposed an approach that combines a parsimonious representation of the seasonal states up to a weekly period in an innovation state space model. With these models, components can change, and decomposition is possible. However, the implementation is not flexible, the use of exogenous regressors is not supported, and the computational time could be high. An alternative approach for forecasting series with MSC is TBATS ([De Livera et al., 2011](#) see also Section 2.2.2). TBATS uses a combination of Fourier terms with an exponential smoothing state space model and a Box–Cox transformation (see Section 2.2.1), in an entirely automated manner. It allows for terms to evolve over time and produce accurate forecasts. Some drawbacks of TBATS, however, are that it is not flexible, can be slow, and does not allow for covariates.

In response to shortcomings in current models, Forecasting with Additive Switching of Seasonality, Trend and Exogenous Regressors (FASSTER) has been proposed by [O’Hara-Wild and Hyndman \(2020\)](#). FASSTER is fast, flexible and support the use of exogenous regressors into a state space model. It extends state space models such as TBATS by introducing a switching component to the measurement equation which captures groups of irregular multiple seasonality by switching between states.

In recent years, Machine Learning (ML; see Section 2.7.10) approaches have also been recommended for forecasting time series with MSC.

MultiLayer Perceptron (MLP: [Dudek, 2013](#), [Zhang and Qi, 2005](#)), Recurrent Neural Networks (RNN: [Lai, Chang, Yang, & Liu, 2018](#)), Generalised Regression Neural Network (GRNN: [Dudek, 2015](#)), and Long Short-Term Memory Networks (LSTM: [Zheng, Xu, Zhang, & Li, 2017](#)) have been applied on real data ([Bandara et al., 2020a](#), [Xie and Ding, 2020](#)) with promising results. These approaches are flexible, allow for any exogenous regressor and suitable when non-

linearity exists in series, however interpretability might be an issue for users (Makridakis, Spiliotis, & Assimakopoulos, 2018).

2.3.6. State-space models¹³

State Space (SS) systems are a very powerful and useful framework for time series and econometric modelling and forecasting. Such systems were initially developed by engineers, but have been widely adopted and developed in Economics as well (Durbin and Koopman, 2012, Harvey, 1990). The main distinguishing feature of SS systems is that the model is formulated in terms of *states* (x_t), which are a set of variables usually unobserved, but which have some meaning. Typical examples are trends, seasonal components or time varying parameters.

A SS system is built as the combination of two sets of equations:

(i) *state* or *transition* equations which describe the dynamic law governing the states between two adjacent points in time; and (ii) *observation* equations which specify the relation between observed data (both inputs and outputs) and the unobserved states. A linear version of such a system is shown in Eq. (1).(1)

In this equations x_t and y_t are the state and observational vectors of zero mean Gaussian noises with covariance Σ_x , Σ_y , Σ_{xy} , and Σ_{yx} are the so-called (time-varying) system matrices, and x_0 and Σ_{x0} are the initial state and state covariance matrix, respectively. Note that x_t and y_t may be parameterised to include some input variables as linear or non-linear relations to the output variables y_t .

The model in Eq. (1) is a *multiple error SS model*. A different formulation is the *single error SS model* or the *innovations SS model*. This latter is similar to (1), but replacing x_t and y_t by x_t and y_t , respectively. Then, naturally, the innovations form may be seen as a restricted version of model (1), but, conversely, under weak assumptions, (1) may also be written as an observationally equivalent *innovations form* (see, for example, Casals, Garcia-Hiernaux, Jerez, Sotoca, & Trindade, 2016 pp. 12–17).

Once a SS system is fully specified, the core problem is to provide optimal estimates of states and their covariance matrix over time. This can be done in two ways, either by looking back in time using the well-known Kalman filter (useful for online applications) or taking into account the whole sample provided by smoothing algorithms (typical of offline applications) (Anderson & Moore, 1979).

Given any set of data and a specific model, the system is not fully specified in most cases because it usually depends on unknown parameters scattered throughout the system

matrices that define the SS equations. Estimation of such parameters is normally carried out by Maximum Likelihood defined by prediction error decomposition (Harvey, 1990). Non-linear and non-Gaussian models are also possible, but at the cost of a higher computational burden because more sophisticated recursive algorithms have to be applied, like the extended Kalman filters and smoothers of different orders, particle filters (Doucet & Gordon, 2001), Unscented Kalman filter and smoother (Julier & Uhlmann, 1997), or simulation of many kinds, like Monte Carlo, bootstrapping or importance sampling (Durbin & Koopman, 2012).

The paramount advantage of SS systems is that they are not a particular model or family of models strictly speaking, but a container in which many very different model families may be implemented, indeed many treated in other sections of this paper. The following is a list of possibilities, not at all exhaustive:

- •
Univariate models with or without inputs: regression (Section 2.3.2), ARIMAX (Section 2.3.4), transfer functions, exponential smoothing (Section 2.3.1), structural unobserved components, Hodrick–Prescott filter, spline smoothing.
- •
Fully multivariate: natural extensions of the previous ones plus echelon-form VARIMAX, Structural VAR, VECM, Dynamic Factor models, panel data (Section 2.3.9).
- •
Non-linear and non-Gaussian: TAR, ARCH, GARCH (Section 2.3.11), Stochastic Volatility (Durbin & Koopman, 2012), Dynamic Conditional Score (Harvey, 2013), Generalised Autoregressive Score (Creal, Koopman, & Lucas, 2013), multiplicative unobserved components.
- •
Other: periodic cubic splines, periodic unobserved components models, state dependent models, Gegenbauer long memory processes (Dissanayake, Peiris, & Proietti, 2018).

Once any researcher or practitioner becomes acquainted to a certain degree with the SS technology, some important advanced issues in time series forecasting may be comfortably addressed (Casals et al., 2016). It is the case, for example, of systems block concatenation, systems nesting in errors or in variables, treating errors in variables,

continuous time models, time irregularly spaced data, mixed frequency models, time varying parameters, time aggregation, hierarchical and group forecasting (Villegas & Pedregal, 2018) (time, longitudinal or both), homogeneity of multivariate models (proportional covariance structure among perturbations), etc.

All in all, the SS systems offer a framework capable of handling many modelling and forecasting techniques available nowadays in a single environment. Once the initial barriers are overcome, a wide panorama of modelling opportunities opens up.

2.3.7. Models for population processes¹⁴

Over the past two centuries, formal demography has established its own, discipline-specific body of methods for predicting (or *projecting*¹⁵) populations. Population sciences, since their 17th century beginnings, have been traditionally very empirically focused, with strong links with probability theory (Courgeau, 2012). Given the observed regularities in population dynamics, and that populations are somewhat better predictable than many other socio-economic processes, with reasonable horizons possibly up to one generation ahead (Keyfitz, 1972, Keyfitz, 1981), demographic forecasts have become a bestselling product of the discipline (Xie, 2000). Since the 20th century, methodological developments in human demography have been augmented by the work carried out in mathematical biology and population ecology (Caswell, 2019a). The theoretical strength of demography also lies almost exclusively in the formal mathematical description of population processes (Burch, 2018), typically growth functions and structural changes. Historically, such attempts started from formulating the logistic model of population dynamics, inspired by the Malthusian theory (Pearl and Reed, 1920, Verhulst, 1845). Lotka's (1907) work laid the foundations of the stable population theory with asymptotic stability under constant vital rates, subsequently extended to modelling of interacting populations by using differential equations (Lotka, 1925, Volterra, 1926b). By the middle of the 20th century, the potential and limitations of demographic forecasting methods were already well recognised in the literature (Brass, 1974, Hajnal, 1955).

In the state-of-the-art demographic forecasting, the core engine is provided by matrix algebra. The most common approach relies on the cohort-component models, which combine the assumptions on fertility, mortality and migration, in order to produce future population by age, sex, and other characteristics. In such models, the deterministic mechanism of population renewal is known, and results from the

following demographic accounting identity (population balancing equation, see [Bryant and Zhang, 2018](#), [Rees and Wilson, 1973](#)):

where N_t denotes population aged a at time t , $D_{a,t}$ refer to deaths between ages a and $a+1$ in the time interval t to $t+1$, with $I_{a,t}$ and $E_{a,t}$ respectively denoting immigration (and other entries) and emigration (and other exits). In addition, for the youngest age group, births B_t need to be added. The equation above can be written up in the general algebraic form: $N_{t+1} = M N_t + B_t$, where N_t is the population vector structured by age (and other characteristics), and M is an appropriately chosen growth matrix (Leslie matrix), closely linked with the life table while reflecting the relationship above, expressed in terms of rates rather than events ([Caswell, 2019a](#), [Leslie, 1945](#), [Leslie, 1948](#), [Preston et al., 2000](#)).

In the cohort-component approach, even though the mechanism of population change is known, the individual components still need forecasting. The three main drivers of population dynamics — fertility, mortality, and migration — differ in terms of their predictability ([National Research Council, 2000](#)): mortality, which is mainly a biological process moderated by medical technology, is the most predictable; migration, which is purely a social and behavioural process is the least; while the predictability of fertility — part-biological, part-behavioural — is in the middle (for component forecasting methods, see [Sections 3.6.3 Forecasting mortality](#), [3.6.4 Forecasting fertility](#), [3.6.5 Forecasting migration](#)). In practical applications, the components can be either projected deterministically, following judgment-based or expert assumptions (for example, [Lutz, Butz, & Samir, 2017](#)), or extrapolated by using probabilistic methods, either for the components or for past errors of prediction ([Alho and Spencer, 1985](#), [Alho and Spencer, 2005](#), [De Beer, 2008](#)). An impetus to the use of stochastic methods has been given by the developments in the [UN World Population Prospects](#) ([Azose et al., 2016](#), [Gerland et al., 2014](#)). Parallel, theoretical advancements included a stochastic version of the stable population theory ([Keiding & Hoem, 1976](#)), as well as coupling of demographic uncertainty with economic models ([Alho, Hougaard Jensen, & Lassila, 2008](#)). Since its original formulation, the cohort-component model has been subject to several extensions (see, for example, [Stillwell & Clarke, 2011](#)). The multiregional model ([Rogers, 1975](#)) describes the dynamics of multiple regional populations at the same time, with regions linked through migration. The multistate model ([Schoen, 1987](#)) generalises the multiregional analysis to any arbitrary set of states (such as educational, marital, health, or employment statuses, and so on; see also state-space models in [Section 2.3.6](#)). The

multiregional model can be in turn generalised to include multiple geographic levels of analysis in a coherent way (Kupiszewski & Kupiszewska, 2011). Recent developments include multifocal analysis, with an algebraic description of kinship networks (Caswell, 2019b, Caswell, 2020). For all these extensions, however, data requirements are very high: such models require detailed information on transitions between regions or states in a range of different breakdowns. For pragmatic reasons, microsimulation-based methods offer an appealing alternative, typically including large-sample Monte Carlo simulations of population trajectories based on available transition rates (Bélanger and Sabourin, 2017, Zaidi et al., 2009).

Aside of a few extensions listed above, the current methodological developments in the forecasting of human populations are mainly concentrated on the approaches for predicting individual demographic components (see Sections 3.6.3 Forecasting mortality, 3.6.4 Forecasting fertility, 3.6.5 Forecasting migration), rather than the description of the population renewal mechanism. Still, the continuing developments in population ecology, for example on the algebraic description of transient and asymptotic population growth (Nicol-Harper, Dooley, Packman, Mueller, Bijak, Hodgson, Townley, & Ezard, 2018), bear substantial promise of further advancements in this area, which can be additionally helped by strengthened collaboration between modellers and forecasters working across the disciplinary boundaries on the formal descriptions of the dynamics of human, as well as other populations.

2.3.8. Forecasting count time series¹⁶

Probabilistic forecasts based on predictive mass functions are the most natural way of framing predictions of a variable that enumerates the occurrences of an event over time; i.e. the most natural way of predicting a time series of *counts*. Such forecasts are both *coherent*, in the sense of being consistent with the discrete support of the variable, and capture all distributional – including tail – information. In contrast, point forecasts based on summary measures of central location (e.g., a (conditional) mean, median or mode), convey no such distributional information, and potentially also lack coherence as, for example, when the mean forecast of the integer-valued count variable assumes non-integer values. These comments are even more pertinent for *low count* time series, in which the number of *rare* events is recorded, and for which the cardinality of the support is small. In this case, point forecasts of any sort can be misleading, and

continuous (e.g., Gaussian) approximations (sometimes adopted for high count time series) are particularly inappropriate.

These points were first elucidated in [Freeland and McCabe \(2004\)](#), and their subsequent acceptance in the literature is evidenced by the numerous count data types for which discrete predictive distributions are now produced; including counts of: insurance claims ([McCabe & Martin, 2005](#)), medical injury deaths ([Bu & McCabe, 2008](#)), website visits ([Bisaglia & Canale, 2016](#)), disease cases ([Bisaglia and Gerolimetto, 2019](#), [Mukhopadhyay and Sathish, 2019](#), [Rao and McCabe, 2016](#)), banking crises ([Dungey, Martin, Tang, & Tremayne, 2020](#)), company liquidations ([Homburg, Weiß, Alwan, Frahm, & Göb, 2020](#)), hospital emergency admissions ([Sun, Sun, Zhang, & McCabe, 2021](#)), work stoppages ([Weiß, Homburg, Alwan, Frahm, & Göb, 2021](#)), and the intermittent product demand described in Section 2.8 ([Berry and West, 2020](#), [Kolassa, 2016](#), [Snyder et al., 2012](#)).

The nature of the predictive model for the count variable, together with the paradigm adopted (Bayesian or frequentist), determine the form of the probabilistic forecast, including the way in which it does, or does not, accommodate parameter and model uncertainty. As highlighted in Sections [2.4.1 Foundations of Bayesian forecasting](#), [2.4.2 Implementation of Bayesian forecasting](#), the Bayesian approach to forecasting is *automatically* probabilistic, no matter what the data type. It also factors parameter uncertainty into the predictive distribution, plus model uncertainty if Bayesian model averaging is adopted, producing a distribution whose location, shape and degree of dispersion reflect all such uncertainty as a consequence. See [Berry and West, 2020](#), [Bisaglia and Canale, 2016](#), [Frazier et al., 2019](#), [Lu, 2021](#), [McCabe and Martin, 2005](#) and [Neal and Kypraios \(2015\)](#), for examples of Bayesian probabilistic forecasts of counts.

In contrast, frequentist probabilistic forecasts of counts typically adopt a ‘plug-in’ approach, with the predictive distribution conditioned on estimates of the unknown parameters of a given count model. Sampling variation in the estimated predictive (if acknowledged) is quantified in a variety of ways. [Freeland and McCabe \(2004\)](#), for instance, produce confidence intervals for the true (point-wise) predictive probabilities, exploiting the asymptotic distribution of the (MLE-based) estimates of those probabilities. [Bu and McCabe \(2008\)](#) extend this idea to (correlated) estimates of sequential probabilities, whilst [Jung and Tremayne \(2006\)](#) and [Weiß et al. \(2021\)](#) exploit

bootstrap techniques to capture point-wise sampling variation in the forecast distribution. [McCabe, Martin, and Harris \(2011\)](#), on the other hand, use subsampling methods to capture sampling fluctuations in the *full* predictive distribution, retaining the non-negativity and summation to unity properties of the probabilities (see also [Harris, Martin, Perera, & Poskitt, 2019](#) for related, albeit non-count data work). Model uncertainty is catered for in a variety of ways: via nonparametric ([McCabe et al., 2011](#)) or bootstrapping ([Bisaglia & Gerolimetto, 2019](#)) methods; via (frequentist) model averaging ([Sun et al., 2021](#)); or via an informal comparison of predictive results across alternative models ([Jung & Tremayne, 2006](#)). Methods designed explicitly for calibrating predictive mass functions to observed count data – whether those functions be produced using frequentist *or Bayesian methods* – can be found in [Czado, Gneiting, and Held \(2009\)](#) and [Wei and Held \(2014\)](#); see also Sections [2.12.4 Evaluating probabilistic forecasts](#), [2.12.5 Assessing the reliability of probabilistic forecasts](#). Finally, whilst full probabilistic forecasts are increasingly common, point, interval and quantile forecasts are certainly still used. The need for such summaries to be coherent with the discrete nature of the count variable appears to be now well-accepted, with recent work emphasising the importance of this property ([Bu and McCabe, 2008](#), [Homburg et al., 2019](#), [Homburg et al., 2020](#), [Mukhopadhyay and Sathish, 2019](#)).

2.3.9. Forecasting with many variables¹⁷

Multivariate models – regression models with multiple explanatory variables – are often based on available theories regarding the determination of the variable to be forecast, and are often referred to as *structural models*. In a stationary world without structural change, then it would be anticipated that the best structural model would provide the best forecasts, since it would provide the conditional mean of the data process (see, for example, [Clements & Hendry, 1998](#)). In a non-stationary world of unit roots and structural breaks, however, this need not be the case. In such situations, often simple forecast models can outperform structural models, especially at short forecast horizons (see, for example, [Hendry & Clements, 2001](#)). Multivariate forecast models require that explanatory variables also be forecast – or at least, scenarios be set out for them. These may be simplistic scenarios, for example all explanatory variables take their mean values. Such scenarios can play a useful role in formulating policy making since they illustrate in some sense the outcomes of different policy choices.

Since the 1980s and [Sims \(1980\)](#), vector autoregressive (VAR) models have become ubiquitous in macroeconomics, and common in finance (see, for example, [Hasbrouck, 1995](#)). A VAR model is a set of linear regression equations (see also Section [2.3.2](#)) describing the evolution of a set of endogenous variables. Each equation casts each variable as a function of lagged values of all the variables in the system.

Contemporaneous values of system variables are not included in VAR models for identification purposes; some set of identifying restrictions are required, usually based on economic theory, and when imposed the resulting model is known as a structural VAR model. VAR models introduce significantly greater levels of parameterisation of relationships, which increases the level of estimation uncertainty. At the same time VAR models afford the forecaster a straightforward way to generate forecasts of a range of variables, a problem when forecasting with many variables. As with autoregressive methods, VAR models can capture a significant amount of variation in data series that are autocorrelated, and hence VAR methods can be useful as baseline forecasting devices. VAR-based forecasts are often used as a benchmark for complex models in macroeconomics like DSGE models (see, for example, [Del Negro & Schorfheide, 2006](#)). The curse of dimensionality in VAR models is particularly important and has led to developments in factor-augmented VAR models, with practitioners often reducing down hundreds of variables into factors using principal component analysis (see, for example, [Bernanke, Boivin, & Elias, 2005](#)). Bayesian estimation is often combined with factor-augmented VAR models.

Often, significant numbers of outliers and structural breaks require many indicator variables to be used to model series (see also Sections [2.2.3 Anomaly detection and time series forecasting](#), [2.2.4 Robust handling of outliers in time series forecasting](#)). Indicator saturation is a method of detecting outliers and structural breaks by saturating a model with different types of indicator, or deterministic variables ([Castle et al., 2015c](#), [Johansen and Nielsen, 2009](#)). The flexibility of the approach is such that it has been applied in a wide variety of contexts, from volcanic eruptions ([Pretis, Schneider, & Smerdon, 2016](#)) to prediction markets and social media trends ([Vaughan Williams & Reade, 2016](#)).

A particularly important and ever-expanding area of empirical analysis involves the use of panel data sets with long time dimensions: panel time series ([Eberhardt, 2012](#)). The many variables are then extended across many cross sectional units, and a central

concern is the dependence between these units, be they countries, firms, or individuals. At the country level one approach to modelling this dependence has been the Global VAR approach of, for example, [Dees, Mauro, Pesaran, and Smith \(2007\)](#). In more general panels, the mean groups estimator has been proposed to account for cross-section dependence ([Pesaran, Shin, & Smith, 1999](#)).

Outliers, structural breaks, and split trends undoubtedly also exist in panel time series. The potential to test for common outliers and structural changes across cross sectional units would be useful, as would the ability to allow individual units to vary individually, e.g., time-varying fixed effects. [Nymoen and Sparrman \(2015\)](#) is the first application of indicator saturation methods in a panel context, looking at equilibrium unemployment dynamics in a panel of [OECD](#) countries, but applications into the panel context are somewhat constrained by computer software packages designed for indicator saturation (Section [3.3.3](#) discusses further the case of forecasting unemployment). The *gets* R package of [Pretis et al., 2017](#), [Pretis et al., 2018](#) can be used with panel data.

2.3.10. Functional time series models¹⁸

Functional time series consist of random functions observed at regular time intervals. Functional time series can be classified into two categories depending on if the continuum is also a time variable. On the one hand, functional time series can arise from measurements obtained by separating an almost continuous time record into consecutive intervals (e.g., days or years, see [Horváth & Kokoszka, 2012](#)). We refer to such data structure as sliced functional time series, examples of which include daily precipitation data ([Gromenko, Kokoszka, & Reimherr, 2017](#)). On the other hand, when the continuum is not a time variable, functional time series can also arise when observations over a period are considered as finite-dimensional realisations of an underlying continuous function (e.g., yearly age-specific mortality rates, see [Li, Robinson, & Shang, 2020e](#)).

Thanks to recent advances in computing storage, functional time series in the form of curves, images or shapes is common. As a result, functional time series analysis has received increasing attention. For instance, [Bosq \(2000\)](#) and [Bosq and Blanke \(2007\)](#) proposed the functional autoregressive of order 1 (FAR(1)) and derived one-step-ahead forecasts that are based on a regularised Yule–Walker equations. FAR(1) was later extended to FAR(), under which the order can be determined via [Kokoszka and Reimherr's \(2013\)](#) hypothesis testing. [Horváth, Liu, Rice, and Wang \(2020\)](#) compared

the forecasting performance between FAR(1), FAR(), and functional seasonal autoregressive models of [Chen, Marron, and Zhang \(2019b\)](#).

To overcome the curse of dimensionality (see also Sections [2.2.5 Exogenous variables and feature engineering](#), [2.5.2 Model complexity](#), [2.5.3 Variable selection](#)), a dimension reduction technique, such as functional principal component analysis (FPCA), is often used. [Aue, Norinho, and Hörmann \(2015\)](#) showed asymptotic equivalence between a FAR and a VAR model (for a discussion of VAR models, see Section [2.3.9](#)). Via an FPCA, [Aue et al. \(2015\)](#) proposed a forecasting method based on the VAR forecasts of principal component scores. This approach can be viewed as an extension of [Hyndman and Shang \(2009\)](#), in which principal component scores are forecast via a univariate time series forecasting method. With the purpose of forecasting, [Kargin and Onatski \(2008\)](#) proposed to estimate the FAR(1) model by using the method of predictive factors. [Klepsch and Klüppelberg \(2017\)](#) proposed a functional moving average process and introduced an innovations algorithm to obtain the best linear predictor. [Klepsch, Klüppelberg, and Wei \(2017\)](#) extended the VAR model to the vector autoregressive moving average model and proposed the functional autoregressive moving average model. The functional autoregressive moving average model can be seen as an extension of autoregressive integrated moving average model in the univariate time series literature (see Section [2.3.4](#)).

Extending short-memory to long-memory functional time series analysis, [Li et al. \(2020e\)](#) and [Li, Robinson, and Shang \(2021\)](#) considered local Whittle and rescale-range estimators in a functional autoregressive fractionally integrated moving average model. The models mentioned above require stationarity, which is often rejected. [Horváth, Kokoszka, and Rice \(2014\)](#) proposed a functional KPSS test for stationarity. [Chang, Kim, and Park \(2016\)](#) studied nonstationarity of the time series of state densities, while [Beare, Seo, and Seo \(2017\)](#) considered a cointegrated linear process in Hilbert space. [Nielsen, Seo, and Seong \(2019\)](#) proposed a variance ratio-type test to determine the dimension of the nonstationary subspace in a functional time series. [Li, Robinson, and Shang \(2020d\)](#) studied the estimation of the long-memory parameter in a functional fractionally integrated time series, covering the functional unit root. From a nonparametric perspective, [Besse, Cardot, and Stephenson \(2000\)](#) proposed a functional kernel regression method to model temporal dependence via a similarity measure characterised by semi-metric, bandwidth and kernel function. [Aneiros-Pérez](#)

and Vieu (2008) introduced a semi-functional partial linear model that combines linear and nonlinear covariates. Apart from conditional mean estimation, Hörmann, Horváth, and Reeder (2013) considered a functional autoregressive conditional heteroscedasticity model for estimating conditional variance. Rice, Wirjanto, and Zhao (2020) proposed a conditional heteroscedasticity test for functional data. Kokoszka, Rice, and Shang (2017) proposed a portmanteau test for testing autocorrelation under a functional generalised autoregressive conditional heteroscedasticity model.

2.3.11. ARCH/GARCH models¹⁹

Volatility has been recognised as a primary measure of risks and uncertainties (Gneiting, 2011a, Markowitz, 1952, Sharpe, 1964, Taylor et al., 2009); for further discussion on uncertainty estimation, see Section 2.3.21. Estimating future volatility for measuring the uncertainty of forecasts is imperative for probabilistic forecasting. Yet, the right period in which to estimate future volatility has been controversial as volatility based on too long a period will make irrelevant the forecast horizon of our interests, whereas too short a period results in too much noise (Engle, 2004). An alternative to this issue is the dynamic volatility estimated through the autoregressive conditional heteroscedasticity (ARCH) proposed by Engle (1982), and the generalised autoregressive conditional heteroscedasticity (GARCH) model proposed by Bollerslev (1987). The ARCH model uses the weighted average of the past squared forecast error whereas the GARCH model generalises the ARCH model by further adopting past squared conditional volatilities. The GARCH model is the combination of (i) a constant volatility, which estimates the long-run average, (ii) the volatility forecast(s) in the last steps, and (iii) the new information collected in the last steps. The weightings on these components are typically estimated with maximum likelihood. The models assume a residual distribution allowing for producing density forecasts. One of the benefits of the GARCH model is that it can model heteroscedasticity, the volatility clustering characteristics of time series (Mandelbrot, 1963), a phenomenon common to many time series where uncertainties are predominant. Volatility clustering comes about as new information tends to arrive time clustered and a certain time interval is required for the time series to be stabilised as the new information is initially recognised as a shock.

The GARCH model has been extended in the diverse aspects of non-linearity, asymmetry and long memory. Among many such extensions, the Exponential GARCH

(EGARCH) model by [Nelson \(1991\)](#) uses log transformation to prevent negative variance; the Threshold GARCH (TGARCH) model by [Zakoian \(1994\)](#) allows for different responses on positive and negative shocks. A small piece of information can have more impact when the time series is under stress than under a stable time series ([Engle, 2004](#)). Another pattern often observed in the volatility time series is slowly decaying autocorrelation, also known as a long memory pattern, which [Baillie, Bollerslev, and Mikkelsen \(1996\)](#) capture using a slow hyperbolic rate of decay for the ARCH terms in the fractionally integrated GARCH (FIGARCH) model. Separately, in a further approach to directly estimating long term volatility, the GARCH-MIDAS (Mixed Data Sampling) model proposed by [Engle, Ghysels, and Sohn \(2013\)](#) decomposes the conditional volatility into the short-term volatility, as captured by the traditional GARCH, and the long-term volatility represented by the realised volatilities. The Heterogeneous Autoregressive (HAR) model by [Corsi \(2009\)](#) considers the log-realised volatility as a linear function of the log-realised volatility of yesterday, last week and last month to reflect traders' preferences on different horizons in the past. This model is extended by [Wilms, Rombouts, and Croux \(2021\)](#) to incorporate information about future stock market volatility by further including option-implied volatility. A different approach to volatility modelling, discussed in Section 2.3.14, is the use of low and high prices in the range-based volatility models.

The univariate GARCH models surveyed so far have been extended to multivariate versions, in order to model changes in the conditional covariance in multiple time series, resulting in such examples as the VEC ([Bollerslev, 1987](#)) and BEKK ([Engle & Kroner, 1995](#)), an acronym derived from Baba, Engle, Kraft, and Kroner. The VEC model, a direct generalisation of the univariate GARCH, requires more parameters in the covariance matrices and provides better fitness at the expense of higher estimation costs than the BEKK. The VEC model has to ensure the positivity of the covariance matrix with further constraints, whereas the BEKK model and its specific forms, e.g., factor models, avoid this positivity issue directly at the model specification stage. In an effort to further reduce the number of parameters to be estimated, the linear and non-linear combinations of the univariate GARCH models, such as the constant conditional correlation model of [Bollerslev \(1990\)](#) and the dynamic conditional correlation models of [Tse and Tsui \(2002\)](#) and of [Engle \(2002\)](#), were investigated.

2.3.12. Markov switching models²⁰

Since the late 1980s, especially in macroeconomics and finance, the applications of dynamic econometric modelling and forecasting techniques have increasingly relied on a special class of models that accommodate regime shifts, Markov switching (MS) models. The idea of MS is to relate the parameters of otherwise standard dynamic econometric frameworks (such as systems of regressions, vector autoregressions, and vector error corrections) to one or more unobserved state variables (see Section 2.3.6 for a definition), say, s_t , that can take K values and capture the notion of systems going through phases or “regimes”, which follow a simple, discrete stochastic process and are independent of the shocks of the model.

For instance, an otherwise standard AR(1) model can be extended to $AR(1|s_t)$, where all the parameters in the conditional mean as well as the variance of the shocks may be assumed to take different, estimable values as a function of s_t . Similarly, in a K -regime MS VAR(), the vector of intercepts and the autoregressive matrices may be assumed to depend on s_t . Moreover, the covariance matrix of the system shocks may be assumed to depend on some state variable, either the same as the mean parameters (s_t) or an additional, specific one (z_t), which may depend on lags of s_t . When a MS VAR model is extended to include exogenous regressors, we face a MS VARX, of which MS regressions are just a special case.

Even though multivariate MS models may suffer from issues of over-parameterisations that must be kept in check, their power of fitting complex non-linearities is unquestioned because, as discussed by [Marron and Wand \(1992\)](#), mixtures of normal distributions provide a flexible family that can be used to approximate many distributions. Moreover, MS models are known ([Timmermann, 2000](#)) to capture key features of many time series. For instance, differences in conditional means across regimes enter the higher moments such as variance, skewness, and kurtosis; differences in means in addition to differences in variances across regimes may generate persistence in levels and squared values of the series.

The mainstream literature (see, e.g., [Hamilton 1990](#), or the textbook treatments by [Kim, Charles, & Nelson 1999](#), and [Guidolin & Pedio 2018](#)) initially focused on time-homogeneous Markov chains (where the probabilities of the state transitions are constant). However, the finance and business cycles literatures ([Gray, 1996](#)) has moved towards time-heterogeneous MS models, in which the transition matrix of the regimes

may change over time, reacting to lagged values of the endogenous variables, to lagged exogenous variables, or to the lagged values of the state (in a self-exciting fashion). MS models may be estimated by maximum likelihood, although other estimation methods cannot be ruled out, like GMM (Lux, 2008). Typically, estimation and inference are based on the Expectation-Maximisation algorithm proposed by Dempster, Laird, and Rubin (1977), a filter that allows the iterative calculation of the one-step ahead forecast of the state vector given the information set and a simplified construction of the log-likelihood of the data. However, there is significant evidence of considerable advantages offered by Bayesian approaches based on Monte Carlo Markov chain techniques to estimating multivariate MS models (see, for example, Hahn, Frühwirth-Schnatter, & Sass, 2010).

Notably, MS models have been recently generalised in a number of directions, such as including regimes in conditional variance functions, for example of a GARCH or DCC type (see Pelletier, 2006 and Section 2.3.11).

2.3.13. Threshold models²¹

It is a well-known fact that financial and economic time series often display non-linear patterns, such as structural instability, which may appear in the form of recurrent regimes in model parameters. In the latter case, such instability is stochastic, it displays structure, and as such, it can be predicted. Accordingly, modelling economic and financial instability has become an essential goal for econometricians since the 1970s. One of the first and most popular models is the threshold autoregressive (TAR) model developed by Tong (1978). A TAR model is an autoregressive model for the time series in which the parameters are driven by a state variable (see Section 2.3.6 for a definition), which is itself a random variable taking distinct integer values (i.e., s). In turn, the value assumed by s_t depends on the value of the threshold variable z_t when compared to threshold levels, τ . For instance, if only two regimes exists, it is $s_t = 1$ if $z_t > \tau$ and otherwise. The threshold variable z_t can be exogenous or can be a lagged value of y_t . In the latter case, we speak of self-exciting threshold autoregressive (SETAR) models. Other choices of z_t include linear (Chen et al., 2003, Chen and So, 2006, Gerlach et al., 2006) or non-linear (Chen, 1995, Wu and Chen, 2007) combinations of the lagged dependent variable or of exogenous variables.

The TAR model has also been extended to account for different specifications of the conditional mean function leading to the development of the threshold moving

average (TMA – see, for example, [De Gooijer, 1998](#), [Ling et al., 2007](#), [Tong, 1990](#)) and the threshold autoregressive moving average (TARMA – see, for example, [Amendola et al., 2006](#), [Ling, 1999](#)) models. Those models are similar to the ones described in Section 2.3.4, but their parameters depend on the regime .

A criticism of TAR models is that they imply a conditional moment function that fails to be continuous. To address this issue, [Chan and Tong \(1986\)](#) proposed the smooth transition autoregressive (STAR) model. The main difference between TAR and STAR models is that, while a TAR imposes an abrupt shift from one regime to the others at any time that the threshold variable crosses above (or below) a certain level, a STAR model allows for gradual changes among regimes.

In its simplest form, a STAR is a two-regime model where the dependent variable is determined as the weighted average of two autoregressive (AR) models, i.e., where x_t is the transition variable and $F(\cdot)$ is some transformation of the transition variable . Regime probabilities are assigned through the transition function $F(\cdot)$, with $F(\cdot)$ being a cumulative density function of choice. The transition variable x_t can be the lagged endogenous variable, for [\(Teräsvirta, 1994\)](#), a (possibly non-linear) function of it, or an exogenous variable. The transition variable can also be a linear time trend t , which generates a model with smoothly changing parameters ([Lin & Teräsvirta, 1994](#)). Popular choices for the transition function are the logistic function (which gives rise to the LSTAR model) and the exponential function (ESTAR). Notably, the simple STAR model we have described can be generalised to have multiple regimes ([Van Dijk, Franses, & Lucas, 1999](#)).

Threshold models are also applied to modelling and forecasting volatility; for instance, the GJR-GARCH model of [Glosten, Jagannathan, and Runkle \(1993\)](#) can be interpreted as a special case of a threshold model. A few multivariate extensions of threshold models also exist, such as vector autoregressive threshold models, threshold error correction models ([Balke & Fomby, 1997](#)), and smooth transition vector error correction models ([Granger & Swanson, 1996](#)).

2.3.14. Low and high prices in volatility models²²

Volatility models of financial instruments are largely based solely on closing prices (see Section 2.3.11); meanwhile, daily low and high (LH) prices significantly increase the amount of information about the variation of returns during a day. LH prices were used for the construction of highly efficient estimators of variance, so called the range-based

(RB) estimators (e.g., [Fiszeder and Perczak, 2013](#), [Garman and Klass, 1980](#), [Magdon-Ismail and Atiya, 2003](#), [Parkinson, 1980](#), [Rogers and Satchell, 1991](#), [Yang and Zhang, 2000](#)). Recently, [Riedel \(2021\)](#) analysed how much additional information about LH reduces the time averaged variance in comparison to knowing only open and close. RB variance estimators, however, have a fundamental drawback, as they neglect the temporal dependence of returns (like conditional heteroscedasticity) and do not allow for the calculation of multi-period dynamic volatility forecasts.

In the last dozen or so years, numerous univariate dynamic volatility models have been constructed based on LH prices. Some of them were presented in the review paper of [Chou, Chou, and Liu \(2015\)](#). These models can be divided into four groups. The first one comprises simple models, used traditionally to describe returns, but they are based on the price range or on the mentioned earlier RB variance estimators. They include such models as random walk, moving average, exponentially weighted moving average (EWMA), autoregressive (AR), autoregressive moving average (ARMA; see [Section 2.3.4](#)), and heterogeneous autoregressive (HAR). The second group contains models which describe the conditional variance (or standard deviation) of returns. It comprises models like GARCH-PARK-R ([Mapa, 2003](#)), GARCH-TR ([Fiszeder, 2005](#)), REGARCH ([Brandt & Jones, 2006](#)), RGARCH ([Molnár, 2016](#)). The third group includes models which describe the conditional mean of the price range. It means that in order to forecast variance of returns the results have to be scaled. This group contains models like RB SV ([Alizadeh, Brandt, & Diebold, 2002](#)), CARR ([Chou, 2005](#)), TARR ([Chen, Gerlach, & Lin, 2008](#)), CARGPR ([Chan, Lam, Yu, Choy, & Chen, 2012](#)), STARR ([Lin, Chen, & Gerlach, 2012](#)), and MSRB ([Miao, Wu, & Su, 2013](#)). The last group is methodologically different because the estimation of model parameters is based on the sets of three prices, i.e., low, high and closing. This approach comprises the GARCH models ([Fiszeder and Perczak, 2016](#), [Lildholdt, 2002](#), [Venter et al., 2005](#)) and the SV model ([Horst, Rodriguez, Gzyl, & Molina, 2012](#)).

The development of multivariate models with LH prices has taken place in the last few years. They can be divided into three groups. The first one includes models, used traditionally to describe returns or prices, but they are based on the price range or RB variance estimators. They comprise such models like multivariate EWMA, VAR, HVAR, and vector error correction (VEC). It is a simple approach, however most models omit modelling the covariance of returns. The second group is formed by the multivariate RB

volatility models like RB-DCC (Chou, Wu, & Liu, 2009), DSTCC-CARR (Chou & Cai, 2009), RR-HGADCC (Asai, 2013), RB-MS-DCC (Su & Wu, 2014), DCC-RGARCH (Fiszeder, Fałdziński, & Molnár, 2019), RB-copula (Chiang and Wang, 2011, Wu and Liang, 2011). The third group includes the multivariate co-range volatility models like multivariate CARR (Fernandes, de Sá Mota, & Rocha, 2005), BEKK-HL (Fiszeder, 2018) and co-range DCC (Fiszeder & Fałdziński, 2019). These models apply LH prices directly not only for the construction of variances of returns but also for covariances. Section 3.3.9 discusses the use of the range-based volatility models in financial time series forecasting.

2.3.15. Forecasting with DSGE models²³

Dynamic Stochastic General Equilibrium (DSGE) models are the workhorse of modern macroeconomics employed by monetary and fiscal authorities to explain and forecast comovements of aggregate time series over the business cycle and to perform quantitative policy analysis. These models are studied in both academia and policy-making institutions (for details, see: Christiano et al., 2018, Del Negro and Schorfheide, 2013, Paccagnini, 2017). For example, the European Central Bank uses the New Area-Wide Model introduced by Warne, Coenen, and Christoffel (2010) and the Federal Reserve Board has created the Estimated, Dynamic, Optimisation-based model (FRB/EDO) as discussed in Chung, Kiley, and Laforte (2010). For an application on forecasting GDP and inflation, see Section 3.3.2. Developed as a response to Lucas (1976) critique of structural macroeconometrics models, DSGEs introduced microfoundations to describe business cycle fluctuations. Initially calibrated, estimated DSGEs have been employed in shocks identification and forecasting horseraces for the last 15 years. Estimation became possible thanks to computational progress and adoption of Bayesian techniques (for technical details, see: An and Schorfheide, 2007, Fernández-Villaverde and Guerrón-Quintana, 2020, Herbst and Schorfheide, 2016). Bayesian estimation allows for attributing prior distributions, instead of calibrating, and computing the posterior distribution for selected model parameters as well as drawing from predictive density. The Smets and Wouters (2007) DSGE is the most popular framework referred to in both research and policy literature. Proposed for the US economy, this medium-scale model is a closed economy composed of households, labor unions, a productive sector, and a monetary policy authority that sets the short-term interest rate according to a Taylor rule. These ingredients are

mathematically represented by a system of linear rational expectation equations. Using a solution algorithm (for example, [Blanchard and Kahn, 1980](#), [Sims, 2002](#)), researchers can write the model using the state-space representation composed by the transition equation and the measurement equation. The latter matches the observed data (in the Smets and Wouters: output growth rate, consumption, investment, wages, worked hours, inflation, and short-term interest rate) with the model latent variables. The solved model is employed for quantitative policy analysis and to predict and explain the behavior of macroeconomic and financial indicators.

DSGE models forecasting performance is investigated along two dimensions: point forecast and density forecast (see Sections [2.12.2 Point, interval, and pHDR forecast error measures](#), [2.12.4 Evaluating probabilistic forecasts](#) for discussions on their evaluation).

The point forecast is implemented by conducting both static and dynamic analysis, as described in [Cardani, Paccagnini, and Villa \(2019\)](#). If the static analysis provides a unique forecast value, the dynamic analysis describes the evolution of the prediction along the time dimension to investigate possible time-varying effects. Usually, point predictions are compared using the [Diebold and Mariano \(1995\)](#) and the [Clark and West \(2006\)](#) tests that compare predictions from two competing models. The accuracy of the static analysis is based mainly on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). MAE and RMSE are used to provide a relative forecasting evaluation compared to other competitors. Following [Clements and Hendry \(1998\)](#), [Kolasa, Rubaszek, and Skrzypczyński \(2012\)](#) apply the standard forecast unbiased test to assess if DSGEs are good forecasters in the absolute sense. The accuracy of the dynamic analysis is based on the Fluctuation Test (for some DSGE applications, see: [Boneva et al., 2019](#), [Cardani et al., 2019](#), [Giacomini and Rossi, 2016](#)). This test is based on the calculation of RMSEs that are assessed to investigate if the forecasting performance can be influenced by instabilities in the model parameters.

The density forecast is based on the uncertainty derived by the Bayesian estimation and it is commonly evaluated using the probability integral transform and the log predictive density scores (as main references, [Kolasa and Rubaszek, 2015a](#), [Wolters, 2015](#)). The statistical significance of these predictions is evaluated using the [Amisano and Giacomini \(2007\)](#) test that compares log predictive density scores from two competing models.

2.3.16. Robust equilibrium-correction forecasting devices²⁴

The use of equilibrium-correction models is ubiquitous in forecasting. [Hendry \(2010\)](#) notes that this class commonly includes models with explicit equilibrium-correction mechanisms such as vector equilibrium-correction models (VEqCM) as well as models with implicit equilibrium-correction (or long-run mean reversion) mechanisms such as vector auto-regressions (VARs; see Section [2.3.9](#)), dynamic factor models (DFMs), dynamic stochastic general-equilibrium (DSGE) models (see Section [2.3.15](#)), most models of the variance (see Section [2.3.11](#)), and almost all regression equations (see Sections [2.3.2 Time-series regression models](#), [2.3.4 Autoregressive integrated moving average \(ARIMA\) models](#)). This class of forecast model is prevalent across most disciplines. For example, [Pretis \(2020\)](#) illustrates that there is an equivalence between physical energy balance models, which are used to explain and predict the evolution of climate systems, and VEqCMs.

Despite their wide-spread use in economic modeling and forecasting, equilibrium-correction models often produce forecasts that exhibit large and systematic forecast errors. [Clements and Hendry, 1998](#), [Clements and Hendry, 1999](#) showed that forecasts from equilibrium-correction models are not robust to abrupt changes in the equilibrium. These types of regime changes are very common in macroeconomic time series (see [Hamilton, 2016](#) as well as Section [2.3.12](#)) and can cause the forecasts from many models to go off track. Therefore, if for example, there is a change in the equilibrium towards the end of the estimation sample, forecasts from this class of models will continue to converge back to the previous equilibrium.

In general, the forecasts from equilibrium-correction models can be robustified by estimating all model parameters over smaller or more flexible sub-samples. Several studies have proposed general procedures that allow for time-variation in the parameters; see, for example, [Giraitis et al., 2013](#), [Inoue et al., 2017](#) and [Pesaran, Pick, and Pranovich \(2013\)](#). This allows for an optimal or more adaptive selection of model estimation windows in order generate forecasts after structural breaks have occurred. An alternative approach for robustifying forecasts from equilibrium-correction models is to focus on the formulation and estimation of the equilibrium. [Hendry \(2006\)](#) shows that differencing the equilibrium-correction mechanism can improve the forecasts by removing aspects which are susceptible to shifts. However, differencing the equilibrium also induces additional forecast-error variance. [Castle, Fawcett, and Hendry](#)

(2010) show that it is beneficial to update the equilibrium or to incorporate the underlying structural break process. Alternatively, [Castle, Clements, and Hendry \(2015a\)](#) show that there can be large forecast gains from smoothing over estimates of the transformed equilibrium. Building on this, [Martinez, Castle, and Hendry \(2021\)](#) show that there are many possible transformations of the equilibrium that can improve the forecasts. Several of these transformations imply that the equilibrium-correction model collapses to different naive forecast devices whose forecasts are often difficult to beat. By replacing the equilibrium with smooth estimates of these transformations, it is possible to outperform the naive forecasts at both short and long forecast horizons while retaining the underlying economic theory embedded within the equilibrium-correction model. Thus, it is possible to dramatically improve forecasts from equilibrium-correction models using targeted transformations of the estimated equilibrium so that it is less susceptible to the shifts which are so damaging to the model forecasts.

2.3.17. Forecasting with data subject to revision²⁵

When a forecast is made today of the future value of a variable, the forecast is necessarily ‘real time’ – only information available at the time the forecast is made can be used. The forecasting ability of a model can be evaluated by mimicking this setup – generating forecasts over some past period (so outcomes known) only using data known at each forecast origin. As noted by [Clements and Hendry \(2005\)](#), out-of-sample forecast performance is the gold standard. Sometimes the analysis is pseudo real time. At a given forecast origin, forecasts are constructed only using data up to the period, but the data are taken from the latest-available vintage at the time the study is undertaken. Using revised data to estimate the forecasting model – instead of the data available at the time the forecast was made – may exaggerate forecast performance, and present a misleading picture of how well the model might perform in real time. The improved availability of real-time databases has facilitated proper real-time studies.²⁶ At time the data are taken from the vintage available at time. Data revisions are often important, and occur because statistical agencies strive to provide timely estimates which are based on incomplete source data (see, for example, [Fixler and Grimm, 2005](#), [Fixler and Grimm, 2008](#), [Zwijnenburg, 2015](#)).

There are a number of possible real-time approaches. The conventional approach is to estimate the forecasting model using the latest vintage of data available at time.

Suppose the vintage- contains data for time periods up to , denoted . The observation for time is a first estimate, for a second estimate, and so on, such that data for earlier periods will have been revised many times. Hence the model will be estimated on data of different maturities, much of which will have been revised many times. But the forecast will typically be generated by feeding into the model ‘lightly-revised’ data for the most recent time periods. The accuracy of the resulting forecasts can be improved upon (in principle) by taking into account data revisions (see, for example, [Clements and Galvão, 2013b](#), [Kishor and Koenig, 2012](#), [Koenig et al., 2003](#)). In the following two paragraphs, we consider alternative real-time approaches which solve the problem of estimating the model on mainly revised data, and feeding in mainly unrevised forecast origin data. [Koenig et al. \(2003\)](#) suggest using real-time-vintage (RTV) data to estimate the model. The idea is to use early estimates of the data to estimate the model, so that the model is estimated on ‘lightly-revised’ data that matches the maturity of the forecast-origin data that the forecast is conditioned on.

Other approaches seek to model the data revisions process along with the fully-revised true values of the data, as in [Cunningham et al., 2009](#), [Jacobs and van Norden, 2011](#) and [Kishor and Koenig \(2012\)](#). Reduced form models that avoid the necessity of estimating unobserved components have adapted the vector autoregression (VAR; see also Section 2.3.9) of [Sims \(1980\)](#) to jointly model different observed vintages of data. Following [Patterson \(1995\)](#), [Garratt, Lee, Mise, and Shields \(2008\)](#) work in terms of the level of the log of the variable, , and model the vector given by . [Carriero, Clements, and Galvão \(2015\)](#) and [Clements and Galvão, 2012](#), [Clements and Galvão, 2013a](#) minimise the effects of benchmark revisions and re-basing by modelling ‘same-vintage-growth rates’, namely , where , and denotes the greatest data maturity.

[Galvão \(2017\)](#) shows how forecasts of fully-revised data can be generated for dynamic stochastic general equilibrium (DSGE; Section 2.3.15) models (for example, [Del Negro & Schorfheide, 2013](#)), by applying the approach of [Kishor and Koenig \(2012\)](#). [Clements \(2017\)](#) argues that improvements in forecast accuracy might be expected to be greater for interval or density forecasts than point forecasts, and this is further explored by [Clements and Galvão \(2017\)](#).

Surveys on data revisions and real-time analysis, including forecasting, are provided by [Clements and Galvão \(2019\)](#) and [Croushore, 2006](#), [Croushore, 2011a](#), [Croushore, 2011b](#); see also Section 3.3.1.

2.3.18. Innovation diffusion models²⁷

Forecasting the diffusion of innovations is a broad field of research, and influential reviews on the topic have highlighted its importance in many disciplines for strategic or anticipative reasons (Mahajan et al., 1990, Meade and Islam, 2006, Peres et al., 2010). Large-scale and fast diffusion processes of different nature, ranging from the spread of new epidemics to the adoption of new technologies and products, from the fast diffusion of news to the wide acceptance of new trends and social norms, are demanding a strong effort in terms of forecasting and control, in order to manage their impact into socio-economic, technological and ecological systems.

The formal representation of diffusion processes is often based on epidemic models, under the hypothesis that an innovation spreads in a social system through communication among people just like an epidemics does through contagion. The simplest example is represented by the (cumulative) logistic equation that describes a pure epidemic process in a homogeneously mixing population (Verhulst, 1838). The most famous and employed evolution of the logistic equation is the Bass model (Bass, 1969), developed in the field of quantitative marketing and soon become a major reference, due to its simple and powerful structure.

The Bass model (BM) describes the life-cycle of an innovation, depicting its characterising phases of launch, growth/maturity, and decline, as result of the purchase decisions of a given cohort of potential adopters. Mathematically, the model is represented by a first order differential equation, describing a diffusion process by means of three parameters: the maximum market potential, M , assumed to be constant along the whole diffusion process, and parameters α and β , referring respectively to two distinct categories of consumers, the *innovators*, identified with parameter α , adopting first, and the *imitators*, adopting at a later stage by imitating others' behaviour and thus responsible for *word-of-mouth* dynamics. In strategic terms, crucial forecasts are referred to the point of maximum growth of the life cycle, the *peak*, and the point of market saturation. For a general description of new product forecasting please refer to Section 3.2.6.

Innovation diffusion models may be also used for *post-hoc* explanations, helping understand the evolution of a specific market and its response to different external factors. Indeed, one of the most appealing characteristics of this class of models is the possibility to give a simple and nice interpretation to all the parameters involved. In this

perspective, a valuable generalisation of the BM was proposed in [Bass et al. \(1994\)](#) with the Generalised Bass Model (GBM). The GBM enlarges the BM by multiplying its hazard rate by a very general intervention function, assumed to be non-negative, which may account for exogenous shocks able to change the temporal dynamics of the diffusion process, like marketing strategies, incentive mechanisms, change in prices and policy measures.

Another generalisation of the BM and the GBM, relaxing the assumption of a constant market potential was proposed in [Guseo and Guidolin \(2009\)](#) with the GGM. This model postulates a time-dependent market potential, which is function of the spread of knowledge about the innovation, and thus assumes that a diffusion process is characterised by two separate phases, information and adoption. The GGM allows a significant improvement in forecasting over the simpler BM, especially through a more efficient description of the first part of the time series, often characterised by a slowdown pattern, as noticed by [Guseo and Guidolin \(2011\)](#).

Other generalisations of innovation diffusion models, considering competition between products, are treated in Section [2.3.20](#). Applications of innovation diffusion models are presented in Sections [3.2.6 New product forecasting](#), [3.4.5 Forecasting renewable energy technologies](#).

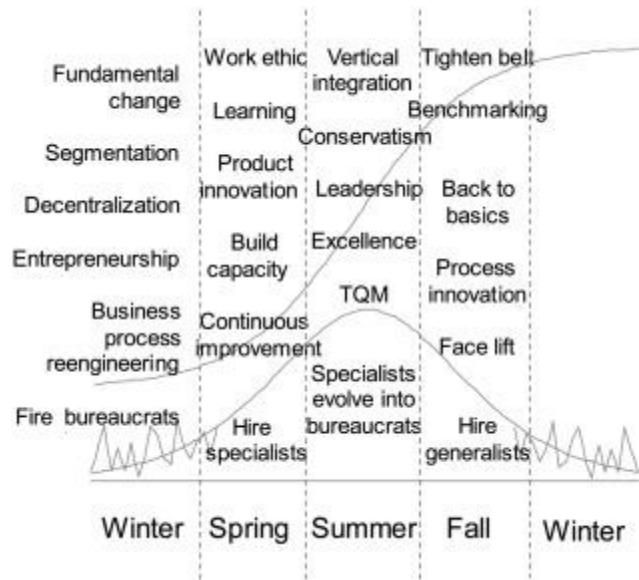
2.3.19. The natural law of growth in competition²⁸

As early as in 1925 Alfred J. Lotka demonstrated that manmade products diffuse in society along S-shaped patterns similar to those of the populations of biological organisms ([Lotka, 1925](#)). Since then S curve logistic descriptions have made their appearance in a wide range of applications from biology, epidemiology and ecology to industry, competitive substitutions, art, personal achievement and others ([Fisher and Pry, 1971](#), [Marchetti, 1983](#), [Meade, 1984](#), [Modis, 1992](#)). The reader is also referred to Sections [2.3.18 Innovation diffusion models](#), [3.4.5 Forecasting renewable energy technologies](#). In fact, logistic growth can be detected whenever there is growth in competition, and competition can be generalised to a high level of abstraction, e.g. diseases competing for victims and all possible accidents competing for the chance to be materialised.

S-curves enter as modular components in many intricate natural patterns. One may find S curves inside S curves because logistics portray a fractal aspect; a large S curve can be decomposed in a cascade of smaller ones ([Modis, 1994](#)). One may also find chaos by

rendering the logistic equation discrete (Modis & Debecker, 1992). Finally, logistics sit in the heart of the Lotka–Volterra equations, which describe the predator–prey relations and other forms of competition. In its full generality, the logistic equation, in a discrete form, with cross terms to account for all interrelations between competing species, would give a complete picture in which growth in competition, chaos, self-organisation, complex adaptive systems, autopoiesis, and other such academic formulations, all ensue as special cases (Modis, 1997).

Each S curve has its own life cycle, undergoing good and bad “seasons” (see Fig. 1). A large set of behaviours have been tabulated, each one best suited for a particular season (Modis, 1998). Becoming conservative – seeking no change – is appropriate in the summer when things work well. But excellence drops in second place during the difficult times of winter – characterised by chaotic fluctuations – when fundamental change must take place. Learning and investing are appropriate for spring, but teaching, tightening the belt, and sowing the seeds for the next season’s crop belong in the fall. Focusing on *what* to do is appropriate in spring, whereas in fall the emphasis shifts to the *how*. For example, the evolution of classical music followed a large-timeframe S curve beginning in the fifteenth century and reaching a ceiling in the twentieth century; see Fig. 2 (Modis, 2013). In Bach’s time composers were concerned with *what* to say. The value of their music is in its architecture and as a consequence it can be interpreted by any instrument, even by simple whistling. But two hundred years later composers such as Debussy wrote music that depends crucially on the interpretation, the *how*. Classical music was still “young” in Bach’s time but was getting “old” by Debussy’s time. No wonder Chopin is more popular than Bartók. Chopin composed during the “summer” of music’s S curve when public preoccupation with music grew fastest. Around that time composers were rewarded more handsomely than today. The innovations they made in music – excursions above the curve – were assimilated by the public within a short period of time because the curve rose steeply and would rapidly catch up with each excursion/innovation. But today the curve has flattened and composers are given limited space. If they make an innovation and find themselves above the curve, there won’t be any time in the future when the public will appreciate their work; see Fig. 3 (Modis, 2007). On the other hand, if they don’t innovate, they will not be saying anything new. In either case today’s composers will not be credited with an achievement.



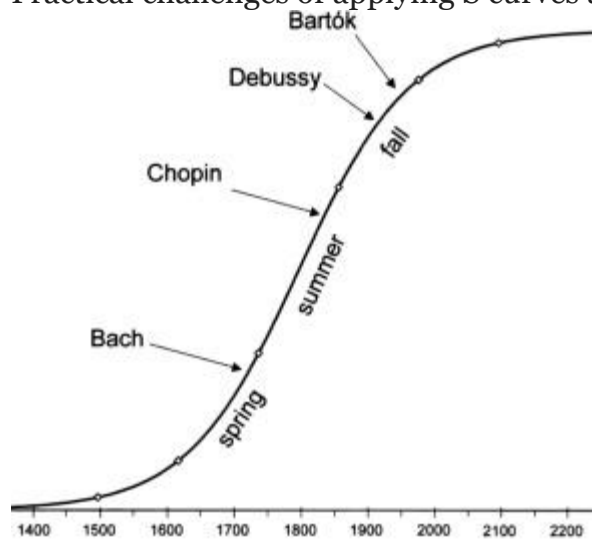
1. [Download: Download high-res image \(247KB\)](#)
2. [Download: Download full-size image](#)

Fig. 1. Typical attributes of a growth cycle's "seasons".

Adopted from [Modis \(1998\)](#) with the permission from the author.

S curves constructed only qualitatively can be accurate, informative, and insightful.

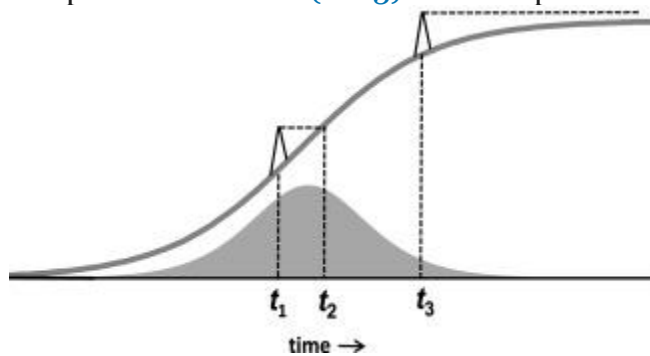
Practical challenges of applying S curves are discussed in Section [3.8.12](#).



1. [Download: Download high-res image \(73KB\)](#)
2. [Download: Download full-size image](#)

Fig. 2. The evolution of classical music. The vertical axis could be something like "importance", "public appreciation", or "public preoccupation with music" (always cumulative).

Adopted from [Modis \(2013\)](#) with the permission from the author.



1. [Download: Download high-res image \(53KB\)](#)
2. [Download: Download full-size image](#)

Fig. 3. An upward excursion at t_3 reaches the same level as the logistic curve at t_2 and can be considered as a “natural” deviation. The same-size excursion at time t_1 has no corresponding point on the curve. The grey life cycle delimits the position and size of all “natural” deviations.

Adapted from [Modis \(2007\)](#) with permission from the author.

2.3.20. Synchronic and diachronic competition²⁹

Synchronic and diachronic competition models account for critical life cycle perturbations due to interactions, not captured by univariate innovation diffusion models (see Section 2.3.18) or other time series models, such as ARIMA and VAR. This is important especially in medium and long-term prediction.

Competition in a natural or socio-economic system generally refers to the presence of antagonists that contend for the same resource. This typically occurs in nature, where multiple species struggle with each other to survive, or in socio-economic contexts where products, technologies and ideas concur to be finally accepted within a market and compete for the same market potential. These competition dynamics are reflected in separate time series – one for each concurrent – characterised by a locally oscillatory behaviour with nonlinear trends, unexpected deviations and saturating effects.

The analytic representation of competition has followed different approaches. A first approach has been based on complex systems analysis ([Boccaro, 2004](#)), which refers to a class of agents (see also Section 2.7.3) that interact, through local transition rules, and produce competition as an emergent behaviour. This approach may be frequently reduced to a system of differential equations, with suitable mean field approximations.

A second approach, systems analysis, has been based on systems of ordinary differential equations (ODE).

In this domain, competition may be a synchronic process, if competitors are present in the environment at the same time; for example, two products may enter the market in the same period. Instead, it is diachronic if competitors come at a later stage; for example, one product enters a market in a given period and just subsequently other similar products enter the same market and start to compete. Pioneering contributions of this competition modelling are due to [Lotka \(1920\)](#) and [Volterra \(1926a\)](#), who independently obtained a class of synchronic predator–prey models; see also Section [2.3.19](#). A generalised version of the Lotka–Volterra (LV) model has been provided by [Abramson and Zanette \(1998\)](#).

[Morris and Pratt \(2003\)](#) proposed an extended LV model for a duopolistic situation, by making explicit the role of carrying capacities or market potentials, and the inhibiting strength of the competitors in accessing the residual resource. LV models typically do not have closed form solutions. In such cases, a stacked form of equations allows a first-stage inference based on nonlinear least squares (NLS), with no strong assumptions on the stochastic distributions of error component. Short-term refining may be grounded on a Seasonal Autoregressive Moving Average with exogenous input (SARMAX) representation. Outstanding forecasts may be obtained including the estimated first-stage solution as ‘exogenous’ input (see Sections [2.2.3 Anomaly detection and time series forecasting](#), [2.2.5 Exogenous variables and feature engineering](#)).

A different synchronic model, termed Givon-Bass (GB) model, extending the univariate innovation diffusion models described in Section [2.3.18](#) ([Bass, 1969](#), [Givon et al., 1995](#)), has been presented in [Bonaldo \(1991\)](#), introducing parametric components of global interaction. In this model, the residual market (or carrying capacity) is completely accessible to all competitors, and the rate equations introduce distributed seeding effects. The GB model has a closed form solution that was independently published by [Krishnan, Bass, and Kummar \(2000\)](#). The more general model by [Savin and Terwiesch \(2005\)](#) and related advances by [Guseo and Mortarino \(2010\)](#) were extended to the diachronic case in [Guseo and Mortarino \(2012\)](#), defining a competition and regime change diachronic (CRCD) model with a closed form solution. A relevant improvement of CRCD has been proposed in [Guseo and Mortarino \(2014\)](#), by introducing within-brand and cross-brand word-of-mouth effects, not present in

standard LV models. The unrestricted unbalanced diachronic competition (unrestricted UCRC) model, is defined with these new factors. The model assumes, among other specifications, a constant market potential. In [Guseo and Mortarino \(2015\)](#) this assumption is relaxed, by introducing a dynamic market potential ([Guseo and Guidolin, 2009](#), [Guseo and Guidolin, 2011](#)). Some applications are summarised in Section 3.8.8.

2.3.21. Estimation and representation of uncertainty³⁰

Forecasting uncertainty consists in estimating the possible range of forecast errors (or true values) in the future and the most widely adopted representation is a forecast interval ([Patel, 1989](#)). The forecast interval indicates a range of values and the respective probability, which is likely to contain the true value (which is yet to be observed) of the response variable. Since for a specific lead-time the forecast interval only encompasses information from a marginal distribution, it can also be named marginal forecast interval (MFI). A MFI can be obtained from: parametric distribution, such as a Gaussian distribution with conditional mean and variance estimated with a Generalised ARCH model ([Baillie & Bollerslev, 1992](#)); non-parametric distribution, e.g., obtained with conditional kernel density estimation ([Hyndman, Bashtannyk, & Grunwald, 1996](#)); directly estimated with statistical learning methods, such as quantile regression ([Taylor & Bunn, 1999](#)) or bootstrapping ([Masarotto, 1990](#)), or with machine learning algorithms like quantile random forests ([Meinshausen, 2006](#)). For a combination of density forecasts from different models, see Section 2.6.2.

For multi-step ahead forecasting problems (see also Section 2.7.7), in particular when information about forecast uncertainty is integrated in multi-period stochastic optimisation ([Dantzig & Infanger, 1993](#)), information about the temporal dependency structure of forecast errors (or uncertainty) is a fundamental requirement. In this case, the concept of simultaneous forecast intervals (SFI) can be found in the statistical literature ([Chew, 1968](#)). SFI differ from MFI since take into account the temporal interdependency of forecast errors and are constructed to have the observed temporal trajectory of the response variable fully contained inside the forecast intervals during all lead-times of the time horizon. The number of works that cover SFI is lower when compared to the MFI, but some examples are: methods based on Bonferroni- and product-type inequalities applied time series forecasting models like ARIMA and Holt-Winters ([Ravishanker, yen Wu, & Glaz, 1991](#)); combination of bootstrap replications and an heuristic optimisation procedure to find an envelope of the temporal trajectories

closest to the deterministic forecast (Staszewska-Bystrova, 2011 see also Section 2.7.5); sampling forecast errors at different horizons and estimate the SFI with the empirical Mahalanobis distance (Jordá, Knüppelc, & Marcellino, 2013).

Advances in Operations Research for decision-making problems under uncertainty imposed new requirements in forecast uncertainty estimation and representation. On one hand, stochastic optimisation requires a scenario representation for forecast uncertainty (Powell, 2019). This motivated research in methods that generate uncertainty forecasts represented by random vectors (term used in statistics) or path forecasts (term used in econometrics), such as parametric copula combined with MFI (Pinson, Madsen, Nielsen, Papaefthymiou, & Klöckl, 2009), parametric dynamic stochastic model (Li & Chan, 2011) or epi-spline basis functions (Rios, Wets, & Woodruff, 2015). On the other hand, robust optimisation does not make specific assumptions on probability distributions and the uncertain parameters are assumed to belong to a deterministic uncertainty set. Hence, some authors proposed new methods to shape forecast uncertainty as polyhedral or ellipsoidal regions to enable a direct integration of forecasts in this type of optimisation problem (Bertsimas and Pachamanova, 2008, Golestaneh et al., 2019).

Finally, communication of forecast uncertainty (e.g., MFI, SFI, random vectors) to decision-makers requires further attention since it remains as a major bottleneck for a wide adoption by industry, particularly in uses cases with multivariate time series (Akram, Binning, & Maih, 2015) and adverse weather events (Ramos, Mathevet, Thielen, & Pappenberger, 2010). Please also see Section 3.7.5.

2.3.22. Forecasting under fat tails³¹

A non-negative continuous random variable is fat-tailed, if its survival function is regularly varying, that is to say if $\bar{F}(x) \sim L(x)x^{-\alpha}$, where L is a slowly varying function, for which $\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$ for (Embrechts, Klüppelberg, & Mikosch, 2013). The parameter α is known as the tail parameter, and it governs the thickness of the tail – the smaller α the fatter the tail – and the existence of moments, so that $E[X^k] < \infty$ if and only if $k < \alpha$. Often α is re-parametrised as $\frac{1}{\beta}$. Fat tails are omnipresent in nature, from earthquakes to floods, and they are particularly common in human-related phenomena like financial markets, insurance, pandemics, and wars (see, for example, Mandelbrot, 1983, Taleb, 2020 and references therein).

Forecasting fat-tailed random variables is therefore pivotal in many fields of life and science. However, while a basic coverage of the topic is available in most time series and risk management manuals (e.g., [McNeil et al., 2015](#), [Shumway and Stoffer, 2017](#)), the profound implications of fat tails are rarely taken into consideration, and this can generate substantial errors in forecasts.

As observed in [Taleb, Bar-Yam, and Cirillo \(2020\)](#), any statistical forecasting activity about the mean – or another quantity – of a phenomenon needs the law of large numbers (LLN), which guarantees the convergence of the sample mean at a given known rate, when the number of observations grows.

Fat-tailed phenomena with tail parameter $\alpha < 2$ are trivially not predictable. Since their theoretical mean is not defined, the LLN does not work, for there is nothing the sample mean can converge to. This also applies to apparently infinite-mean phenomena, like pandemics and wars, i.e., extremely tricky objects of study, as discussed in [Cirillo and Taleb \(2016a\)](#). In similar situations, one can rely on extreme value theory to understand tail risk properties, but should refrain from standard forecasting.

For random variables with $\alpha > 2$, the LLN can be extremely slow, and an often unavailable number of observations is needed to produce reliable forecasts. Even for a well-behaved and non-erratic phenomenon, we all agree that a claim about the fitness or non-fitness of a forecasting approach, just on the basis of one single observation ($n=1$), would be considered unscientific. The fact is that with fat-tailed variables that “ $n=1$ ” problem can be made with $n > 1$ observations ([Embrechts et al., 2013](#), [Taleb, 2020](#)). In the case of events like operational losses, even a larger n can still be just anecdotal ([Cirillo & Taleb, 2016a](#)). According to [Taleb \(2020\)](#), owing to preasymptotics, a conservative heuristic is to manage variables with $\alpha < 2$ as practically unpredictable (for example, see Sections [3.6.2 Epidemics and pandemics](#), [3.6.6 Forecasting risk for violence and wars](#)). Their sample average is indeed too unstable and needs too many observations for forecasts to be reliable in a reasonable period of time. For $\alpha > 2$, conversely, forecasting can take place, and the higher α the better. In any case, more research is strongly needed and desirable (see also Section [4](#)).

Observe that even discussing the optimality of any alarm system (see, for example, [Svensson et al., 1996](#), [Turkman and Turkman, 1990](#)) based on average forecasts would prove meaningless under extremely fat tails ($\alpha < 2$), when the LLN works very slowly or does not work. In fact, even when the expected value is well-defined

(i.e.,), the non-existence of the variance would affect all the relevant quantities for the verification of optimality (De Mare, 1980), like for instance the chance of undetected events. For all these quantities, the simple sample estimates commonly used would indeed be misleading.

2.4. Bayesian forecasting

2.4.1. Foundations of Bayesian forecasting³²

The Bayesian approach to forecasting produces, by default, a probabilistic forecast (see also Section 2.6.2 and Section 2.12.4) describing the uncertainty about future values of the phenomenon of interest, conditional on all *known quantities*; and with uncertainty regarding all *unknown quantities* having been integrated out. In order to produce such forecasts, the Bayesian approach requires (i) a predictive model for the future value of the relevant phenomenon, conditional on the observed data and all model unknowns; (ii) a model for the observed data; and (iii) a distribution describing the (subjective or objective) prior beliefs regarding the model unknowns. Using these quantities, the standard calculus of probability distributions, and Bayes' theorem, then yield the Bayesian predictive (equivalently, forecast) density function (where density is used without loss of generality).

Stated more formally, given observed data up to time t , \mathcal{D}_t , and denoting the model unknowns by θ , Bayesian forecasting describes the behaviour of the future random variable y_{t+1} via the predictive density: (2) where y denotes a value in the support of y_{t+1} and $f(y|\mathcal{D}_t, \theta)$ is the predictive model for y_{t+1} conditional on \mathcal{D}_t and the model unknowns θ . Critically, and in contrast with frequentist approaches to forecasting, parameter uncertainty has been factored into $f(y|\mathcal{D}_t, \theta)$ via the process of integration with respect to the posterior probability density function (pdf) for θ , $\pi(\theta|\mathcal{D}_t)$. The posterior pdf is given, by Bayes' theorem, as $\pi(\theta|\mathcal{D}_t) \propto \pi(\theta) L(\mathcal{D}_t|\theta)$, where $L(\mathcal{D}_t|\theta)$ defines the assumed model for y_{t+1} (equivalently, the likelihood function), and the prior pdf $\pi(\theta)$ captures prior beliefs about θ . Moreover, uncertainty about the assumed predictive model itself can be easily accommodated using Bayesian model averaging, which involves taking a weighted average of *model-specific* predictives, with posterior model probabilities (also obtained via Bayes' theorem) serving as the weights. See Greenberg (2008), Koop (2003) and O'Hagan and Forster (2004) for textbook illustrations of all of these steps.

No matter what the data type, the form of predictive model, or the dimension of the unknowns, the basic manner in which all Bayesian forecast problems are framed is the same. What differs however, from problem to problem, is the way in which the forecasting problem is *solved*. To understand why, it is sufficient to recognise that in order to obtain the predictive density we must be able to (somehow) perform the integration that defines this quantity. In almost any practical setting, this integration is infeasible analytically and we must rely on *computational methods* to access the predictive density. Therefore, the evolution of the practice of Bayesian forecasting has gone hand in hand with developments in Bayesian computation ([Martin, Frazier, & Robert, 2020](#)). Through the lens of computation, in Section 2.4.2 we briefly describe the methods of implementing Bayesian forecasting.

2.4.2. Implementation of Bayesian forecasting³³

If the posterior is accessible via methods of *exact simulation* – e.g., Monte Carlo simulation, importance sampling, Markov chain Monte Carlo (MCMC) sampling, pseudo-marginal MCMC ([Andrieu et al., 2011](#), [Andrieu and Roberts, 2009](#)) – an estimate of the predictive density in (2) can be produced using draws of the unknown from the posterior pdf, . In most cases, this simulation-based estimate of can be rendered arbitrarily accurate by choosing a very large number of posterior draws; hence the use of the term ‘exact predictive’ to reference this estimate. See [Geweke and Whiteman \(2006\)](#) for an early review of Bayesian forecasting implemented using exact simulation, Sections 2.3.15 Forecasting with DSGE models, 2.4.3 Bayesian forecasting with copulas, 2.5.3 Variable selection, 3.3.2 Forecasting GDP and inflation, 3.3.10 Copula forecasting with multivariate dependent financial times series for further discussions and a range of relevant applications, and Chapters 3, 7, and 9 in [Geweke, Koop, and van Dijk \(2011\)](#) for applications in (general) state space, macroeconometric and finance settings respectively. In addition, a 2008 special issue of *International Journal of Forecasting* on Bayesian Forecasting in Economics provides coverage of forecasting applications (and methods of computation) that exploit exact simulation methods, as do selected chapters in [Brooks, Gelman, Jones, and Meng \(2011\)](#) and [O’Hagan and West \(2010\)](#).

In cases where the posterior is not readily accessible, due to either the intractability of the likelihood function or the high dimension of the model unknowns, or both, methods of *approximation* are required. [Frazier et al. \(2019\)](#), for instance, produce an

‘approximate predictive’ by replacing the exact posterior in (2), , with an ‘approximate posterior’ constructed using *approximate Bayesian computation* (ABC – [Sisson, Fan, & Beaumont, 2019](#)). In large samples, this approximate predictive is shown to be equivalent to the exact predictive. The approximate and exact predictives are also shown to be numerically indistinguishable in finite samples, in the cases investigated; see also [Canale and Ruggiero \(2016\)](#), and [Kon Kam King, Canale, and Ruggiero \(2019\)](#). Related work produces an approximate predictive by exploiting *variational Bayes* ([Blei, Kucukelbir, & McAuliffe, 2017](#)) approximations of the posterior ([Chan and Yu, 2020](#), [Koop and Korobilis, 2018](#), [Loaiza-Maya, Smith et al., 2020](#), [Quiroz et al., 2018](#), [Tran et al., 2017](#)). The flavour of this work is broadly similar to that of [Frazier et al. \(2019\)](#); that is, computing the predictive via an approximation to the posterior does not *significantly* reduce predictive accuracy.

The Bayesian paradigm thus provides a very natural and coherent approach to prediction that can be implemented successfully via one of any number of computational methods. Inherent to the approach, however, is the assumption that the model we are using to make predictions is an accurate description of the data generating process (DGP) that has generated the observed data; or if we are averaging across models using Bayesian model averaging, we must assume that the assumed set of models spans the DGP for the observed data. In response to this limitation, allied with the desire to drive prediction by user-specified measures of predictive loss, new approaches to Bayesian prediction have recently been proposed, and we briefly discuss two such classes of methods.

First are methods for combining predictives in which the weights are not equated to the posterior model probabilities (as in standard Bayesian model averaging) but, rather, are updated via problem-specific predictive criteria, or via predictive calibration ([Dawid, 1982](#), [Dawid, 1985](#), [Gneiting et al., 2007](#)); see [Bassetti et al., 2018](#), [Basturk et al., 2019](#), [Billio et al., 2013](#), [Casarin et al., 2015](#), [McAlinn et al., 2020](#), [McAlinn and West, 2019](#), and [Pettenuzzo and Ravazzolo \(2016\)](#), for a selection of approaches, and [Section 2.6.2](#) for related discussion. Importantly, these methods do not assume that the true model is spanned by the constituent model set. Second are methods in which the standard Bayesian posterior, which is itself based on a potentially misspecified model, is replaced by a generalised version that is designed for the specific predictive task at hand (e.g., accurate prediction of extreme values); with the overall goal then being to produce

predictions that are accurate according to the particular measure of interest. See [Frazier et al., 2021](#), [Loaiza-Maya, Martin et al., 2020](#) and [Syring and Martin \(2020\)](#), for specific examples of this methodology, as well as proofs of its theoretical validity.

2.4.3. Bayesian forecasting with copulas³⁴

Copulas provide intrinsic multivariate distribution structure that allows for modelling multivariate dependence with marginal distributions as the input, making it possible to forecast dependent time series and time series dependence. This review focuses on the Bayesian approach for copula forecasting. For rigorous topics on copula introduction ([Joe, 1997](#), [Nelsen, 2006](#), [Trivedi and Zimmer, 2007](#)), copula modelling techniques ([Durante & Sempì, 2015](#)), vine copulas ([Joe, 2014](#)), review on frequentist approaches for copula-based forecasting ([Patton, 2013](#)), see the aforementioned references and the references therein.

The advantages of the Bayesian copula approach compared to the frequentist treatments are (i) the Bayesian approach allows for jointly modelling the marginal models and the copula parameters, which improves the forecasting efficiency ([Joe, 2005](#), [Li and Kang, 2018](#)), (ii) probabilistic forecasting is naturally implied with the Bayesian predictive density, and (iii) experts information can be seamlessly integrated into forecasting via the priors' setting.

Forecasting with copulas involves selecting between a different class of copulas. Common approaches include Bayesian hypothesis testing (see Section 2.4.1) where copula parameters are treated as nuisance variables ([Huard, Évin, & Favre, 2006](#)), or parsimonious modelling of covariance structures using Bayesian selection and model averaging ([Min and Czado, 2011](#), [Pitt et al., 2006](#), [Smith, 2010](#)).

One particular interest in copulas is to forecast the dynamic dependencies between multivariate time series. Time-varying copula construction is possible via (i) an autoregressive or GARCH form (see Section 2.3.11) of dependence parameters ([Lucas et al., 2014](#), [Patton, 2006](#)), (ii) factor copula construction ([Oh and Patton, 2018](#), [Tan et al., 2019](#)) that simplifies the computation, (iii) the stochastic copula autoregressive model ([Almeida & Czado, 2012](#)) that dependence is modelled by a real-valued latent variable, and (iv) covariate-dependent copulas approach by parameterising the dependence as a function of possible time-dependent covariates ([Li & Kang, 2018](#)) that also improves the forecasting interpretability. ARMA-like and GARCH-like dependences in the tail can be considered as special cases of [Li and Kang \(2018\)](#).

In multivariate time series forecasting (see also Section 2.3.9), unequal length of data is a common issue. One possible approach is to partition the copula parameters into elements relating only to the marginal distributions and elements only relating to the copula (Patton, 2006a). For mixed frequency data, it is possible to decompose the copula dependence structure into linear and nonlinear components. Then the high and low frequency data is used to model the linear and nonlinear dependencies, respectively (Oh & Patton, 2016). Bayesian data augmentation is also used to forecast multivariate time series with mixed discrete and continuous margins (Smith & Khaled, 2012). For other treatments for discrete and continuous time series (see, for example, Panagiotelis et al., 2012, Panagiotelis et al., 2017).

The Bayesian approach for lower dimensional copula forecasting () is straightforward with traditional Gaussian copulas, Student's- copulas, Archimedean copulas, or pair copula combinations. In higher dimensional settings, special considerations are required to save the computational burden, such as low rank approximation of covariates matrix (Salinas, Michael, Laurent, Roberto & Jan, 2019) or factor copula models with stochastic loadings (Creal & Tsay, 2015).

In the Bayesian setup, forecasting model performance is typically evaluated based on a - fold out-of-sample log predictive score (LPS: Geweke & Amisano, 2010), and out-of-sample Value-at-Risk (VaR) or Expected Shortfall (ES) are particularly used in financial applications. The LPS is an overall forecasting evaluation tool based on predictive densities, serving out-of-sample probabilistic forecasting. LPS is ideal for decision makers (Geweke, 2001, Geweke and Amisano, 2010). The VaR gives the percentile of the conditional distribution, and the corresponding ES is the expected value of response variable conditional on it lying below its VaR.

2.5. Variable and model selection

2.5.1. Leading indicators and Granger causality³⁵

Leading (economic) indicators are variables that try to capture changes in the development of economic activity before those changes materialise. Typically, market participants, policy makers, the general public, etc. are interested in which direction the economy of a country is developing ahead of the publication date of the respective quarterly GDP figures (see also Section 3.3.2), for which these leading

indicators play a crucial role. This is of particular relevance for short-term forecasting and nowcasting of such a (low-frequency) variable characterised by a publication lag. For some leading indicators, single variables are combined into a composite index such as The Conference Board Leading Economic Index published by The Conference Board for the United States. It consists of averages of ten variables, including average weekly hours in manufacturing, manufacturers' new orders, private housing building permits, and average consumer expectations ([The Conference Board, 2020](#)). Other leading indicators consist of one variable only and are purely survey-based, such as the monthly Ifo Business Climate Index published by the Ifo Institute for Germany, which covers the appraisal of current and future expected business of executives of approximately 9,000 German firms across industries ([Ifo Institute, 2020](#)).

If a specific leading indicator is able to capture certain economic changes before they happen, such a leading indicator is said to Granger-cause (or predictively cause) a forecast variable ([Granger, 1969](#)), implying that a "cause" cannot happen after an effect. In (theoretical) econometric terms and following the notation of [Lütkepohl \(2005\)](#), is Granger-caused by if for at least one forecast horizon the following inequality holds: In other words, the mean square error of the optimal h -step-ahead predictor of the forecast variable, which includes the information contained in the leading indicator in the set of all relevant information to the forecaster available at the forecast origin, must be smaller than the mean square error of that h -step-ahead predictor of the forecast variable without the information contained in said leading indicator ([Lütkepohl, 2005](#)). Nonetheless, the concept of Granger causality also has its limitations. The notion that a "cause" cannot happen after an effect is only a necessary, yet not a sufficient condition for causality, hence the reason why "cause" has been written with quotation marks in this section. If one ignored this fact and equated causality with Granger causality, they would commit an informal logical fallacy called Post hoc ergo propter hoc (i.e., after this, therefore because of this; [Walton, Reed, & Macagno, 2008](#)). A bold example of this fallacy would be, "Because the birds migrate, winter is coming". This is a fallacy, as winter would come at about the same time every year, no matter if there were migrating birds or not. Moreover, hardly any economic activity is monocausal.

There are also different types of formal statistical Granger causality tests available for different data structures that are implemented in typical statistics/econometrics software packages. For the simple case of two variables ([Granger, 1969](#)), say the forecast

variable and the leading indicator, the null hypothesis of a bivariate Granger causality test is that the leading indicator does not Granger-cause the forecast variable. Under this null hypothesis, the F -test statistic on the joint impact of the coefficients of the past realisations of the leading indicator employed as explanatory variables in a multiple linear regression with the forecast variable as dependent variable and its past realisations as additional explanatory variables will not be statistically significantly different from zero. The maximum lag for past realisations would be optimally determined, for instance, by some information criterion (e.g., AIC, BIC; see also Section 2.5.4). Practical applications of Granger causality and leading indicators in tourism demand forecasting can be found, for instance, in Section 3.8.1.

2.5.2. Model complexity³⁶

A simple model must be easily understood by decision-makers. On the contrary, relationships in a complex model are opaque for its users (see also Section 3.7.4). In this context, complexity is not measured solely by the number of parameters, but also by the functional form of the relationships among variables.

Complex models are commonly believed to deliver better forecasts, as they well describe sophisticated economic structures and offer good fit the data. Consequently, they are favoured by researchers, decision-makers and academic journals. However, empirical studies provide little evidence about their forecasting superiority over simple models. This can be explained using the bias–variance trade-off framework, in which the mean squared error can be decomposed into Noise is driven by the random term in the model. It cannot be avoided, even if we know the true DGP. Variance is caused by the need to estimate the model parameters, hence its value increases with model complexity and declines with the sample size. Bias is predominantly related to model mis-specification, which is most likely to occur for simple methods. The implications of the above framework are twofold: (i) the relationship between model complexity and MSE is U-shaped and (ii) the optimal model complexity increases with the sample size.

The illustration of the bias–variance trade-off for a simple autoregressive model (see Section 2.3.4) is provided by [Ca' Zorzi, Muck, and Rubaszek \(2016\)](#). They present analytical proof that for any persistent DGP of the form characterised by half-life of over two years, the accuracy of forecasts from the random walk or the AR(1) model with parameter fixed at an arbitrary value consistent with half-life of five years tend to be higher than that from the estimated AR(1) model. This result explains why in numerous

studies the random walk is a tough benchmark as well as why a simple, calibrated AR(1) model can be successful in forecasting inflation (Faust & Wright, 2013), exchange rates (Ca' Zorzi, Kolasa, & Rubaszek, 2017) or oil prices (Rubaszek, 2020) compared to a number of complex competitors.

Wide support to the view that model simplicity improves forecasting performance is presented by Green and Armstrong (2015), in an introductory article to the special issue of *Journal of Business Research* “Simple versus complex forecasting”, as well as the results of M1 and M2 competitions (Makridakis et al., 1982, Makridakis et al., 1993). Stock and Watson (1998) also show that for most US monthly series complex non-linear autoregressive models deliver less accurate forecasts than their linear counterparts. On the contrary, the results of M4 competition tend to favour more complex models (Makridakis, Spiliotis & Assimakopoulos, 2020 and Section 2.12.7). Why are then complex models preferred to simple ones, if the latter deliver more accurate forecasts? Brighton and Gigerenzer (2015) claim that there is a tendency to overemphasize the bias component and downplay the role of variance. This behaviour is leading to an implicit preference towards more complex forecasting methods, which is called by the authors as “bias bias”. To avoid it, one can follow the golden rules of forecasting, which says: be conservative in the choice of over-ambitious models and be wary of the difficulty to forecast with complex methods (Armstrong, Green, & Graefe, 2015). The alternative is to use methods, which explicitly account for the bias–variance trade-off, e.g. machine learning (see Section 2.7.10).

2.5.3. Variable selection³⁷

References to ‘big data’ have become somewhat ubiquitous both in the media and in academic literature in recent years (see Section 2.7.1 but also Section 2.2.5). Whilst in some disciplines (for example Finance) it has become possible to observe time series data at ever higher frequency, it is in the cross section that the amount of data available to analysts has seen exponential growth.

Ordinary Least Squares (OLS) is the standard tool for data analysis and prediction, but is well known to perform poorly when there many potential explanatory variables; Bishop (2006) sets out clearly why this is so. In situations where there is no obvious underlying model to suggest which of a potentially large set of candidate variables to focus on, the researcher needs to add new tools to the tool kit.

There are two principal sets of approaches to this problem. The first seeks to reduce the model dimensions by summarising the predictors in to a much smaller number of aggregated indices or factors. The second is to employ a regularisation strategy to reduce the effective dimension of the problem by shrinking the size of the regression coefficients. Some strategies reduce a subset of these coefficients to zero, removing some variables from the model entirely and these can hence be described as variable selection procedures.

Such procedures may be applicable either when a problem is believed to be truly sparse, with only a small number of variables having an effect, or alternatively when a sparse model can provide effective forecasts, even though the underlying system is in fact more complex (see also Section 2.5.2).

In the frequentist framework, the Least Absolute Shrinkage and Selection Operator (LASSO) procedure of Tibshirani (1996) has proven effective in many applications. The LASSO requires choice of an additional regularisation parameter (usually selected by some statistical criteria or via a cross validation process). Various refinements to the original LASSO procedure have been developed, see in particular Rapach, Strauss, Tu, and Zhou (2019) for a recent forecasting application. An alternative frequentist approach to variable selection and shrinkage is the Complete Subset Regression (CSR) model of Elliott (2015), where separate OLS regressions are run on all possible combinations of potential regressors, with forecasts generated by averaging across the entire set of models. Kotchoni, Leroux, and Stevanovic (2019) combine CSR with LASSO in a comprehensive empirical economic forecasting exercise.

Where the underlying process is not properly sparse (see Giannone & Primiceri, 2017 for a discussion), it is perhaps more natural to work in a Bayesian framework where samples can be drawn from the variable selection part of model reflecting the estimated probability of inclusion of each variable. The appropriate degree of regularisation can also be selected in a similar way. Forecasts are then constructed as weighted averages over several sparse models. This approach has proven to be very effective in practice, for example in fitting Bayesian Vector Auto Regressions to forecast economic time series. Early examples include George and McCulloch (1993) and Mitchell and Beauchamp (1988), which use binary indicators to select variables in to the model. More recent approaches use continuous random variables to achieve a similar effect, making computation more tractable. Examples include the Horseshoe Prior (Carvalho et al.,

2010, Piironen and Vehtari, 2017) and the LN-CASS prior of Thomson, Jabbari, Taylor, Arlt, and Smith (2019). Cross (2020) is a recent example of an economic forecasting exercise using several such models.

2.5.4. Model selection³⁸

Taxonomies of all possible sources of forecast errors from estimated models that do not coincide with their data generating process (DGP) have revealed that two mistakes determine forecast failures (i.e., systematic departures between forecasts and later outcomes), namely mis-measured forecast origins and unanticipated location shifts (Clements and Hendry, 1999, Hendry and Mizon, 2012). The former can be addressed by nowcasting designed to handle breaks (Bańbura et al., 2011, Castle and Hendry, 2010, Castle et al., 2018). It is crucial to capture the latter as failure to do so will distort estimation and forecasts, so must be a focus when selecting models for forecasting facing an unknown number of in-sample contaminating outliers and multiple breaks at unknown times.

Consequently, selection must be jointly over both observations and variables, requiring computer learning methods (Castle et al., 2020b, Hendry and Doornik, 2014 see also Section 3.6.2). *Autometrics*, a multiple-path block-search algorithm (see Doornik, 2018), uses impulse (IIS: Hendry, Johansen, & Santos, 2008a and Johansen & Nielsen, 2009) and step (SIS: Castle, Doornik, Hendry, & Pretis, 2015b) indicator saturation for discovering outliers and breaks, as well as selecting over other variables. Although knowing the in-sample DGP need not suffice for successful forecasting after out-of-sample shifts, like financial crises and pandemics, ‘robust variants’ of selected models can then avoid systematic mis-forecasts (see, for example, Doornik et al., 2020a, Martinez et al., 2021 and Section 3.3.4).

Saturation estimation has approximately candidates for regressors, with possible models, requiring selection with more candidate variables, , than observations, (also see Section 2.7 and Section 2.3.9). Selection criteria like AIC (Akaike, 1973), BIC (Schwarz, 1978), and HQ (Hannan & Quinn, 1979) are insufficient in this setting. For saturation estimation, we select at a tight significance level, , retaining subject-theory variables and other regressors. When forecasting is the aim, analyses and simulations suggest loose significance for then selecting other regressors, close to the 10% to 16% implied significance level of AIC, regardless of location shifts at or near the forecast origin (Castle, Doornik, & Hendry, 2018a). At loose levels, *Autometrics* can find multiple

undominated terminal models across paths, and averaging over these, a univariate method and a robust forecasting device can be beneficial, matching commonly found empirical outcomes. The approach applies to systems, whence selection significance of both indicators and variables is judged at the system level. Capturing in-sample location shifts remains essential (Doornik, Castle, & Hendry, 2020b).

There are costs and benefits of selection for forecasting (also see Sections 2.5.3 Variable selection, 2.11.3 Judgmental model selection). Selection at a loose significance level implies excluding fewer relevant variables that contribute to forecast accuracy, but retaining more irrelevant variables that are adventitiously significant, although fewer than by simply averaging over all sub-models (Hoeting, Madigan, Raftery, & Volinsky, 1999). Retaining irrelevant variables that are subject to location shifts worsens forecast performance, but their coefficient estimates are driven towards zero when updating estimates as the forecast origin moves forward. Lacking omniscience about future values of regressors that then need to be forecast, not knowing the DGP need not be costly relative to selecting from a general model that nests it (Castle et al., 2018a). Overall, the costs of model selection for forecasting are small compared to the more fundamental benefit of finding location shifts that would otherwise induce systematic forecast failure.

2.5.5. Cross-validation for time-series data³⁹

When building a predictive model, its purpose is usually not to predict well the already known samples, but to obtain a model that will generalise well to new, unseen data. To assess the out-of-sample performance of a predictive model we use a test set that consists of data not used to estimate the model (for a discussion of different error measures used in this context see Section 2.12.2). However, as we are now using only parts of the data for model building, and other parts of the data for model evaluation, we are not making the best possible use of our data set, which is a problem especially if the amount of data available is limited. Cross-validation (CV), first introduced by Stone (1974), is a widely used standard technique to overcome this problem (Hastie et al., 2009) by using all available data points for both model building and testing, therewith enabling a more precise estimation of the generalisation error and allowing for better model selection. The main idea of (-fold) cross-validation is to partition the data set randomly into subsets, and then use each of the subsets to evaluate a model that has been estimated on the remaining subsets. An excellent overview of different cross-validation techniques is given by Arlot and Celisse (2010).

Despite its popularity in many fields, the use of CV in time series forecasting is not straightforward. Time series are often non-stationary and have serial dependencies (see also Section 2.3.4). Also, many forecasting techniques iterate through the time series and therewith have difficulties dealing with missing values (withheld for testing). Finally, using future values to predict data from the past is not in accordance with the normal use case and therewith seems intuitively problematic. Thus, practitioners often resort to out-of-sample evaluation, using a subset from the very end of a series exclusively for evaluation, and therewith falling back to a situation where the data are not used optimally.

To overcome these problems, the so-called time-series cross-validation (Hyndman & Athanasopoulos, 2018) extends the out-of-sample approach from a fixed origin to a rolling origin evaluation (Tashman, 2000). Data is subsequently moved from the out-of-sample block from the end of the series to the training set. Then, the model can be used (with or without parameter re-estimation) with the newly available data. The model re-estimation can be done on sliding windows with fixed window sizes or on expanding windows that always start at the beginning of the series (Bell & Smyl, 2018).

However, these approaches extending the out-of-sample procedure make again not optimal use of the data and may not be applicable when only small amounts of data are available. Adapting the original CV procedure, to overcome problems with serial correlations, blocked CV approaches have been proposed in the literature, where the folds are chosen in blocks (Bergmeir and Benítez, 2012, Racine, 2000) and/or data around the points used for testing are omitted (Burman et al., 1994, Racine, 2000). Finally, it has been shown that with purely autoregressive models, CV can be used without modifications, i.e., with randomly choosing the folds (Bergmeir, Hyndman, & Koo, 2018). Here, CV estimates the generalisation error accurately, as long as the model errors from the in-sample fit are uncorrelated. This especially holds when models are overfitting. Underfitting can be easily detected by checking the residuals for serial correlation, e.g., with a Ljung–Box test (Ljung & Box, 1978). This procedure is implemented in the *forecast* package (Hyndman et al., 2020) in R (R Core Team, 2020), in the function `CVar`.

2.6. Combining forecasts

2.6.1. Forecast combination: a brief review of statistical approaches⁴⁰

Given forecasts of the same event, forecast combination involves estimation of so called combination weights assigned to each forecast, such that the accuracy of the combined forecast generally outperforms the accuracy of the forecasts included. Early statistical approaches adopted a range of strategies to estimate combination weights including (i) minimising in-sample forecast error variance among forecast candidates (Bates and Granger, 1969, Min and Zellner, 1993, Newbold and Granger, 1974), (ii) formulation and estimation via ordinary least squares regression (Granger and Ramanathan, 1984, MacDonald and Marsh, 1994), (iii) use of approaches based on Bayesian probability theory (Bordley, 1982, Bunn, 1975, Clemen and Winkler, 1986, Diebold and Pauly, 1990, Raftery, 1993 and Section 2.4), (iv) and the use of regime switching and time varying weights recognising that weights can change over time (Diebold and Pauly, 1987, Elliott et al., 2005, Lütkepohl, 2011, Tian and Anderson, 2014 and Section 2.3.12). Barrow and Kourentzes (2016) contains a very good documentation and empirical evaluation of a range of these early approaches, while De Menezes, Bunn, and Taylor (2000) and Armstrong (2001a) contain guidelines on their use.

Recent statistical approaches use a variety of techniques to generate forecasts and/or derive weights. Kolassa (2011) apply so called Akaike weights based on Akaike Information Criterion (Sakamoto, Ishiguro, & Kitagawa, 1986), while bootstrapping has been used to generate and combine forecast from exponential smoothing (Barrow et al., 2020, Bergmeir et al., 2016, Cordeiro and Neves, 2009 but also Section 2.7.5 and Section 2.7.6), artificial neural networks (Barrow and Crone, 2016a, Barrow and Crone, 2016b and Section 2.7.8), and other forecast methods (Athanasopoulos et al., 2018, Hillebrand and Medeiros, 2010, Inoue and Kilian, 2008). Barrow and Crone (2016b) developed cross-validation and aggregating (Crogging) using cross-validation to generate and average multiple forecasts, while more recently, combinations of forecasts generated from multiple temporal levels has become popular (Athanasopoulos et al., 2017, Kourentzes and Athanasopoulos, 2019, Kourentzes et al., 2014 and Section 2.10.2). These newer approaches recognise the importance of forecast generation in terms of uncertainty reduction (Petropoulos, Hyndman, & Bergmeir, 2018a), the creation of diverse forecasts (Brown et al., 2005b, Lemke and Gabrys, 2010), and the pooling of forecasts (Kourentzes et al., 2019, Lichtendahl Jr and Winkler, 2020).

Now nearly 50 years on from the seminal work of [Bates and Granger \(1969\)](#), the evidence that statistical combinations of forecasts improves forecasting accuracy is near unanimous, including evidence from competitions ([Makridakis et al., 1982](#), [Makridakis and Hibon, 2000](#), [Makridakis, Spiliotis et al., 2020](#) and Section 2.12.7), and empirical studies ([Andrawis et al., 2011](#), [Elliott et al., 2005](#), [Jose and Winkler, 2008](#), [Kourentzes et al., 2019](#)). Still, researchers have tried to understand why and when combinations improve forecast accuracy ([Atiya, 2020](#), [Palm and Zellner, 1992](#), [Petropoulos et al., 2018a](#), [Timmermann, 2006](#)), and the popularity of the simple average ([Chan and Pauwels, 2018](#), [Claeskens et al., 2016](#), [Smith and Wallis, 2009a](#)). Others have investigated properties of the distribution of the forecast error beyond accuracy considering issues such as normality, variance, and in out-of-sample performance of relevance to decision making ([Barrow and Kourentzes, 2016](#), [Chan et al., 1999](#), [Makridakis and Winkler, 1989](#)).

Looking forward, evidence suggests that the future lies in the combination of statistical and machine learning generated forecasts ([Makridakis, Spiliotis et al., 2020](#)), and in the inclusion of human judgment ([Gupta, 1994](#), [Petropoulos et al., 2018b](#), [Wang and Petropoulos, 2016](#) but also Section 2.11.1). Additionally, there is need to investigate such issues as decomposing combination accuracy gains, constructing prediction intervals ([Grushka-Cockayne and Jose, 2020](#), [Koenker, 2005](#)), and generating combined probability forecasts ([Clements and Harvey, 2011](#), [Hall and Mitchell, 2007](#), [Raftery et al., 1997](#), [Ranjan and Gneiting, 2010](#) and Section 2.6.2). Finally, there is need for the results of combined forecasts to be more interpretable and suitable for decision making ([Barrow and Kourentzes, 2016](#), [Bordignon et al., 2013](#), [Graefe et al., 2014](#), [Todini, 2018](#)).

2.6.2. Density forecast combinations⁴¹

Density forecasts provide an estimate of the future probability distribution of a random variable of interest. Unlike point forecasts (and point forecasts supplemented by prediction intervals) density forecasts provide a complete measure of forecast uncertainty. This is particularly important as it allows decision makers to have full information about the risks of relying on the forecasts of a specific model. Policy makers like the Bank of England, the European Central Bank and Federal Reserve Banks in the US routinely publish density forecasts of different macroeconomic variables such as inflation, unemployment rate, and GDP. In finance, density forecasts find application, in

particular, in the areas of financial risk management and forecasting of stock returns (see, for example, [Berkowitz, 2001](#), [Guidolin and Timmermann, 2006](#), [Shackleton et al., 2010](#), [Tay and Wallis, 2000](#) *inter alia*). The reader is referred to Section 3.3 for a discussion of relevant applications.

Initial work on forecast combination focused on the combination of point forecasts (see Section 2.6.1). In recent years attention has shifted towards evaluation, comparison and combination of density forecasts with empirical applications that are mostly encountered in the areas of macroeconomics and finance. The improved performance of combined density forecasts stems from the fact that pooling of forecasts allows to mitigate potential misspecifications of the individual densities when the true population density is non-normal. Combining normal densities yields a flexible mixture of normals density which can accommodate heavier tails (and hence skewness and kurtosis), as well as approximate non-linear specifications ([Hall and Mitchell, 2007](#), [Jore et al., 2010](#)).

The predictive density combination schemes vary across studies and range from simple averaging of individual density forecasts to complex approaches that allow for time-variation in the weights of prediction models, also called experts (see [Aastveit, Mitchell, Ravazzolo, & van Dijk, 2019](#) for a comprehensive survey of density forecast combination methods). A popular approach is to combine density forecasts using a convex combination of experts' predictions, so called 'linear opinion pools' (see, for example, [Geweke and Amisano, 2011](#), [Hall and Mitchell, 2007](#), [Kascha and Ravazzolo, 2010](#)). In order to determine the optimal combination weights this method relies on minimising Kullback–Leibler divergence of the true density from the combined density forecast. These linear approaches have been extended by [Gneiting and Ranjan \(2013\)](#) to allow non-linear transformations of the aggregation scheme and by [Kapetanios, Mitchell, Price, and Fawcett \(2015\)](#), whose 'generalised pools' allow the combination weights to depend on the (forecast) value of the variable of interest.

[Billio et al. \(2013\)](#) developed a combination scheme that allows the weights associated with each predictive density to be time-varying, and propose a general state space representation of predictive densities and combination schemes. The constraint that the combination weights must be non-negative and sum to unity implies that linear and Gaussian state-space models cannot be used for inference and instead Sequential Monte Carlo methods (particle filters) are required. More recently, [McAlinn and West \(2019\)](#) developed a formal Bayesian framework for forecast combination (Bayesian

predictive synthesis) which generalises existing methods. Specifying a dynamic linear model for the synthesis function, they develop a time-varying (non-convex/nonlinear) synthesis of predictive densities, which forms a dynamic latent (agent) factor model. For a discussion on methods for evaluating probabilistic forecasts, see Sections [2.12.4 Evaluating probabilistic forecasts](#), [2.12.5 Assessing the reliability of probabilistic forecasts](#).

2.6.3. Ensembles and predictive probability post processors⁴²

Improved rational decisions are the final objective of modelling and forecasting. Relatively easy decisions among a number of alternatives with predefined and known outcomes become hard when they are conditioned by future unknown events. This is why one resorts to modelling and forecasting, but this is insufficient. To be successful, one must account for the future conditioning event uncertainty to be incorporated into the decision-making process using appropriate Bayesian approaches (see also Section [2.4](#)), as described by the decision theory literature ([Berger, 1985](#), [Bernardo, 1994](#), [DeGroot, 2004](#)). The reason is that taking a decision purely based on model forecasts is equivalent to assuming the future event (very unlikely, as we know) to equal the forecasted value. Therefore, the estimation of the predictive probability density is the essential prerequisite to estimating the expected value of benefits (or of losses) to be compared and traded-off in the decision-making process ([Draper & Krnjajić, 2013](#)). This highly increases the expected advantages together with the likelihood of success and the robustness of the decision ([Todini, 2017](#), [Todini, 2018](#)).

In the past, the assessment of the prediction uncertainty was limited to the evaluation of the confidence limits meant to describe the quality of the forecast. This was done using continuous predictive densities, as in the case of the linear regression, or more frequently in the form of predictive ensembles. These probabilistic predictions, describing the uncertainty of the model forecasts given (knowing) the observations can be used within the historical period to assess the quality of the models ([Todini, 2016](#)). When predicting into the future, observations are no more available and what we look for (known as predictive probability) is the probability of occurrence of the unknown “future observations” given (knowing) the model forecasts. This can be obtained via Bayesian inversion, which is the basis of several uncertainty post-processors used in economy ([Diebold, Gunther, & Tay, 1998](#) and Section [3.3](#)), hydrology ([Krzysztofowicz, 1999](#), [Schwanenberg et al., 2015](#), [Todini, 1999](#), [Todini, 2008](#) and Section [3.5.4](#)),

meteorology ([Economou et al., 2016](#), [Granger and Pesaran, 2000](#), [Katz and Lazo, 2011](#), [Reggiani and Boyko, 2019](#) see also Section 3.5.2), etc. Accordingly, one can derive a predictive probability from a single model forecast to estimate the expected value of a decision by integrating over the entire domain of existence all the possible future outcomes and their effects, weighted with their appropriate probability of occurrence. When several forecasts are available to a decision maker, the problem of deciding on which of them one should rely upon becomes significant. It is generally hard to choose among several forecasts because one model could be best under certain circumstances but rather poor under others. Accordingly, to improve the available knowledge on a future unknown event, predictive densities are extended to multiple forecasts to provide decision makers with a single predictive probability, conditional upon several model's forecasts ([Coccia and Todini, 2011](#), [Raftery et al., 1997](#) see also Section 2.6.1 and Section 2.6.2).

A number of available uncertainty post processors can cope with multi-model approaches, such as Bayesian model averaging ([Raftery, 1993](#), [Raftery et al., 1997](#)), model output statistics ([Glahn and Lowry, 1972](#), [Wilkd, 2005](#)), ensemble model output statistics ([Gneiting, Raftery, Westveld, & Goldman, 2005](#)), and model conditional processor ([Coccia and Todini, 2011](#), [Todini, 2008](#)).

Finally, important questions such as: (i) “what is the probability that an event will happen within the next hours?” and (ii) “at which time interval it will most likely occur?” can be answered using a multi-temporal approach ([Coccia, 2011](#), [Krzysztofowicz, 2014](#) see also Section 2.10.2) and results of its applications were presented in [Barbetta et al., 2017](#), [Coccia, 2011](#), and [Todini \(2017\)](#).

2.6.4. The wisdom of crowds⁴³

Multiple experts' forecasts are collected in a wide variety of situations: medical diagnostics, weather prediction, forecasting the path of a hurricane, predicting the outcome of an election, macroeconomic forecasting, and more. One of the central findings from the forecasting literature is that there is tremendous power in combining such experts' forecasts into a single forecast. The simple average, or what Surowiecki refers to as ‘the wisdom of crowds’ ([Surowiecki, 2005](#)), has been shown to be a surprisingly robust combined forecast in the case of point forecasting ([Armstrong, 2001b](#), [Clemen, 1989](#), [Clemen and Winkler, 1986](#) and Section 2.6.1). The average forecast is more accurate than choosing a forecast from the crowd at random and is

sometimes even more accurate than nearly all individuals (Mannes, Larrick, & Soll, 2012). The average point forecast also often outperforms more complicated point aggregation schemes, such as weighted combinations (Smith and Wallis, 2009b, Soule et al., 2020).

Mannes et al. (2012) highlight two crucial factors that influence the quality of the average point forecast: individual expertise and the crowd's diversity. Of the two: "The benefits of diversity are so strong that one can combine the judgments from individuals who differ a great deal in their individual accuracy and still gain from averaging" (Mannes et al., 2012 page 234).

Larrick and Soll (2006) define the idea of 'bracketing': In the case of averaging, two experts can either bracket the realisation (the truth) or not. When their estimates bracket, the forecast generated by taking their average performs better than choosing one of the two experts at random; when the estimates do not bracket, averaging performs equally as well as the average expert. Thus, averaging can do no worse than the average expert, and with some bracketing, it can do much better. Modern machine learning algorithms such as the random forest exploit this property by averaging forecasts from hundreds of diverse experts (here, each "expert" is a regression tree; Grushka-Cockayne, Jose & Lichtendahl, 2017).

Only when the crowd of forecasts being combined has a high degree of dispersion in expertise, some individuals in the crowd might stand out, and in such cases, there could be some benefits to chasing a single expert forecaster instead of relying on the entire crowd. Mannes, Soll, and Larrick (2014) suggest that combining a small crowd can be especially powerful in practice, offering some diversity among a subset of forecasters with an minimum level of expertise.

When working with probabilistic forecasting (see also Sections 2.6.2 Density forecast combinations, 2.6.3 Ensembles and predictive probability post processors, 2.12.4 Evaluating probabilistic forecasts), averaging probabilities is the most widely used probability combination method (Clemen, 2008, Cooke, 1991, Hora, 2004). Stone (1961) labelled such an average the linear opinion pool. O'Hagan, Buck, Daneshkhah, Richard Eiser, Garthwaite, Jenkinson, Oakley, and Rakow (2006) claimed that the linear opinion pool is: "hard to beat in practice".

Although diversity benefits the average point forecast, it can negatively impact the average probability forecast. As the crowd's diversity increases, the average probability

forecast becomes more spread out, or more underconfident (Dawid et al., 1995, Hora, 2004, Ranjan and Gneiting, 2010). Averaging quantiles, instead of probabilities, can offer sharper and better calibrated forecasts (Lichtendahl, Grushka-Cockayne, & Winkler, 2013). Trimmed opinion pools can be applied to probability forecasts, also resulting in better calibrated forecasts (Jose, Grushka-Cockayne, & Lichtendahl, 2014 see also Section 2.12.5).

The ubiquity of data and the increased sophistication of forecasting methods results in more use of probabilistic forecasts. While probabilities are more complex to elicit, evaluate, and aggregate compared to point estimates, they do contain richer information about the uncertainty of interest. The wisdom of combining probabilities, however, utilises diversity and expertise differently than combining point forecasts. When relying on a crowd, eliciting point forecasts versus eliciting probabilities can significantly influence the type of aggregation one might choose to use.

2.7. Data-driven methods

2.7.1. Forecasting with big data⁴⁴

The last two decades have seen a proliferation of literature on forecasting using big data (Hassani and Silva, 2015, Swanson and Xiong, 2018, Varian, 2014) but the evidence is still uncertain as to whether the promised improvements in forecast accuracy can systematically be realised for macroeconomic phenomena. In this section we question whether big data will significantly increase the forecast accuracy of macroeconomic forecasts. Athey (2018) argues that machine learning methods are seen as an efficient approach to dealing with big data sets, and we present these methods before questioning their success at handling non-stationary macroeconomic data that are subject to shifts. Section 2.7.2 discusses big data in the context of distributed systems, and Section 2.7.11 evaluates a range of machine learning methods frequently applied to big data.

The tools used to analyse big data focus on regularization techniques to achieve dimension reduction, see Kim and Swanson (2014) for a summary of the literature. This can be achieved through selection (such as *Autometrics*, Doornik, 2018 but also see Section 2.5.3 and Section 2.5.4), shrinkage (including Ridge Regression, LASSO, and Elastic Nets, see Section 2.7.11 but also Section 3.3.13 for an applied example), variable

combination (such as Principal Components Analysis and Partial Least Squares), and machine learning methods (including Artificial Neural Networks, see Section 2.7.8). Many of these methods are ‘black boxes’ where the algorithms are not easily interpretable, and so they are mostly used for forecasting rather than for policy analysis. Big data has been effectively used in nowcasting, where improved estimates of the forecast origin lead to better forecasts, absent any later shifts. Nowcasting can benefit from large data sets as the events have happened and the information is available, see [Castle et al. \(2018\)](#) for a nowcasting application, and Section 2.5.1 on leading indicators. However, the benefits of big data are not as evident in a forecasting context where the future values of all added variables also need to be forecast and are as uncertain as the variable(s) of interest.

Macroeconomic time series data are highly non-stationary, with stochastic trends and structural breaks. The methods of cross-validation and hold-back, frequently used to handle big data, often assume that the data generating process does not change over time. Forecasting models that assume the data are drawn from a stationary distribution (even after differencing) do not forecast well *ex ante*. So while there seems to be lots of mileage in improving forecasts using big data, as they allow for more flexible models that nest wider information sets, more general dynamics and many forms of non-linearity, the statistical problems facing ‘small’ data forecasting models do not disappear ([Doornik and Hendry, 2015](#), [Harford, 2014](#)). [Castle, Doornik, and Hendry \(2020a\)](#) do not find improvements in forecasting from big data sets over small models. It is essential to keep in mind the classical statistical problems of mistaking correlation for causation, ignoring sampling biases, finding excess numbers of false positives and not handling structural breaks and non-constancies both in- and out-of-sample, in order to guard against these issues in a data abundant environment.

2.7.2. Forecasting on distributed systems⁴⁵

Big data is normally accompanied by the nature that observations are indexed by timestamps, giving rise to big data time series characterised by high frequency and long-time span. Processing big data time series is obstructed by a wide variety of complications, such as significant storage requirements, algorithms’ complexity and high computational cost ([Galicia et al., 2018](#), [L’heureux et al., 2017](#), [Wang, Kang, Hyndman, and Li, 2020](#), [Wang, Yang, et al., 2018](#)). These limitations accelerate the great demand for scalable algorithms. Nowadays, increasing attention has been paid to

developing data mining techniques on distributed systems for handling big data time series, including but not limited to processing (Mirko & Kantelhardt, 2013), decomposition (Bendre & Manthalkar, 2019), clustering (Ding, Wang, Dang, Fu, Zhang, & Zhang, 2015), classification (Triguero, Peralta, Bacardit, García, & Herrera, 2015), and forecasting (Galicia et al., 2018). For forecasting problems based on big data sets and/or large sets of predictors, please refer to Sections 2.7.1 Forecasting with big data, 3.3.13 Forecasting stock returns.

Distributed systems, initially designed for independent jobs, do not support to deal with dependencies among observations, which is a critical obstacle in time series processing (Li et al., 2014, Wang, Kang, Hyndman, and Li, 2020). Various databases (e.g., InfluxDB,⁴⁶ OpenTSDB,⁴⁷ RRDtool,⁴⁸ and Timely⁴⁹) can function as storage platforms for time series data. However, none of these databases supports advanced analysis, such as modelling, machine learning algorithms and forecasting. Additional considerations are therefore required in further processing time series on distributed systems. Mirko and Kantelhardt (2013) developed the Hadoop.TS library for processing large-scale time series by creating a time series bucket. Li et al. (2014) designed an index pool serving as a data structure for assigning index keys to time series entries, allowing time series data to be sequentially stored on HDFS (Hadoop Distributed File System: Shvachko, Kuang, Radia, & Chansler, 2010) for MapReduce (Dean & Ghemawat, 2008) jobs. Chen, Li, Rong, Bilal, Li, and Philip (2019a) proposed a data compression and abstraction method for large-scale time series to facilitate the periodicity-based time series prediction in a parallel manner.

The evolution of the algorithms for efficiently forecasting big data time series on distributed systems is largely motivated by a wide range of applications including meteorology, energy, finance, transportation and farming (Chen et al., 2019a, Galicia et al., 2018, Hong and Pinson, 2019, Sommer et al., 2020). Researchers have made several attempts to make machine learning techniques available for big data time series forecasting on distributed systems (Galicia et al., 2019, Li et al., 2014, Talavera-Llames et al., 2016, Xu et al., 2020). Talavera-Llames et al. (2016) presented a nearest neighbours-based algorithm implemented for Apache Spark (Zaharia, Xin, Wendell, Das, Armbrust, Dave, Meng, Rosen, Venkataraman, Franklin, Ghodsi, Gonzalez, Shenker, & Stoica, 2016) and achieved satisfactory forecasting performance. Galicia et al. (2018) proposed a scalable methodology which enables Spark's MLib (Meng et al.,

2016) library to conduct multi-step forecasting by splitting the multi-step forecasting problem into sub-problems (is the forecast horizon).

Another strand of the literature on forecasting big data time series is to improve time-consuming estimation methods using a MapReduce framework. [Sheng, Zhao, Leung, and Wang \(2013\)](#) learned the parameters of echo state networks for time series forecasting by designing a parallelised extended Kalman filter involving two MapReduce procedures. Recently, [Sommer et al. \(2020\)](#) accurately estimated coefficients of a high-dimensional ARX model by designing two online distributed learning algorithms. [Wang, Kang, Hyndman, and Li \(2020\)](#) resolved challenges associated with forecasting ultra-long time series from a new perspective that global estimators are approximated by combining the local estimators obtained from subseries by minimising a global loss function. Besides, inspired by the *no-free-lunch* theorem ([Wolpert & Macready, 1997](#)), model selection (see Section 2.5.4) and model combination (see Section 2.6) are involved in finalisation of algorithms for forecasting on distributed systems (e.g., [Bendre and Manthalkar, 2019](#), [Galicia et al., 2019](#), [Li et al., 2014](#), [Xu et al., 2020](#)).

2.7.3. Agent-based models⁵⁰

Time series forecasting involves use of historical data to predict values for a specific period time in future. This approach assumes that recent and historical patterns in the data will continue in the future. This assumption is overly ingenuous. However, this is not reliable in some situations. For example, (i) forecasting COVID-19 cases (see also Section 3.6.2) where, due to interventions and control measures taken by the governments and due to the change in personal behaviour, the disease transmission pattern changes rapidly, and (ii) forecasting sales of a new product (see also Section 3.2.6): external factors such as advertisement, promotions (see Section 3.2.5), social learning, and imitation of other individuals change the system behaviour.

In such circumstances to make reliable forecasts it is important to take into account all information that might influence the variable that is being forecast. This information includes a variety of environmental-level and individual-level factors. An agent-based modelling is a powerful tool to explore such complex systems. Agent-based modelling approach is useful when, (i) data availability is limited, (ii) uncertainty of various interventions in place and a rapidly changing social environment, and (iii) limited understanding of the dynamics of the variable of interest.

Agent-based modelling disaggregates systems into individual level and explores the aggregate impact of individual behavioural changes on the system as a whole. In other words, the key feature of agent-based modelling is the bottom-up approach to understand how a system's complexity arises, starting with individual level (see also Section 2.10.1). As opposed to this, the conventional time series forecasting approaches are considered top-down approaches.

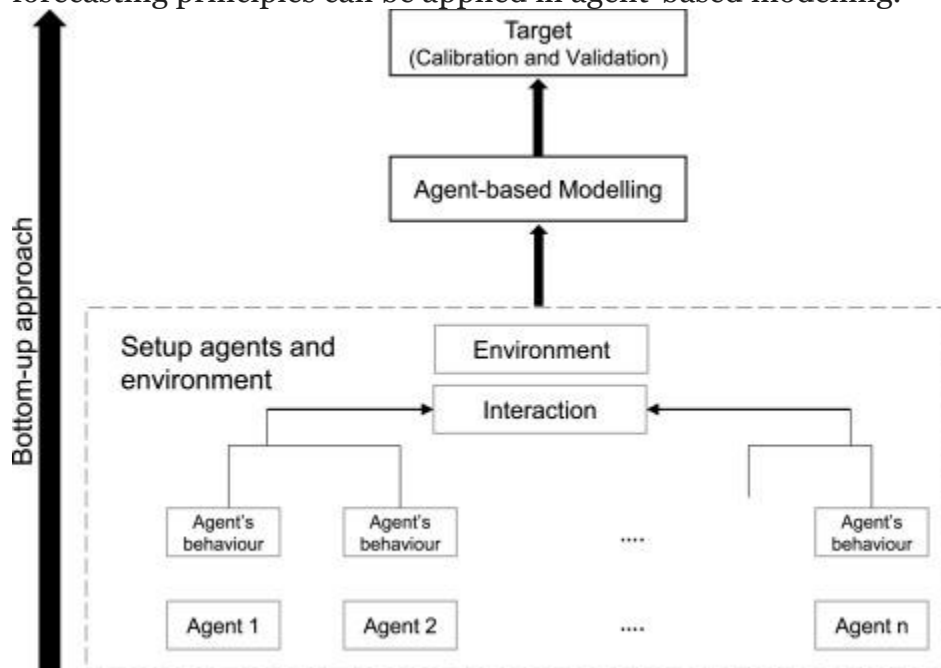
Agent-based models have two main components: (i) Agents, and (ii) Rules, sometimes referred as procedures and interactions. Agents are individuals with autonomous behaviour. Agents are heterogeneous. Each agent individually assesses on the basis of a set of rules. An agent-based modelling approach simulates how heterogeneous agents interact and behave to assess the role of different activities on the target variable.

According to [Farmer and Foley \(2009\)](#), "An agent-based model is a computerised simulation of a number of decision-makers (agents) and institutions, which interact through prescribed rules". Their paper highlights the importance of adopting agent-based models as a better way to help guide financial policies.

A general framework for agent-based modelling involves three main stages (See [Fig. 4](#)): (i) setup environments and agents, (ii) agent-based modelling, and (iii) calibration and validation. The first two steps are self-explanatory. The final step involves calibration of the model with empirical data and then evaluates whether the agent-based model mirrors the real-world system/target. The validation step involves testing the significance of the difference between agent-based model results and real data collected about the target. One of the main challenges in designing an agent-based model is finding a balance between model simplicity and model realism (see also Section 2.5.2). The KISS principle (keep it simple, stupid), introduced by [Axelrod \(1997\)](#) is often cited as an effective strategy in agent-based modelling. A high level of expertise in the area of the subject is necessary when developing an agent-based model.

Despite these limitations and challenges, agent-based modelling has been used extensively to model infectious disease transmission and forecasting ([Tracy et al., 2018](#), [Venkatramanan et al., 2018](#)). Agent-based modelling approaches have been widely used in early phases of the COVID-19 outbreak, to assess the impact of different interventions on disease spread and forecasts ([Wallentin, Kaziyeva, & Reibersdorfer-Adelsberger, 2020](#)). In a review paper, [Weron \(2014\)](#) states some applications of agent-based models for electricity demand forecasting. [Xiao and Han \(2016\)](#) use agent-based

models to forecast new product diffusion. Furthermore, thinking the other way around, [Hassan, Arroyo, Galán Ordax, Antunes, and Pavón Mestras \(2013\)](#) explain how forecasting principles can be applied in agent-based modelling.



1. [Download: Download high-res image \(183KB\)](#)
2. [Download: Download full-size image](#)

Fig. 4. Framework for Agent-based modelling.

2.7.4. Feature-based time series forecasting⁵¹

A time series *feature* can be any statistical representation of time series characteristics. A vast majority of time series mining tasks are based on similarity quantification using their feature representations, including but not limited to time series clustering ([Kang et al., 2014](#), [Kang et al., 2015](#), [Wang et al., 2006](#) and Section 2.7.12), classification ([Fulcher et al., 2013](#), [Nanopoulos et al., 2001](#)), anomaly detection ([Kang, 2012](#), [Talagala, Hyndman, and Smith-Miles, 2020](#) and Section 2.2.3), and forecasting ([Kang et al., 2017](#), [Montero-Manso et al., 2020](#) see also Section 2.2.5). The choice of features depends on the nature of the data and the application context. The state-of-the-art time series feature representation methods quantify a wide range of time series characteristics, including simple summary statistics, stationarity ([Montero-Manso et al., 2020](#), [Wang, Kang et al., 2021](#)), model fits ([Christ et al., 2018](#), [Fulcher and Jones, 2014](#)), time series imaging ([Li, Kang & Li, 2020](#)), and others. In the forecasting community, two lines of forecasting approaches have been developed using time series features,

namely feature-based model selection and feature-based model combination. The motivation behind them is no single model always performs the best for all time series. Instead of choosing one model for all the data, features can be used to obtain the most appropriate model or the optimal combination of candidate models, per series. As early as in 1972, [Reid \(1972\)](#) argues that time series characteristics provide valuable information in forecast model selection, which is further echoed by [Makridakis and Hibon \(1979\)](#). One way to forecast an extensive collection of time series is to select the most appropriate method per series according to its features. Pioneer studies focus on rule-based methods (for example, [Arinze, 1994](#), [Wang, Smith-Miles et al., 2009](#)) to recommend the “best” forecasting model per series based on its features. Another line of approaches apply regression to study how useful features are in predicting which forecasting method performs best (for example, [Meade, 2000](#), [Petropoulos et al., 2014](#)). With the advancement of machine learning (see also Section [2.7.10](#)), more recent literature uses “meta-learning” to describe the process of automatically acquiring knowledge for forecast model selection. The first such study is by [Prudêncio and Ludermir \(2004\)](#), who apply decision trees for forecast model selection. [Lemke and Gabrys \(2010\)](#) compare different meta-learning approaches to investigate which model works best in which situation. [Kang et al. \(2017\)](#) propose using feature spaces to visualise the strengths and weaknesses of different forecasting methods. Other algorithms such as neural networks (see also Section [2.7.8](#)) and random forecasts are also applied to forecast model selection ([Kück et al., 2016](#), [Talagala et al., 2018](#)). One of the pioneering studies in feature-based forecast combination is the rule-based approach by [Collopy and Armstrong \(1992\)](#), who develop 99 rules for forecast combination based on 18 features. Recently, [Kang, Hyndman and Li \(2020\)](#) use 26 features to predict the performances of nine forecasting methods with nonlinear regression models, and obtain the combination weights as a tailored softmax function of the predicted forecasting errors. The feature-based forecast model averaging (FFORMA) framework proposed by [Montero-Manso et al. \(2020\)](#) employ 42 features to estimate the optimal combination weights of nine forecasting methods based on extreme gradient boosting (XGBoost, [Chen & Guestrin, 2016](#)). [Li, Kang et al. \(2020\)](#) first transform time series into images, and use features extracted from images to estimate the optimal combination weights. For feature-based interval forecasting, [Wang, Kang et al. \(2021\)](#) investigate how time series features affect the relative performances of prediction

intervals from different methods, and propose a general feature-based interval forecasting framework to provide reliable forecasts and their uncertainty estimation.

2.7.5. Forecasting with bootstrap⁵²

The bootstrap methodology has been widely applied in many areas of research, including time series analysis. The bootstrap procedure (Efron, 1979) is a very popular methodology for independent data because of its simplicity and nice properties. It is a computer-intensive method that presents solutions in situations where the traditional methods fail or are very difficult to apply. However, Efron's bootstrap (iid bootstrap) has revealed itself inefficient in the context of dependent data, such as in the case of time series, where the dependence structure arrangement has to be kept during the resampling scheme.

Most of the resampling for dependent data consider segments of the data to define blocks, such that the dependence structure within each block can be kept. Different versions of blocking differ in the way as blocks are constructed: the nonoverlapping block bootstrap (Carlstein, 1990), the moving block bootstrap (Künsch, 1989), the circular block bootstrap (Politis & Romano, 1992), and the stationary block bootstrap (Politis & Romano, 1994). But, if the time series process is driven from iid innovations, another way of resampling can be used.

The iid Bootstrap can then be easily extended to a dependent setup. That was the spirit of sieve bootstrap proposed by Bühlmann (1997). This method is based on the idea of fitting parametric models first and resampling from the residuals. Such models include, for example, the linear regression (Freedman, 1981) and autoregressive time series (Efron & Tibshirani, 1986). This approach is different from the previous bootstrap methods for dependent data; the sample bootstrap is (conditionally) stationary and does not present a structure of dependence. Another different feature is that the sieve bootstrap sample is not a subsample from the original data, as in the previous methods. Observe that even if the sieve bootstrap is based on a parametric model, it is nonparametric in its spirit. The AR model (see Section 2.3.4) here is just used to filter the residuals series.

A few years ago, the sieve bootstrap was used for estimating forecast intervals (Andrés et al., 2002, Zagdański, 2001). Motivated by these works, Cordeiro and Neves, 2006, Cordeiro and Neves, 2009, Cordeiro and Neves, 2010 developed a procedure to estimate point forecasts. The idea of these authors was to fit an exponential smoothing

model (see Section 2.3.1) to the time series, extract the residuals and then apply the sieve bootstrap to the residuals. Further developments of this procedure include the estimation of forecast intervals (Cordeiro & Neves, 2014) and also the detection, estimation and imputation of missing data (Cordeiro & Neves, 2013). In a recent work (Bergmeir et al., 2016), a similar approach was also considered, the residuals were extracted and resampled using moving block bootstrap (see Section 2.7.6 for further discussion).

Bickel and Freedman (1981) and later in Angus (1992) showed that in extreme value theory, the bootstrap version for the maximum (or minimum) does not converge to the extremal limit laws. Zelterman (1993) pointed out that “to resample the data for approximating the distribution of the largest observations would not work because the ‘pseudo-samples’ would never have values greater than ”⁵³. A method considering to resample a smaller size than the original sample was proposed in Hall (1990).

Recently, Neves and Cordeiro (2020) used this idea and developed a preliminary work in modelling and forecasting extremes in time series.

2.7.6. Bagging for time series forecasting⁵⁴

The term *bagging* was proposed by Breiman (1996) to describe the generation of several versions of a predictor, via Bootstrap procedures (introduced in Section 2.7.5), with a final aggregation stage. Thus, “**bootstrap aggregating**” was established as bagging. The main idea is to improve predictors’ accuracy once the data sets, drawn randomly with replacement, will approximate the original distribution. The author argues that bagging works for unstable procedures, but it was not tested for time series. Years after, Kilian and Inoue (2004) suggested the first attempts for temporal dependent data. For data-driven methods, to forecasting and simulation time series and deal with predictors ensembles, bagging has shown as a powerful tool.

A general framework for ensemble forecasting methods involves four main stages: (i) data treatment, (ii) resampling, (iii) forecasting, and (iv) aggregation. However, for time series, bootstrap should be done carefully, as the serial dependence and non-stationarity must be considered.

As mentioned in Section 2.7.5, this led Bergmeir et al. (2016) to propose a bagging version for exponential smoothing, the Bagged ETS. As pre-treatment, after a Box–Cox transformation, the series is decomposed into trend, seasonal, and remainder components via STL decomposition (Cleveland et al., 1990). The resampling stage uses

moving block bootstrap (MBB: [Lahiri & Lahiri, 2003](#)), applied to the remainder. There are several discussions in the literature about this procedure, mainly regarding the size of the blocks. MBB resampling the collection of overlapping (consecutive) blocks of observations. The idea is to keep the structure still present in the remainder. The forecasts are obtained via ETS methods (see Section [2.3.1](#)) and, for the final aggregation, the authors adopted the median. Their method is evaluated on the M3 data set and outperformed the original benchmarks. The work of [Bergmeir et al. \(2016\)](#) inspired many others: [Dantas, Cyrino Oliveira, and Varela Repolho \(2017\)](#) applied the idea for air transport demand data and [de Oliveira and Cyrino Oliveira \(2018\)](#) for energy consumption, proposing the so-called remainder sieve bootstrap (RSB).

[Dantas and Cyrino Oliveira \(2018\)](#) proposed an extension to the Bagged ETS where bagging and exponential smoothing are combined with clustering methods (clustering-based forecasting methods are discussed in Section [2.7.12](#)). The approach aims to consider and reduce the covariance effects among the ensemble time series, creating clusters of similar forecasts – since it could impact the variance of the group. A variety of forecasts are selected from each cluster, producing groups with reduced variance. In light of the aforementioned, there are several possibilities for each stage of the mentioned framework. In this context, to investigate the reasons why bagging works well for time series forecasting, [Petropoulos et al. \(2018a\)](#) explored three sources of uncertainty: model form, data, and parameter. While arguably bagging can handle all of them, [Petropoulos et al. \(2018a\)](#) showed that simply tackling model uncertainty is enough for achieving a superior performance, leading to the proposal of a Bootstrap Model Combination (BMC) approach, where different model forms are identified in the ensemble and fitted to the original data.

Finally, [Meira, Cyrino Oliveira, and Jeon \(2020\)](#) proposed “treating and pruning” strategies to improve the performance of prediction intervals for both model selection and forecast combinations. Testing over a large set of real time series from the M forecasting competitions (see also Section [2.12.7](#)), their results highlighted the importance of analysing the prediction intervals of the ensemble series before the final aggregation.

2.7.7. Multi-step ahead forecasting⁵⁵

Given a univariate time series comprising observations, y_t , multi-step ahead point forecasting involves producing point estimates of the future values y_{t+h} , where h is the forecast horizon (Ben Taieb, 2014).

The (naive) recursive strategy estimates a one-step-ahead autoregressive model to predict y_{t+h} from y_t , by minimising the one-step-ahead forecast errors. Each forecast is then obtained dynamically by iterating the model h times, and by plugging in the missing lagged values with their respective forecasts. The direct strategy builds separate h -step-ahead models to predict y_{t+h} from y_t for $h = 1, \dots, H$, by minimising h -step-ahead forecast errors, and forecasts are computed directly by the estimated models.

In theory, with linear models, model misspecification plays an important role in the relative performance between the recursive and direct strategy (Chevillon, 2007). If the model is correctly specified, the recursive strategy benefits from more efficient parameter estimates, while the direct strategy is more robust to model misspecification. With nonlinear models, recursive forecasts are known to be asymptotically biased (Fan and Yao, 2005, Lin and Granger, 1994, Teräsvirta et al., 2010), and the direct strategy is often preferred over the recursive strategy since it avoids the accumulation of forecast errors. In practice, the results are mixed (Atiya et al., 1999, Kline, 2004, Marcellino et al., 2006, Pesaran et al., 2011, Sorjamaa et al., 2007), and depend on many interacting factors including the model complexity (see also Section 2.5.2), the (unknown) underlying data generating process, the number of observations, and the forecast horizon (see also Section 2.7.4).

Hybrids and variants of both recursive and direct strategies have been proposed in the literature. For example, one of the hybrid strategies (Sorjamaa and Lendasse, 2006, Zhang and Hutchinson, 1994, Zhang et al., 2013) first produce recursive forecasts, then adjust these forecasts by modelling the multi-step forecast errors using a direct strategy (Ben Taieb & Hyndman, 2014). Variants of the recursive strategy match the model estimation and forecasting loss functions by minimising the implied h -step-ahead recursive forecast errors (Bhansali and Kokoszka, 2002, Bontempi et al., 1999, McNames, 1998). Variants of the direct strategy exploit the fact that the errors of different models are serially correlated (Chen et al., 2004, Franses and Legerstee, 2009c, Lee and Billings, 2003, Pesaran et al., 2011). The idea is to reduce the forecast variance of independently selected models by exploiting the relatedness between the forecasting tasks, as in multi-task learning (Caruana, 1997). For example, a multi-

horizon strategy will measure forecast accuracy (see Section 2.12.2) by averaging the forecast errors over multiple forecast horizons (Bontempi and Ben Taieb, 2011, Kline, 2004). Different multi-horizon strategies can be specified, with different formulation of the objective function (Ben Taieb, Sorjamaa, & Bontempi, 2010). One particular case is the multi-output strategy which estimates a single model for all horizons by minimising the average forecast error over the entire forecast horizon (Bontempi & Ben Taieb, 2011).

Forecasting strategies are often model-dependent, especially with machine learning models (see Section 2.7.10). Furthermore, model architecture and parameters are often trained by taking into account the chosen forecasting strategy. For example, we can naturally implement and train recursive forecasting models using recurrent neural networks (see also Section 2.7.8). Also, different specifications of the decoder in sequence-to-sequence models will induce different forecasting strategies, including variants of direct and multi-horizon strategies. For more details, we refer the reader to Hewamalage, Bergmeir, and Bandara (2021) and Section 4.2 in Benidis et al. (2020). Which forecasting strategy is best is an empirical question since it involves a tradeoff between forecast bias and variance (Ben Taieb and Atiya, 2015, Ben Taieb et al., 2012). Therefore, the forecasting strategy should be part of the design choices and the model selection procedure of any multi-step-ahead forecasting model.

2.7.8. Neural networks⁵⁶

Neural Networks (NNs) or Artificial Neural Networks (ANNs) are mathematical formulations inspired by the work and functioning of biological neurons. They are characterized by their ability to model non-stationary, nonlinear and high complex datasets. This property along with the increased computational power have put NNs in the frontline of research in most fields of science (De Gooijer and Hyndman, 2006, Zhang et al., 1998).

A typical NN topology is consisted by three types of layers (input, hidden and output) and each layer is consisted by nodes. The first layer in every NN, is the input layer and the number of its nodes corresponds to the number of explanatory variables (inputs). The last layer is the output layer and the number of nodes corresponds to the number of response variables (forecasts). Between the input and the output layer, there is one or more hidden layers where the nodes define the amount of complexity the model is capable of fitting. Most NN topologies in the input and the first hidden layer contain an

extra node, called the bias node. The bias node has a fixed value of one and serves a function similar to the intercept in traditional regression models. Each node in one layer has connections (weights) with all or a subset (for example, for the convolutional neural network topology) of the nodes of the next layer.

NNs process the information as follows: the input nodes contain the explanatory variables. These variables are weighted by the connections between the input and the first hidden nodes, and the information reaches to the hidden nodes as a weighted sum of the inputs. In the hidden nodes, there is usually a non-linear function (such as the sigmoid or the ReLU) which transform the information received. This process is repeated until the information reaches the output layer as forecasts. NNs are trained by adjusting the weights that connect the nodes in a way that the network maps the input value of the training data to the corresponding output value. This mapping is based on a loss function, the choice of which depends on the nature of the forecasting problem. The most common NN procedure, is the back-propagation of errors (for additional details on training, see Section 2.7.11).

The simpler and most common NN topology, is the Multilayer Forward Perceptron (MLP). In MLP, the hidden nodes contain the sigmoid function and the information moves in forward direction (from the inputs to the output nodes). An another NN topology where the information moves also only in a forward direction is the Radial Basis Function NN (RBF). Now the hidden neurons compute the Euclidean distance of the test case from the neuron's centre point and then applies the Gaussian function to this distance using the spread values. Recurrent Neural Networks (RNNs) are NN topologies that allow previous outputs to be used as inputs while having hidden states. The information moves both forwards and backwards. RNNs have short-term memory and inputs are taken potentially from all previous values. MLPs, RBFs and RNNs are universal function approximators (Hornik, 1991, Park and Sandberg, 1991, Schäfer and Zimmermann, 2006). However, the amount of NN complexity in terms of hidden layers and nodes to reach this property, might make the NN topology computationally unfeasible to train (see also the discussion in Section 2.7.11). For the interaction of NNs with the probability theory, we refer the reader to last part of Section 2.7.9.

2.7.9. Deep probabilistic forecasting models⁵⁷

Neural networks (Section 2.7.8) can be equipped to provide not only a single-valued forecast, but rather the entire range of values possible in a number of ways (see also

Sections 2.6.2 [Density forecast combinations](#), 2.6.3 [Ensembles and predictive probability post processors](#)). We will discuss three selected approaches in the following, but remark that this is a subjective selection and is by far not comprehensive.⁵⁸

- 1.
Analogously to linear regression and Generalised Linear Models, obtaining probabilistic forecasts can be achieved by the neural network outputting not the forecasted value itself but rather parameters of probability distribution or density ([Bishop, 2006](#)). In forecasting, a prominent example is the DeepAR model ([Salinas, Valentin, Jan & Januschowski, 2019](#)), which uses a recurrent neural network architecture and assumes the probability distribution to be from a standard probability density function (e.g., negative binomial or Student's). Variations are possible, with either non-standard output distributions in forecasting such as the multinomial distribution ([Rabanser, Januschowski, Flunkert, Salinas, & Gasthaus, 2020](#)) or via representing the probability density as cumulative distribution function ([Salinas, Michael et al., 2019](#)) or the quantile function ([Gasthaus et al., 2019](#)).
- 2.
An alternative approach is to apply concepts for quantile regression ([Koenker, 2005](#)) to neural networks, e.g., by making the neural network produce values for selected quantiles directly ([Wen, Torkkola, Narayanaswamy, & Madeka, 2017](#)).
- 3.
It is possible to combine neural networks with existing probabilistic models. For example, neural networks can parametrise state space models ([Durbin & Koopman, 2012](#)) as an example for another class of approaches ([Rangapuram, Seeger, Gasthaus, Stella, Wang, & Januschowski, 2018](#)), dynamic factor models ([Geweke, 1977](#)) with neural networks ([Wang, Smola, Maddix, Gasthaus, Foster, & Januschowski, 2019](#)) or deep temporal point processes ([Turkmen, Wang, & Januschowski, 2019](#)).

The appeals of using neural networks for point forecasts carry over to probabilistic forecasts, so we will only comment on the elegance of modern neural network programming frameworks. To the forecasting model builder, the availability of auto gradient computation, the integration with highly-tuned optimisation algorithms and scalability considerations built into the frameworks, means that the time from model

idea to experimental evaluation has never been shorter. In the examples above, we brushed over the need to have loss functions with which we estimate the parameters of the neural networks. Standard negative log-likelihood based approaches are easily expressible in code as are approaches based on non-standard losses such as the continuous ranked probability score ([Gneiting et al., 2007](#) and Section 2.12.4). With open source proliferating in the deep learning community, most of the above examples for obtaining probabilistic forecasts can readily be test-driven (see, for example, [Alexandrov et al., 2019](#)).

For the future, we see a number of open challenges. Most of the approaches mentioned above are univariate, in the following sense. If we are interested in forecasting values for all time series in a panel, we may be interested in modelling the relationship among these time series. The aforementioned approaches mostly assume independence of the time series. In recent years, a number of multivariate probabilistic forecasting models have been proposed ([Rangapuram et al., 2020](#), [Salinas, Michael et al., 2019](#)), but much work remains to obtain a better understanding. Another counter-intuitive challenge for neural networks is to scale them down. Neural networks are highly parametrised, so in order to estimate parameters correctly, panels with lots of time series are needed. However, a large part of the forecasting problem landscape ([Januschowski & Kolassa, 2019](#)) consists of forecasting problems with only a few time series. Obtaining good uncertainty estimates with neural networks in these settings is an open problem.

2.7.10. Machine learning⁵⁹

Categorising forecasting methods into statistical and machine learning (ML) is not trivial, as various criteria can be considered for performing this task ([Januschowski et al., 2020](#)). Nevertheless, more often than not, forecasting methods are categorised as ML when they do not prescribe the data-generating process, e.g., through a set of equations, thus allowing for data relationships to be automatically learned ([Barker, 2020](#)). In this respect, methods that build on unstructured, non-linear regression algorithms (see also Section 2.3.2), such as Neural Networks (NN), Decision Trees, Support Vector Machines (SVM), and Gaussian Processes, are considered as ML ([Makridakis et al., 2018](#)).

Since ML methods are data-driven, they are more generic and easier to be adapted to forecast series of different characteristics ([Spiliotis, Kouloumos, Assimakopoulos & Makridakis, 2020](#)). However, ML methods also display some limitations. First, in order

for ML methods to take full advantage of their capacity, sufficient data are required. Thus, when series are relatively short and display complex patterns, such as seasonality and trend, ML methods are expected to provide sub-optimal forecasts if the data are not properly pre-processed (Makridakis et al., 2018, Zhang et al., 1998). On the other hand, when dealing with long, high-frequency series, typically found in energy (Chae, Horesh, Hwang, & Lee, 2016 but also Section 3.4), stock market (Moghaddam, Moghaddam, & Esfandyari, 2016 and Section 3.3), and demand (Carmo & Rodrigues, 2004 but also Section 3.2) related applications, ML methods can be applied with success. Second, computational intensity may become relevant (Makridakis, Spiliotis et al., 2020), especially when forecasting numerous series at the weekly and daily frequency (Seaman, 2018) or long-term accuracy improvements over traditional methods are insignificant (Nikolopoulos & Petropoulos, 2018). Third, given that the effective implementation of ML methods strongly depends on optimally determining the values of several hyper-parameters, related both with the forecasting method itself and the training process, considerable complexity is introduced, significant resources are required to set up the methods, and high experience and a strong background in other fields than forecasting, such as programming and optimisation, are needed. In order to deal with these limitations, ML methods can be applied in a cross-learning (CL) fashion instead of a series-by-series one (Makridakis, Spiliotis et al., 2020), i.e., allow the methods to learn from multiple series how to accurately forecast the individual ones (see also Section 2.12.7). The key principle behind CL is that, although series may differ, common patterns may occur among them, especially when data are structured in a hierarchical way and additional information, such as categorical attributes and exogenous/explanatory variables (see Section 2.2.5), is provided as input (Fry & Brundage, 2020). The CL approach has several advantages. First, computational time can be significantly reduced as a single model can be used to forecast multiple series simultaneously (Semenoglou, Spiliotis, Makridakis, & Assimakopoulos, 2021). Second, methods trained in a particular dataset can be effectively used to provide forecasts for series of different datasets that display similar characteristics (transfer-learning), thus allowing the development of generalised forecasting methods (Oreshkin, Carпов, Chapados, & Bengio, 2020a). Third, data limitations are mitigated and valuable information can be exploited at global level, thus allowing for patterns shared among the series, such as seasonal cycles (Dekker, van

Donselaar, & Ouwehand, 2004) and special events (Huber & Stuckenschmidt, 2020), to be effectively captured.

Based on the above, CL is currently considered the most effective way of applying ML for batch time series forecasting. Some state-of-the-art implementations of CL include long short-term memory NNs (Smyl, 2020), deep NNs based on backward and forward residual links (Oreshkin, Carпов, Chapados, & Bengio, 2020b), feature-based XGBoost (Montero-Manso et al., 2020), and gradient boosted decision trees (Bojer & Meldgaard, 2020).

2.7.11. Machine learning with (very) noisy data⁶⁰

With the advent of big data, machine learning now plays a leading role in forecasting.⁶¹ There are two primary reasons for this. First, conventional ordinary least squares (OLS) estimation is highly susceptible to *overfitting* in the presence of a large number of regressors (or features); see also Sections 2.5.2 Model complexity, 2.5.3 Variable selection. OLS maximises the fit of the model over the estimation (or training) sample, which can lead to poor out-of-sample performance; in essence, OLS over-responds to *noise* in the data, and the problem becomes magnified as the number of features grows. A class of machine-learning techniques, which includes the popular least absolute shrinkage and selection operator (LASSO: Tibshirani, 1996) and elastic net (ENet: Zou & Hastie, 2005), employs *penalised regression* to improve out-of-sample performance with large numbers of features. The LASSO and ENet guard against overfitting by *shrinking* the parameter estimates toward zero.

Very noisy data – data with a very low signal-to-noise ratio – exacerbate the overfitting problem. In such an environment, it is vital to induce adequate shrinkage to guard against overfitting and more reliably uncover the predictive signal amidst all the noise. For LASSO and ENet estimation, a promising strategy is to employ a stringent information criterion, such as the Bayesian information criterion (BIC, Schwarz, 1978), to select (or tune) the regularisation parameter governing the degree of shrinkage (often denoted by λ). Fan and Tang (2013) and Wang, Li, and Leng (2009) modify the BIC penalty to account for a diverging number of features, while Hui, Warton, and Foster (2015) refine the BIC penalty to include the value of λ . These BIC variants induce a greater degree of shrinkage by strengthening the BIC's penalty term, making them useful for implementing the LASSO and ENet in noisy data environments; see Filippou, Rapach, Taylor, and Zhou (2020) for a recent empirical application.

A second reason for the popularity of machine learning in the era of big data is the existence of powerful tools for accommodating complex predictive relationships. In many contexts, a linear specification appears overly restrictive, as it may neglect important nonlinearities in the data that can potentially be exploited to improve forecasting performance. Neural networks (NNs; see Section 2.7.8) are perhaps the most popular machine-learning device for modelling nonlinear predictive relationships with a large number of features. Under a reasonable set of assumptions, a sufficiently complex NN can approximate any smooth function (for example, [Barron, 1994](#), [Cybenko, 1989](#), [Funahashi, 1989](#), [Hornik et al., 1989](#)).

By design, NNs are extremely flexible, and this flexibility means that a large number of parameters (or weights) need to be estimated, which again raises concerns about overfitting, especially with very noisy data. The weights of a NN are typically estimated via a stochastic gradient descent (SGD) algorithm, such as *Adam* ([Kingma & Ba, 2015](#)). The SGD algorithm itself has some regularising properties, which can be strengthened by adjusting the algorithm's hyperparameters. We can further guard against overfitting by shrinking the weights via LASSO or ENet penalty terms, as well as imposing a dropout rate ([Hinton et al., 2012](#), [Srivastava et al., 2014](#)).

Perhaps the quintessential example of a noisy data environment is forecasting asset returns, especially at short horizons (e.g., monthly). Because many asset markets are reasonably efficient, most of the fluctuations in returns are inherently unpredictable – they reflect the arrival of new information, which, by definition, is unpredictable. This does not mean that we should not bother trying to forecast returns, as even a seemingly small degree of return predictability can be economically significant (e.g., [Campbell & Thompson, 2008](#)). Instead, it means that we need to be particularly mindful of overfitting when forecasting returns in the era of big data. Section 3.3.13 discusses applications of machine-learning techniques for stock return forecasting.

2.7.12. Clustering-based forecasting⁶²

The robustness of the forecasting process depends mainly on the characteristics of the target variable. In cases of high nonlinear and volatile time series, a forecasting model may not be able to fully capture and simulate the special characteristics, a fact that may lead to poor forecasting accuracy ([Pradeepkumar & Ravi, 2017](#)). Contemporary research has proposed some approaches to increase the forecasting performance ([Sardinha-Lourenço, Andrade-Campos, Antunes, & Oliveira, 2018](#)). Clustering-based forecasting

refers to the application of unsupervised machine learning in forecasting tasks. The scope is to increase the performance by employing the information of data structure and of the existing similarities among the data entries (Goia, May, & Fusai, 2010); see also Sections 2.7.4 Feature-based time series forecasting, 2.7.10 Machine learning.

Clustering is a proven method in pattern recognition and data science for deriving the level of similarity of data points within a set. The outputs of a clustering algorithm are the centroids of the clusters and the cluster labels, i.e., integer numbers that denote the number of cluster that a specific data entry belongs to (Xu & Wunsch, 2005).

There are two approaches in clustering-based forecasting: (i) Combination of clustering and supervised machine learning, and (ii) solely application of clustering. In the first case, a clustering algorithm is used to split the training set into smaller sub-training sets. These sets contain patterns with high similarity. Then for each cluster a dedicated forecaster is applied (Chaouch, 2014, Fan et al., 2008). Thus, the number of forecasting algorithms is equal to the number of clusters. This approach enables to train forecasters with more similar patterns and eventually achieve better training process. The forecasting systems that involve clustering are reported to result in lower errors (Fan et al., 2006, Mori and Yuihara, 2001). The combination of clustering and forecasting has been presented in the literature earlier than the sole application of clustering. One of the first articles in the literature on combining clustering and forecasting sets up the respective theoretical framework (Kehagias & Petridis, 1997).

In the second case, a clustering algorithm is used to both cluster the load data set and perform the forecasting (López, Valero, Senabre, Aparicio, & Gabaldon, 2012). In the sole clustering applications, either the centroids of the clusters can be utilised or the labels. Pattern sequence-based forecasting is an approach that employs the cluster labels. In this approach, a clustering of all days prior to the test day is held and this results in sequences of labels of a certain length. Next, the similarity of the predicted day sequence with the historical data sequences is examined. The load curve of the predicted day is the average of the curves of the days following the same sequences (Kang, Spiliotis et al., 2020, Martinez Alvarez et al., 2011).

There is variety of clustering algorithms that have been proposed in forecasting such as the -means, fuzzy C-means (FCM), self-organising map (SOM), deterministic annealing (DA), ant colony clustering ACC, and others. Apart from the clustering effectiveness, a selection criterion for an algorithm is the complexity. -means and FCM are less complex

compared to the SOM that needs a large number of variables to be calibrated prior to its application. Meta-heuristics algorithms, such as ACC, strongly depend on the initialisation conditions and the swarm size. Therefore, a comparison of clustering algorithms should take place to define the most suitable one for the problem under study (Elangasinghe et al., 2014, Li et al., 2008, Mori and Yuihara, 2001, Wang et al., 2015).

The assessment of clustering-based forecasting is held via common evaluation metrics for forecasting tasks (see also Section 2.12). The optimal number of clusters, which is a crucial parameter of a clustering application, is selected via trial-and-error, i.e., the optimal number corresponds to the lowest forecasting error (Nagi, Yap, Nagi, Tiong, & Ahmed, 2011).

2.7.13. Hybrid methods⁶³

Hybrid approaches combine two or more of the above-mentioned advanced methods. In general, when methods based on AI-based techniques, physical, and statistical approaches are combined together, the result is often improved forecasting accuracy as a benefit from the inherent integration of the single methods. The idea is to mix diverse methods with unique features to address the limitations of individual techniques, thus enhancing the forecast performance (Mandal et al., 2012, Nespoli et al., 2019 see also Section 2.6). The performance of the hybrid methods depends on the performance of the single methods, and these single methods should be specifically selected for the problem that has to be addressed.

Hybrid methods can be categorised based on the constituent methods, but also considering that these base methods may not necessarily act only on the forecasting stage but also on data treatment and parameters identification stages. In data pre-processing combining approaches (see also Section 2.2), different methods can be used for decomposing the time series into subseries (Son, Yang, & Na, 2019) or the signal into different frequencies (Zang, Cheng, Ding, Cheung, Liang, Wei, & Sun, 2018), and for classifying the historical data (Huang, Chen, Yang & Kuo, 2015). An advantage of such hybrid methods is robustness against sudden changes in the values of the main parameters. However, they require additional knowledge and understanding of the base methods, and have the disadvantage of slow response time to new data.

The purpose of the parameter selection stage is to optimise the parameters of the model, in terms of extracting nonlinear features and invariant structures (Behera et al.,

2018, Ogliari et al., 2018) but also in terms of estimation of the parameter adopted for the prediction; for example, meteorological factors such as temperature, humidity, precipitation, snowfall, cloud, sunshine, wind speed, and wind direction (Qu, Kang, Zhang, Jiang, & Ma, 2016). Hybrid methods feature straightforward determination of the parameters with relatively basic structures. However, the implementation is sometimes challenging, and depends on the knowledge and expertise of the designer. Finally, the data post-processing hybrid approaches forecast the residual errors resulted from the forecasting model. Since these hybrid methods consider residual errors from the model, they aim in further improving the predictions of the base methods by applying corrections in the forecasts. However, a disadvantage of these hybrid methods is the increased calculation time, as the residual errors must also be estimated. Also, such hybrid methods are not general and will depend on the field of application. In many cases, hybrids approaches outperform other (single) approaches such as NN, NN, and ARIMA-based models (Mellit, Massi Pavan, Ogliari, Leva, & Lugh, 2020). A great example is the hybrid method by Smyl (2020), which achieved the best performance in the M4 forecasting competition (see also Section 2.12.7). In particular, in energy applications (see Section 3.4), a combination of physical and AI-based techniques can lead to improved forecasting performance. Furthermore, machine learning methods (see Section 2.7.10) based on historical data of meteorological variables combined with an optimal learning algorithm and weather classification can further improve the forecasting accuracy of single methods. However, in general, the weak point of such hybrid approaches is that they underperform when meteorological conditions are unstable (Chicco, Cocina, Di Leo, Spertino, & Massi Pavan, 2015).

2.8. Methods for intermittent demand

2.8.1. Parametric methods for intermittent demand forecasting⁶⁴

Demand forecasting is the basis for most planning and control activities in any organisation. Demand will typically be accumulated in some pre-defined ‘time buckets’ (periods), such as a day, a week or a month. On many occasions, demand may be observed in every time period, resulting in what is sometimes referred to as ‘non-intermittent demand’. Alternatively, demand may appear sporadically, with no demand at all in some periods, leading to an intermittent appearance of demand occurrences.

Intermittent demand items monopolise the stock bases in the after sales industry and are prevalent in many other industries, including the automotive, IT, and electronics sectors. Their inventory implications are dramatic and forecasting their requirements is a very challenging task.

Methods to forecast intermittent demand may broadly be classified as parametric and non-parametric. The former suppose that future demand can be well represented by a statistical distribution (say Poisson or Negative Binomial) which has parameters that are unknown but may be forecasted using past data. These methods are discussed in this sub-section. In the latter, the data are not assumed to follow any standard probability distribution. Instead, direct methods are used to assess the distributions required for inventory management (see also Section 3.2.3). Such methods are discussed in Section 2.8.2.

Simple Exponential Smoothing (SES; see Section 2.3.1) is often used in practice to forecast intermittent demand series. However, SES fails to recognise that intermittent demand is built from two constituent elements: (i) the inter-demand intervals, which relate to the probability of demand occurring, and (ii) the demand sizes, when demand occurs. The former indicates the degree of intermittence, whereas the latter relates to the behaviour of the positive demands. Croston (1972) showed that this inherent limitation leads to SES being (positively) biased after a demand occurring period; this is sometimes referred to as an ‘issue point’ bias. Subsequently, he proposed a method that forecasts separately the sizes of demand, when demand occurs, and the inter-demand intervals. Both forecasts are produced using SES, and the ratio of the former over the latter gives a forecast of the mean demand per period. Croston’s method was shown by Syntetos and Boylan (2001) to suffer from another type of bias (inversion bias) and the same researchers (Syntetos & Boylan, 2005) proposed a modification to his method that leads to approximately unbiased estimates. This method is known in the literature as the Syntetos-Boylan Approximation (SBA). It has been found repeatedly to account for considerable empirical inventory forecasting improvements (Eaves and Kingsman, 2004, Gutierrez et al., 2008, Nikolopoulos et al., 2016, van Wingerden et al., 2014) and, at the time of writing, it constitutes the benchmark against which other (new) proposed methodologies in the area of intermittent demand forecasting are assessed.

Croston’s method is based upon the assumption of a Bernoulli demand arrival process. Alternatively, demand may be assumed to arrive according to a Poisson process. It is

also possible to adapt Croston's method so that sizes and intervals are updated based on a simple moving average (SMA) procedure instead of SES. [Boylan and Syntetos \(2003\)](#), [Shale, Boylan, and Johnston \(2006\)](#), and [Syntetos, Babai, and Luo \(2015\)](#) presented correction factors to overcome the bias associated with Croston's approach under a Poisson demand arrival process and/or estimation of demand sizes and intervals using an SMA.

For a detailed review of developments in intermittent demand forecasting interested readers are referred to [Boylan and Syntetos \(2021\)](#).

2.8.2. Non-parametric intermittent demand methods⁶⁵

Two main non-parametric forecasting approaches have dominated the intermittent demand literature: the bootstrapping approach and the Overlapping/Non-Overlapping aggregation Blocks approach ([Boylan & Syntetos, 2021](#)).

The bootstrapping approach relies upon a resampling (with or without replacement) of the historical demand data to build the empirical distribution of the demand over a specified interval. As discussed in Section 2.7.5, this approach was initially introduced by [Efron \(1979\)](#). Since then, it has been developed by [Willemain, Smart, and Schwarz \(2004\)](#) and [Zhou and Viswanathan \(2011\)](#) to deal with intermittent demand items ([Babai, Tsadiras, & Papadopoulos, 2020](#)). [Willemain et al. \(2004\)](#) have proposed a method that resamples demand data by using a Markov chain to switch between no demand and demand periods. The empirical outperformance of this method has been shown when compared to Simple Exponential Smoothing (SES) and Croston's method (see also Section 2.8.1). However, the findings of [Willemain et al. \(2004\)](#)'s work have been challenged by [Gardner and Koehler \(2005\)](#) and some limitations have been addressed by [Syntetos, Zied Babai and Gardner \(2015\)](#). [Zhou and Viswanathan \(2011\)](#) have developed an alternative bootstrapping method. Their method samples separately demand intervals and demand sizes and it has been shown to be associated with a good performance for long lead-times. [Teunter and Duncan \(2009\)](#) and [Hasni, Aguir, Babai, and Jemai \(2019a\)](#) have developed adjustments of the bootstrapping methods, where the lead-time demand forecast is adjusted by assuming that the first period in the lead-time bucket corresponds to a non-zero demand. They have demonstrated the outperformance of the adjusted bootstrapping methods in a periodic order-up-to-level inventory control system. A review of the bootstrapping methods in

the context of intermittent demand is provided by [Hasni, Aguir, Babai, and Jemai \(2019b\)](#).

[Porras and Dekker \(2008\)](#) were the first to consider aggregation with overlapping and non-overlapping blocks (OB and NOB) approach in forecasting the demand of spare parts. In the NOB approach, a demand series is divided into consecutive non-overlapping blocks of time, whereas in OB, at each period the oldest observation is dropped and the newest is included ([Rostami-Tabar, Babai, Syntetos, & Ducq, 2013](#)). [Boylan and Babai \(2016\)](#) have compared the statistical and inventory performance of the OB and NOB methods. They found that, unless the demand history is short, there is a clear advantage of using OB instead of NOB. More recently, based on extreme value theory (EVT), [Zhu, Dekker, van Jaarsveld, Renjie, and Koning \(2017\)](#) have proposed an improvement of the OB method that models better the tail of lead-time demand. They have shown that the empirical-EVT method leads to higher achieved target cycle service levels when compared to the original method proposed by [Porras and Dekker \(2008\)](#). Temporal aggregation is further discussed in Section 2.10.2.

2.8.3. Classification methods⁶⁶

In many application areas, forecasts are required across a wide collection of products, services or locations. In this situation, it is convenient to introduce classification rules that allow subsets of time series to be forecasted using the same approaches and methods. Categorisation rules, such as the ABC inventory classification, serve the forecasting function only coincidentally. They do not necessarily align to the selection of the best forecasting method.

Within certain modelling frameworks, classification of time series is well established. For example, within an ARIMA framework ([Box, Jenkins, & Reinsel, 2008](#) and Section 2.3.4), or within a state-space framework for exponential smoothing ([Hyndman et al., 2002](#) and Section 2.3.1), series may be classified, for example based on the AIC ([Akaike, 1973](#)). It is more challenging to classify series according to their recommended forecasting method if some of the candidate methods, such as Croston's method (see Section 2.8.1), lack a fully satisfactory model base. In the field of intermittent demand forecasting, [Syntetos, Boylan, and Croston \(2005\)](#) proposed the SBC classification scheme, enabling time series to be classified according to their length of average demand interval and coefficient of variation of demand sizes (when demand

occurs). These rules were based on assumptions of independent and identically distributed (iid) demand, and a comparison of expected mean square error between methods. The scheme has been extended by [Kostenko and Hyndman \(2006\)](#) and by [Petropoulos and Kourentzes \(2015\)](#). In an empirical case-study, [Boylan, Syntetos, and Karakostas \(2008\)](#) examined series not necessarily conforming to iid assumptions and found the rules to be robust to inexact specification of cut-off values. [Moon, Simpson, and Hicks \(2013\)](#) used logistic regression to classify time series of demand for spare parts in the South Korean Navy. The classification was designed to identify superior performance (accuracy and inventory costs) of direct and hierarchical forecasting methods, based on the serial correlation of demands, the coefficient of variation of demand volume of spare parts (see also Section [3.2.7](#)), and the functionality of the naval equipment.

[Bartezzaghi, Verganti, and Zotteri \(1999\)](#) identified five factors that contribute towards intermittence and ‘lumpiness’ (intermittence with highly variable demand sizes): number of potential customers, frequency of customers’ requests, heterogeneity of customers, variety of individual customer’s requests, and correlations between customers’ requests. These may contribute towards useful classifications, for example by the number of customers for an item. When this number is low and some of the customers are large, then direct communication with the customers can inform judgmental forecasts. Similarly, if a customer’s requests are highly variable, then ‘advance demand information’ from customers can help to improve judgmental estimates. These strategies can be very useful in a business-to-business environment, where such strategies are feasible.

An alternative approach to classification is combination of forecasts (see Section [2.6](#) for a review, and Section [2.6.1](#) in particular). [Petropoulos and Kourentzes \(2015\)](#) investigated combining standard forecasting methods for intermittent demand (e.g., SES, Croston, Syntetos-Boylan Approximation; see also Section [2.8.1](#)). They did not find this to improve accuracy directly, but obtained good results from the use of combinations of forecasts at different temporal frequencies, using methods selected from the extended SBC classification scheme.

2.8.4. Peak over the threshold⁶⁷

In the forecasting literature, [Nikolopoulos \(2020\)](#) argues that great attention has been given to modelling fast-moving time series with or without cues of information

available (Nikolopoulos, Goodwin, Patelis, & Assimakopoulos, 2007). Less attention has been given to intermittent/count series (see Sections 2.3.8 Forecasting count time series, 2.8.1 Parametric methods for intermittent demand forecasting, 2.8.2 Non-parametric intermittent demand methods), which are more difficult to forecast given the presence of two sources of uncertainty: demand volume, and timing.

Historically there have been few forecasting methods developed specifically for such data (Syntetos, Zied Babai & Gardner, 2015). We believe that through a time series decomposition approach á la Leadbetter (1991) we can isolate ‘peaks over threshold’ (POT) data points, and create new intermittent series from any time series of interest. The derived series present almost identical characteristics with the series that Croston (1972) analysed. In essence one could use such decomposition forecasting techniques to tackle much more difficult phenomena and problems coming from finance, politics, healthcare, humanitarian logistics, business, economics, and social sciences.

Any time series can be decomposed into two sub-series: one containing the baseline (*white swans*) and one containing the extreme values over an arbitrary-set or rule-based-set threshold (*grey* and *black swans*) as proposed by Taleb (2008); see also Section 2.3.22. Unfortunately, major decision-related risks and most of the underlying uncertainty lay with these extremes. So, it is very important to be able to effectively model and forecast them.

It is unlikely that any forecasting approach will accurately give the exact timing of the forthcoming extreme event, but it will instead provide a satisfactory cumulative forecast over a long period of time. The question still stands what can one do with this forecast? For earthquake data, although even if we know that a major earthquake is going to hit a region, it is almost impossible to decide to evacuate cities, but still we can influence and legislate the way structures are built and increase the awareness, training, preparedness and readiness of the public; and also ensure enough capital on hold to cope with the aftermath of the major event. For epidemics/pandemics (see Section 3.6.2) there are clear implications, as we have evidenced with COVID-19, on how proactively we can source and secure human resources, medical supplies, etc.

What is the current doctrine when forecasting in such a context: advanced probabilistic models. These methods typically require a lot of data and reconstruct the distributions of the underlying phenomena. These come with common successes and a plethora of constraints: *big data sets* needed for training the models, high

mathematical *complexity*, and invisibility to practitioners how these methods do actually work and thus *less acceptance in practice*. Yet again, forecasting accuracy is the name of the game and thus these forecasting methods are serious contenders for the task in hand.

Extreme Value Theory (EVT) analyses extreme deviations from statistical measures of central location to estimate the probability of events that are more extreme than anything observed in the time series. This is usually done in the following two ways (Nikolopoulos, 2020): (i) deriving maxima and/or minima series as a first step and then having the Generalised Extreme Value Distribution fitted (often the number of extreme events is limited), and (ii) isolating the values that exceed a threshold (point over threshold) that can also lead to only a few instances extracted – so a very intermittent series in nature. The analysis involves fitting a Poisson distribution for the number of events in a basic time period and a second distribution – usually a Generalised Pareto Distribution – for the size of the resulting POT values.

2.9. Reasoning and mining

2.9.1. Fuzzy logic⁶⁸

The “classical” Boolean logic is not able to handle for uncertainties and/or vagueness that are necessary when dealing with many real world problems. This is in part due to the fact that the Boolean logic is based on only two values (i.e., a statement can only be true or false). Fuzzy logic tries to overcome this issue by admitting that a statement/variable could be partially true or partially false. Mathematically, the fuzzy logic framework is based on the work of Zadeh (1965) who introduced the theory of *fuzzy sets*. The main point of this theory is the definition of two kinds of sets:

- 1.
Crisp sets are the “classical” sets in the Boolean logic. An element can belong (or not) to a certain set.
- 2.
Fuzzy sets, where an element can belong to the sets with a certain *membership grade*, with a value that varies in the interval .

The definition of the fuzzy sets allows the framework to take into account the uncertainty and vagueness of information. An extension of this approach is related to

the fact that a certain variable can assume a crisp value (classical theory) or can belong to different fuzzy sets with different membership grade. For example, in a system implemented to forecast the daily pollutant concentration in atmosphere, one of the inputs could relate to the weather conditions, such as the wind speed. In the classical approach, the system must have as an input the value of the wind speed at a certain day. In the fuzzy approach, the input of the system could be the membership grade of the input variable to three different fuzzy sets: (i) “Not windy”, (ii) “average windy”, and (iii) “strong windy”. On the other hand, the user of the forecasting system may be only interested in a classification of the output variable instead of the crisp value. In this case, the fuzzy approach is applied to the pollutant concentration which could belong with a certain degree to the fuzzy sets (i) “not polluted day”, (ii) “medium polluted day”, (iii) “high polluted day”, and (iv) “critical polluted day”.

In fuzzy theory, each fuzzy set is characterised by a (generally nonlinear) function, called the *membership function*, linking crisp values to the membership of the different sets. The association of a crisp value to its membership for a set is called *fuzzyfication*, while the inverse operation (from a membership value to a crisp value) is called *defuzzification*. As with the logic theory, the *inference system* assumes a key role in the fuzzy theory. A Fuzzy Inference System (FIS) allows the interpretation of the membership grades of the input variable(s) and, given some sets of fuzzy rules, assigns the corresponding values to the output variable(s). In the literature, two main fuzzy inference systems are presented:

- 1.
Mamdani system ([Mamdani & Assilian, 1975](#)), where both the input and output of the inference system are membership functions.
- 2.
Sugeno system ([Sugeno, 1985](#)), where the output of the inference system is a crisp value, usually obtained by applying a linear function to the defuzzified value of the input.

2.9.2. Association rule mining⁶⁹

Association rule mining is an exploratory data-driven approach which is able to automatically and exhaustively extract all existing correlations in a data set of categorical features. It is a powerful but computationally intensive technique,

successfully applied in different forecasting contexts (Acquaviva et al., 2015, Apiletti and Pastor, 2020, Di Corso et al., 2018). Its results are in a human-readable form.

The data set must be in the form of transactions, i.e., a collection of events, each described by categorical features. If the phenomena under analysis are modelled by continuous-valued variables, discretisation can be applied to obtain a suitable data set. Association rule mining core task is the frequent itemset extraction, which consists in finding frequently-occurring relationships among items in a data set (Han, Pei, & Kamber, 2011). Given a data set of records characterised by several attributes, an item refers to a pair of (attribute = value), while a set of items is called itemset. The support count of an itemset is the number of records containing that itemset. The support of an itemset is the percentage of records containing it with respect to the total number of records in the data set. An itemset is frequent when its support is greater than or equal to a minimum support threshold.

An association rule is an implication in the form $X \rightarrow Y$, where X and Y are disjoint itemsets (i.e., $X \cap Y = \emptyset$) (Tan, Steinbach, & Kumar, 2005). X is called rule body or antecedent and Y rule head or consequent.

To evaluate the quality of an association rule, the support, confidence, and lift metrics are commonly exploited (Han et al., 2011). *Rule support* is the fraction of records containing both X and Y , indicating the probability that a record contains every item in these itemsets. The support of the rule is computed as the support of the union of X and Y . *Rule confidence* represents the strength of the implication, and is the conditional probability that a transaction containing X also contains Y , i.e., the proportion of records that contain X with respect to those that also contain Y .

Finally, the *lift* of a rule measures the correlation between antecedent and consequent. It is defined as the ratio between the rule confidence and the support of X . A lift ratio equal to 1.0 implies that itemsets X and Y are not correlated. A lift higher than 1.0 indicates a positive correlation, meaning that the occurrence of X likely leads to the occurrence of Y with the given confidence. The greater the lift, the stronger the association. Finally, a lift lower than 1.0 indicates a negative correlation between X and Y .

The problem of association rule mining consists in the extraction of all the association rules having rule support and confidence greater than the respective support and confidence thresholds, σ and γ , defined as parameters of the mining process (Tan et al., 2005). These thresholds allow to control the statistical relevance of the extracted rules.

The process of rule mining consists of two steps. The first step is the computation of frequent itemsets, i.e., itemsets with support greater or equal to σ . The second step is the extraction of association rules from frequent itemsets. Let X be a frequent itemset, hence having a support higher than σ , pairs X and Y are derived so that the confidence of $X \rightarrow Y$ is higher than τ . The first step of the process is the most computationally expensive. Thus, several algorithms have been proposed to solve the problem of frequent itemset extraction (Zaki, 2000), some specifically addressing high-dimensionality issues (Apiletti et al., 2015, Apiletti et al., 2017). Despite being computationally demanding, association rule mining is an exhaustive approach, i.e., all and only statistically relevant correlations are extracted. Section 3.8.11 offers an example of applying association rule mining to forecast the quality of beverages.

2.9.3. Forecasting with text information⁷⁰

Text data, such as social media posts, scholar articles and company reports, often contains valuable information that can be used as predictors in forecasting models (Aggarwal & Zhai, 2012). Before extracting useful features from the data, a text document needs to be cleaned and normalised for further processing. The first step of preparing the data is to filter out *stop words* – the words that do not add much meaning to a sentence, e.g., “a”, “is”, and “me”. For grammatical reasons, it is necessary for documents to use different forms of words. Stemming and lemmatisation can be applied to reduce inflectional forms or relate different forms of a word to a common base form. Stemming often chops off the end of a word, while lemmatisation uses vocabularies and morphological analysis to return the base form (or the *lemma*) of a word (Lovins, 1968, Manning et al., 2008). For example, the word “industries” will be turned into “industri” or “industry” if stemming or lemmatisation is applied.

To model and analyse text data, we need to transform it into numerical representations so the forecasting models can process them as predictors. One way of transforming the data is through sentiment analysis. Sentiment analysis is often applied to detect polarity within customer materials such as reviews and social media posts (Archak et al., 2011, Das and Chen, 2007). An easy way to obtain the sentiment score of a word is to look it up in a happiness dictionary (for example, the hedonometer dictionary, Hedonometer, 2020). Another common way of representing the sentiment is to use a vector of numeric values that denote the word’s positivity, negativity and neutrality based on existing lexical databases such as the *WordNet* (Baccianella et al.,

2010, Godbole et al., 2007). Once the sentiment of each word is calculated, we can apply an aggregation algorithm (e.g., simple average) to measure the sentiment of an entire sentence or paragraph.

In scholar articles and company reports, context features might be more important than sentiments. The bag-of-words model and word embeddings are often applied to generate numeric representations of such text. A bag-of-words model simply returns a matrix that describes the occurrence of words within a document (Goldberg, 2017).

When we use this matrix as input to a forecasting model, each word count can be considered as a feature. The Word2Vec method is a widely used embedding method that is built based on the context of a word. Specifically, it trains a two-layer neural network that takes each word as an input to predict its surrounding words (see Section 2.7.8 for a discussion of neural networks for forecasting). The weights from the input layer to the hidden layer are then utilised as the numerical representation for the input word (Le & Mikolov, 2014). Once the text is turned into numeric representations, they can be used as predictors in any forecasting models. The most challenging part in this process is to find the right technique to extract features from the text.

In terms of software implementation, the Natural Language Toolkit (NLTK) and SpaCy library in Python can be applied to remove stop words and stem or lemmatise text (Honnibal, 2015, Loper and Bird, 2002). The bag-of-words technique is also available in NLTK. A particular implementation of the Word2Vec model is available on Google code (2013). Moreover, a public data set of movie reviews that is commonly studied in literature is available from the Stanford NLP Group (2013).

2.10. Forecasting by aggregation

2.10.1. Cross-sectional hierarchical forecasting⁷¹

In many applications time series can be aggregated at several levels of aggregation based on geographic or logical reasons to form hierarchical structures. These are called hierarchical time series. In the retail industry (see also Section 3.2.4), for example, individual sales of products at the bottom-level of the hierarchy can be grouped in categories and families of related products at increasing aggregation levels, with the total sales of the shop or distribution centre at the top level (Oliveira and Ramos, 2019, Pennings and van Dalen, 2017, Villegas and Pedregal, 2018). Similarly, cross-

sectional hierarchies can be used for spatial aggregation to help model housing prices or traffic in transportation networks, or otherwise formed geographical demarcations (for example, [Athanasopoulos et al., 2009](#), [Kourentzes and Athanasopoulos, 2019](#)). The forecasts of hierarchical time series produced independently of the hierarchical structure generally will not add up according to the aggregation constraints of the hierarchy, i.e., they are not coherent. Therefore, hierarchical forecasting methods that generate coherent forecasts should be considered to allow appropriate decision-making at the different levels. Actually, by taking advantage of the relationships between the series across all levels these methods have shown to improve forecast accuracy ([Athanasopoulos et al., 2009](#), [Shang and Hyndman, 2017](#), [Yagli et al., 2019](#)). One of the main reasons behind this improved performance is that forecast reconciliation is effectively a special case of forecast combinations ([Hollyman, Petropoulos, & Tipping, 2021](#)); see also Section 2.6.

The most common approaches to hierarchical forecasting are bottom-up and top-down. In the bottom-up approach forecasts for each time series at the bottom-level are first produced and then these are added up to obtain forecasts for all other series at the hierarchy ([Dunn, Williams, & Dechaine, 1976](#)). Since forecasts are obtained at the bottom-level no information is lost due to aggregation. In the top-down approach forecasts for the top-level series are first generated and then these are disaggregated generally using historical proportions to obtain forecasts for the bottom-level series, which are then aggregated ([Gross & Sohl, 1990](#)). [Hyndman, Ahmed, Athanasopoulos, and Shang \(2011\)](#) claim that this approach introduces bias to the forecasts, however [Hollyman et al. \(2021\)](#) showed that it is possible to calculate unbiased top-down forecasts.

Recent research on hierarchical forecasting tackles the problem using a two-stage approach. Forecasts for the series at all levels of the hierarchy are first obtained independently without considering any aggregation constraints (we refer to these as base forecasts). Then, base forecasts are adjusted so that they become coherent (we refer to these as reconciled forecasts). This adjustment is achieved by a matrix that maps the base forecasts into new bottom-level forecasts which are then added up ([Hyndman et al., 2011](#)).

[Wickramasuriya, Athanasopoulos, and Hyndman \(2019\)](#) found the optimal solution for this matrix, which minimises the trace of the covariance matrix of the reconciled

forecast errors (hence MinT reconciliation). This optimal solution is based on the covariance matrix of the base forecast errors which incorporates the correlation structure of the hierarchy. [Wickramasuriya et al. \(2019\)](#) presented several alternative estimators for this covariance matrix: (i) proportional to the identity which is optimal only when base forecast errors are uncorrelated and equivariant (referred to as OLS), (ii) proportional to the sample covariance estimator of the in-sample one-step-ahead base forecast errors with off-diagonal elements null accounts for the differences in scale between the levels of the hierarchy (referred to as WLS), (iii) proportional to the previous estimator unrestricted also accounts for the relationships between the series (referred to as MinT-Sample), and (iv) proportional to a shrinkage estimator based on the two previous estimators, parameterising the shrinkage in terms of variances and correlations, accounts for the correlation across levels (referred as MinT-Shrink). Other researchers focus on simple (equal-weighted) combinations of the forecasts produced at different hierarchical levels ([Abouarghoub et al., 2018](#), [Hollyman et al., 2021](#)). [Pritularga, Svetunkov, and Kourentzes \(2021\)](#) showed that more complex reconciliation schemes result in more variability in the forecasts, due to the estimation of the elements in the covariance matrix, or the implicit combination weights. They provide approximations for the covariance matrix that balance this estimation uncertainty with the benefits of more finely tuned weights.

More recently these techniques were extended to probabilistic forecasting ([Ben Taieb, Taylor, & Hyndman, 2020](#)). When base forecasts are probabilistic forecasts characterised by elliptical distributions, [Panagiotelis, Gamakumara, Athanasopoulos, and Hyndman \(2021\)](#) showed that reconciled probabilistic forecasts also elliptical can be obtained analytically. When it is not reasonable to assume elliptical distributions, a non-parametric approach based on bootstrapping in-sample errors can be used.

2.10.2. Temporal aggregation⁷²

Temporal aggregation is the transformation of a time series from one frequency to another of lower frequency. As an example, a time series of length n that is originally sampled at a monthly frequency can be transformed to a quarterly series of length $n/4$ by using equally-sized time buckets of three periods each. It is usually applied in a non-overlapping manner, but overlapping aggregation can also be considered. The latter is preferred in the case when the original series is short, but has the disadvantage of

applying lower weights on the few first and last observations of the series and introducing autocorrelations (Boylan & Babai, 2016 and Section 2.8.2).

Temporal aggregation is appealing as it allows to investigate the original series through different lenses. By changing the original frequency of the data, the apparent series characteristics also change. In the case of slow-moving series, temporal aggregation leads to decrease of intermittence (Nikolopoulos, Syntetos, Boylan, Petropoulos, & Assimakopoulos, 2011 see also Section 2.8). In the case of fast-moving series, higher levels of aggregation (i.e., lower frequencies) allow for better modelling of trend patterns, while lower aggregation levels (i.e., higher frequencies) are more suitable for capturing seasonal patterns (Kourentzes et al., 2014, Spithourakis et al., 2014).

Research has found evidence of improved forecasting performance with temporal aggregation for both slow (Nikolopoulos et al., 2011) and fast (Spithourakis, Petropoulos, Babai, Nikolopoulos, & Assimakopoulos, 2011) moving time series. This led to characterise temporal aggregation as a “self-improving mechanism”. The good performance of temporal aggregation was reconfirmed by Babai, Ali, and Nikolopoulos (2012), who focused on its utility performance rather than the forecast error. However, one challenge with single levels of aggregation is the choice of a suitable aggregation level for each series (Kourentzes, Rostami-Tabar, & Barrow, 2017).

Instead of focusing on a single aggregation level, Andrawis et al. (2011), Kourentzes et al. (2014), Petropoulos and Kourentzes (2014), and Petropoulos and Kourentzes (2015) suggested the use of multiple levels of aggregation, usually abbreviated as MTA (multiple temporal aggregation). This not only tackles the need to select a single aggregation level, but also partly addresses the issue of model uncertainty, instead of relying on model selection and parametrisation at a single aggregation level. Using this property, Kourentzes et al. (2017) showed that MTA will typically lead to more accurate forecasts, even if in theory suboptimal. Different frequencies allow for better identification of different series patterns, so it is intuitive to consider multiple temporal levels and benefit from the subsequent forecast combination across frequencies. Kourentzes and Petropoulos (2016) showed how multiple temporal aggregation can be extended to incorporate exogenous variables (see also Section 2.2.5). However, forecasting at a single level of aggregation can still result in better performance when the seasonal pattern is strong (Spiliotis, Petropoulos et al., 2019, Spiliotis, Petropoulos et al., 2020).

[Athanasopoulos et al. \(2017\)](#) expressed multiple temporal aggregation within the hierarchical forecasting framework (see Section [2.10.1](#)) using the term “temporal hierarchies”. Temporal hierarchies allow for the application of established hierarchical reconciliation approaches directly to the temporal dimension. [Jeon, Panagiotelis, and Petropoulos \(2019\)](#) show how temporal hierarchies can be used to obtain reconciled probabilistic forecasts, while ([Spiliotis, Petropoulos et al., 2019](#)) explored empirical bias-adjustment strategies and a strategy to avoid excessive seasonal shrinkage. [Nystrup, Lindström, Pinson, and Madsen \(2020\)](#) proposed estimators for temporal hierarchies suitable to account for autocorrelation in the data. Finally, [Kourentzes and Athanasopoulos \(2020\)](#) applied temporal hierarchies on intermittent data, and showed that higher aggregation levels may offer structural information which can improve the quality of the forecasts.

2.10.3. Cross-temporal hierarchies⁷³

In the last two subsections (Sections [2.10.1 Cross-sectional hierarchical forecasting](#), [2.10.2 Temporal aggregation](#)), we saw two complimentary hierarchical structures, cross-sectional and temporal. Although the machinery behind both approaches is similar, often relying on the hierarchical framework by [Athanasopoulos et al. \(2009\)](#) and [Hyndman et al. \(2011\)](#), and work that followed from these (particularly [Wickramasuriya et al., 2019](#)), they address different forecasting problems. Cross-sectional hierarchies change the unit of analysis but are fixed in the period of analysis. For example, a manufacturer may operate using a hierarchy across products. The different nodes in the hierarchy will correspond to different products, product groups, super-groups, and so on, but will all refer to the same period, for example, a specific week. Temporal hierarchies do the opposite, where the unit of analysis is fixed, but the period is not. For example, we may look at the sales of a specific Stock Keeping Unit (SKU) at a daily, weekly, monthly, quarterly and annual levels. However, one can argue that having annual forecasts at the SKU level may not be useful. Similarly, having aggregate sales across an organisation at a weekly frequency is also of little value.

In connecting these to organisational decisions, we can observe that there is only a minority of problems that either cross-sectional or temporal hierarchies are natural, as typically decisions can differ across both the unit and the period (planning horizon) of analysis. In the latter case, both hierarchical approaches are more akin to statistical

devices that can improve forecast accuracy through the use of forecast combinations, rather than satisfy the motivating argument behind hierarchical forecasting that is to provide coherent predictions for decisions at different levels of the hierarchy. Cross-temporal hierarchies attempt to overcome this limitation, providing coherent forecasts across all units and periods of analysis, and therefore a common outlook for the future across decision-makers at different functions and levels within an organisation. The literature remains sparse on how to construct cross-temporal forecasts, as the size of the hierarchy can easily become problematic. [Kourentzes and Athanasopoulos \(2019\)](#) propose a heuristic approach to overcome the ensuing estimation issues. The approach works by compartmentalising the estimation. First, they obtain estimates of the cross-sectional reconciliation weights for each temporal level of the hierarchy. Then, these are combined across temporal levels, to a unique set that satisfies all coherency constraints. Using these combined weights, they obtain the reconciled bottom level forecasts, which can be aggregated as needed. Although they recognise that their approach can result in suboptimal results in terms of reconciliation errors, it guarantees coherent forecasts. Cross-temporal forecasts are more accurate than either temporal or cross-sectional hierarchical forecasts and provide a holistic view of the future across all planning levels and demarcations. [Spiliotis, Petropoulos et al. \(2020\)](#) also identify the problem, however, they do not focus on the coherency of forecasts and propose a sequential reconciliation across the two dimensions. This is shown to again be beneficial, but it does not achieve coherency. Arguably one can adapt the iterative correction algorithm by [Kourentzes and Athanasopoulos \(2020\)](#) to enforce coherency in this approach as well.

2.10.4. Ecological inference forecasting⁷⁴

Ecological inference forecasting (EIF) aims to predict the inner-cells values of a set of contingency tables when only the margins are known. It defines a fundamental problem in disciplines such as political science, sociology and epidemiology ([Salway & Wakefield, 2004](#)). [Cleave, Brown, and Payne \(1995\)](#), [Greiner \(2007\)](#) and [Pavia, Cabrer, and Sala \(2009\)](#) describe other areas of application. The fundamental difficulty of EIF lies in the fact that this is a problem with more unknowns than observations, giving rise to concerns over identifiability and indeterminacy: many sets of substantively different internal cell counts are consistent with a given marginal table. To overcome this issue, a similarity hypothesis (and, sometimes, the use of covariates) is routinely assumed. The

basic hypothesis considers that either conditional row (underlying) probabilities or fractions are similar (related) among contingency tables (Greiner & Quinn, 2010). The covariations among row and column margins of the different tables are then used to learn about the internal cells.

The above hypothesis is not a cure-all to the main drawback of this approach. EIF is exposed to the so-called ecological fallacy (Robinson, 1950): the presence of inconsistencies in correlations and association measures across different levels of aggregation. This is closely related to the well-known Simpson's Paradox (Simpson, 1951). In this setting, the ecological fallacy is manifested through aggregation bias (Wakefield, 2004) due to contextual effects and/or spatial autocorrelation (Achen & Phillips Shively, 1995). This has led many authors to disqualify ecological inference forecasts (see, for example, Anselin and Tam Cho, 2002, Freedman et al., 1998, Herron and Shotts, 2004, Tam Cho, 1998) and many others to study under which circumstances ecological inference predictions would be reliable (Firebaugh, 1978, Forcina and Pellegrino, 2019, Gelman et al., 2001, Guseo, 2010). Despite the criticisms, many algorithms for solving the EIF problem can be found in the literature, mainly from the ecological regression and mathematical programming frameworks (some of them available in functions of the R statistical software).

The ecological regression literature has been prolific since the seminal papers of Duncan and Davis (1953) and Goodman, 1953, Goodman, 1959 and is undergoing a renaissance after King (1997): new methods generalised from 2×2 tables to tables have been proposed (King et al., 1999, Rosen et al., 2001), the geographical dimension of the data is being explicitly considered (Calvo and Escolar, 2003, Puig and Ginebra, 2015), and new procedures combining aggregated and individual level data, including exit polls (see also Section 3.8.5), are introduced (Glynn and Wakefield, 2010, Greiner and Quinn, 2010, Klima et al., 2019). See King, Tanner, and Rosen (2004) for a wide survey and Klima, Thurner, Molnar, Schlesinger, and Küchenhoff (2016) and Plescia and De Sio (2018) for an extensive evaluation of procedures. In mathematical programming exact and inequality constraints for the inner-cell values are incorporated in a natural way. Hence, this approach has shown itself to be a proper framework for generating ecological inference forecasts. The proposals from this approach can be traced back to Hawkes (1969) and Irwin and Meeter (1969). After them, some key references include Corominas et al., 2015, McCarthy and Ryan, 1977, Pavía and Romero,

2021, Romero et al., 2020, and Tzifafetas (1986). Solutions based on other strategies, for instance, entropy maximization, have been also suggested (see, for example, Bernardini Papalia and Fernandez Vazquez, 2020, Johnston and Pattie, 2000).

2.11. Forecasting with judgment

2.11.1. Judgmental forecasting⁷⁵

People may use judgment alone to make forecasts or they may use it in combination with statistical methods. Here the focus is on pure judgmental forecasting (for judgmental adjustments, see Section 2.11.2). Different types of judgment heuristic (mental ‘rules of thumb’) can be used to make forecasts. The heuristic used depends on the nature of the information available to the forecaster (Harvey, 2007).

Consider cases where the only relevant information is held in the forecaster’s memory. For example, someone might be asked whether Manchester United or Burnley will win next week’s match. Here one memory-based heuristic that might be applicable is the recognition heuristic: if one recognises one object but not the other, then one should infer that the recognised object has higher value (Goldstein & Gigerenzer, 2002). In the above example, most people who recognise just one of the teams would be likely to make a correct forecast that Manchester United will win (Ayton, Önkal, & McReynolds, 2011). The availability heuristic is another memory-based heuristic that may be applicable: objects that are most easily brought to mind are those which are more likely. Thus, if we are asked which team is likely to come top of the premier league, we would say Manchester United if that is the one that most easily comes to mind. The availability heuristic is often effective because more likely events are encountered more often and more recently and are hence better remembered. However, it can be disrupted by, for example, greater media coverage of more unlikely (and hence more interesting) events. Consider next cases in which forecasters possess information about values of one or more variables correlated with the variable to be forecast. For example, teachers may wish to forecast the grades of their students in a final examination on the basis of past records of various other measures. Kahneman and Tversky (1973) suggested that people use the representativeness heuristic to deal with this type of situation. Forecasters first select a variable that they think is able to represent the one that must be predicted. For

example, a teacher may consider that frequency in attending voluntary revision classes represents a student's ability in the final examination. Thus, if a student attended 15 of the 20 revision classes, they are likely to obtain 75% in the final examination.

Finally, consider situations in which people forecast future values of a variable on the basis of a record of previous values of that variable. There is some evidence that, when forecasting from time series, people use anchor-and-adjustment heuristics (Hogarth and Makridakis, 1981, Lawrence and O'Connor, 1992). For example, (i) when forecasting from an upward trended series, they anchor on the last data point and then make an upward adjustment to take the trend into account and (ii) when forecasting from an untrended series containing autocorrelation, they anchor on the last data point and make an adjustment towards the mean to take the autocorrelation into account.

Kahneman (2011) and others have divided cognitive processes into those which are intuitive (System 1) and those which are deliberative (System 2). We have discussed only intuitive processes underlying judgmental forecasting (Gigerenzer, 2007). However, they can be supplemented by deliberative (System 2) processes (Theocharis & Harvey, 2019) in some circumstances.

2.11.2. Judgmental adjustments of computer-based forecasts⁷⁶

Judgmental adjustments to algorithmic computer-based forecasts can enhance accuracy by incorporating important extra information into forecasts (Fahimnia et al., 2020, McNees, 1990, Perera et al., 2019). However, cognitive factors (see, for example, Section 2.11.1), and motivational biases (see Section 3.2.2), can lead to the inefficient use of information (Fildes, Goodwin, & Önköl, 2019a), unwarranted adjustments and reductions in accuracy (Fildes et al., 2009, Franses and Legerstee, 2009a).

People may 'egocentrically discount' a computer's forecasts when its rationale is less clear than their own reasoning (Bonaccio & Dalal, 2006). They can also be less tolerant of errors made by algorithms than those made by humans (Dietvorst et al., 2015, Önköl et al., 2009 and Section 3.7.4). The random errors associated with algorithmic forecasts, and the salience of rare large errors, can therefore lead to an unjustified loss of trust in computer forecasts. Adjustments may also give forecasters a sense of ownership of forecasts or be used to justify their role (Önköl & Gönül, 2005).

Computer-based forecasts are designed to filter randomness from time-series. In contrast, humans tend to perceive non-existent systematic patterns in random movements (O'Connor et al., 1993, Reimers and Harvey, 2011 and Section 3.7.3) and

apply adjustments to reflect them. This can be exacerbated by the narrative fallacy (Taleb, 2008), where people invent stories to explain these random movements, and hindsight bias (Fischhoff, 2007), where they believe, in retrospect, that these movements were predictable. Recent random movements, and events, are particularly likely to attract undue attention, so long-run patterns identified by the computer are given insufficient weight (Bolger & Harvey, 1993). Damaging interventions are also probable when they result from political interference (Oliva & Watson, 2009) or optimism bias (Fildes et al., 2009), or when they reflect information already factored into the computer's forecast, leading to double counting (Van den Broeke, De Baets, Vereecke, Baecke, & Vanderheyden, 2019).

How can interventions be limited to occasions when they are likely to improve accuracy? Requiring people to document reasons justifying adjustments can reduce gratuitous interventions (Goodwin, 2000b). Explaining the rationale underlying statistical forecasts also improved adjustment behaviour when series had a simple underlying pattern in a study by Goodwin and Fildes (1999). However, providing guidance on when to adjust was ineffective in an experiment conducted by Goodwin, Fildes, Lawrence, and Stephens (2011), as was a restriction preventing people from making small adjustments. When determining the size of adjustments required, decomposing the judgment into a set of easier tasks improved accuracy in a study by Webby, O'Connor, and Edmundson (2005). Providing a database of past outcomes that occurred in circumstances analogous to those expected in the forecast period also improved adjustments in a study by Lee, Goodwin, Fildes, Nikolopoulos, and Lawrence (2007). Outcome feedback, where the forecaster is informed of the most recent outcome is unlikely to be useful since it contains noise and exacerbates the tendency to over-focus on recent events (Goodwin and Fildes, 1999, Petropoulos et al., 2016). However, feedback on biases in adjustments over several recent periods may improve judgments (Petropoulos, Goodwin, & Fildes, 2017). Feedback will be less useful where interventions are rare so there is insufficient data to assess performance.

The evidence for this section is largely based on laboratory-based studies of adjustment behaviour. Section 3.7.3 gives details of research into forecast adjustment in practice and discusses the role of forecasting support systems in improving the effectiveness of judgmental adjustments.

2.11.3. Judgmental model selection⁷⁷

Forecasters – practitioners and researchers alike – use Forecast Support Systems (FSS) in order to perform their forecasting tasks. Usually, such an FSS allows the forecaster to load their historical data and they can then apply many different types of forecasting techniques to the selected data. The idea is that the forecaster selects the method which leads to highest forecast accuracy. Yet, there is no universally ‘best’ method, as it depends on the data that is being forecasted (see also Section 2.5.4). Thus, selection is important in achieving high accuracy. But how does this selection occur?

Research on judgmental selection of statistical forecasts is limited in quantity. [Lawrence, Goodwin, and Fildes \(2002\)](#) found that participants were not very adept at selecting good forecasting algorithms from a range offered to them by an FSS and had higher error than those who were presented with the optimal algorithm by an FSS. [Petropoulos et al. \(2018b\)](#) compared judgmental selection of forecasting algorithms with automatic selection based on predetermined information criteria. They found that judgmental selection was better than automatic selection at avoiding the ‘worst’ models, but that automatic selection was superior at choosing the ‘best’ ones. In the end, overall accuracy of judgmental selection was better than that of algorithmic selection. If their experiment had included more variation of the data (trends, fractals, different autoregressive factors) and variation of proposed models, this could possibly have led to better algorithmic than judgmental performance ([Harvey, 2019](#)). Time series that are more complex will place a higher cognitive load on judgmental selection. This was confirmed in a study by [Han, Wang, Petropoulos, and Wang \(2019\)](#), who used an electroencephalogram (EEG) for the comparison of judgmental forecast selection versus (judgmental) pattern identification. They found that pattern identification outperformed forecast selection, as the latter required a higher cognitive load, which in turn led to a lower forecasting accuracy.

It is likely that, in practice, judgmental selection is much more common than automatic selection. This preference for human judgment over advice from an algorithm has been shown in an experiment by [Önköl et al. \(2009\)](#). But how apt are forecasters in distinguishing ‘good’ models from ‘bad’ models? This was investigated by [De Baets and Harvey \(2020\)](#) in an experiment. People were asked to select the best performing model out of a choice of two different qualities (accuracies) of models (different combinations of good versus medium versus bad). People’s choice outperformed forecasts made by averaging the model outputs, lending credence to the views of [Fifić and Gigerenzer](#)

(2014). The performance of the participants improved with a larger difference in quality between models and a lower level of noise in the data series. In a second experiment, [De Baets and Harvey \(2020\)](#) found that participants adjusted more towards the advice of what they perceived to be a good quality model than a medium or bad quality one. Importantly, in selecting an algorithm and seeing it err, people are quick to abandon it. This phenomenon is known as ‘algorithm aversion’ ([Dietvorst et al., 2015](#) see also Section 2.11.6) and is due to a ‘perfection schema’ we have in our heads where algorithms are concerned ([Madhavan & Wiegmann, 2007](#)). We do not expect them to ‘fail’ and thus react strongly when they do. While a model may not perform as it should for a particular dataset and may thus elicit algorithm aversion for that particular method, one should not abandon it for all datasets and future forecasts.

2.11.4. Panels of experts⁷⁸

Panels of experts are often used in practice to produce judgmental forecasts (see, for example, Sections 3.2.6 [New product forecasting](#), 3.8.5 [Elections forecasting](#)). This is especially true in cases with limited available quantitative data and with the level of uncertainty being very high. In this section, three methods for eliciting judgmental forecasts from panels of experts are presented: the [Delphi method](#), interaction groups (IG), and structured analogies (SA).

The [Delphi method](#) is centred around organising and structuring group communication ([Rao, Anderson, Sukumar, Beauchesne, Stein, & Frankel, 2010](#)), which aims to achieve a convergence of opinion on a specific real-world issue. It is a multiple-round survey in which experts participate anonymously to provide their forecasts and feedback ([Rowe & Wright, 2001](#)). At the end of each round, the facilitator collects and prepares statistical summaries of the panel of experts’ forecasts. These summaries are presented as feedback to the group, and may be used towards revising their forecasts. This loop continues until a consensus is reached, or the experts in the panel are not willing to revise their forecasts further. In some implementations of the Delphi method, justification of extreme positions (forecasts) is also part of the (anonymous) feedback process. The Delphi method results in a more accurate outcome in the decision-making process ([Dalkey, 1969](#), [Steurer, 2011](#)). [Rowe and Wright \(2001\)](#) mentioned that, by adopting the Delphi method, groups of individuals can produce more accurate forecasts than simply using unstructured methods. A drawback of the Delphi method is the additional cost associated with the need to run multiple rounds, extending the

forecasting process as well as increasing the potential drop-out rates. On the other hand, the anonymity in the Delphi method eliminates issues such as groupthink and the ‘dominant personalities’ effects (Van de Ven & Delbeco, 1971).

The IG method suggests that the members of the panel of experts actively interact and debate their points to the extent they have to reach an agreement on a common forecast (Litsiou, Polychronakis, Karami, & Nikolopoulos, 2019). Sniezek and Henry (1989) found that members of interacting groups provide more accurate judgments compared to individuals. However, there is mixed evidence about the forecasting potential of IG (Boje and Murnighan, 1982, Graefe and Armstrong, 2011, Scott Armstrong, 2006). Besides, the need for arranging and facilitating meetings for the IG makes it a less attractive option.

Another popular approach to judgmental forecasting using panels of experts is SA, which refers to the recollection of past experiences and the use analogies (Green & Armstrong, 2007). In the SA method, the facilitator assembles a panel of experts. The experts are asked to recall and provide descriptions, forecasts, and similarities/differences for cases analogous to the target situation, as well as a similarity ranking for each of these analogous cases. The facilitator gathers the lists of the analogies provided by the experts, and prepares summaries, usually using weighted averages of the recalled cases based on their similarity to the target situation (see also Section 2.6.4). Semi-structured analogies (sSA) have also been proposed in the literature, where the experts are asked to provide a final forecasts based on the analogous cases they recalled, which essentially reduces the load for the facilitator (Nikolopoulos, Litsa, Petropoulos, Bougioukos, & Khammash, 2015). Nikolopoulos et al. (2015) supported that the use of SA and IG could result to forecasts that are 50% more accurate compared to unstructured methods (such as unaided judgment). One common caveat of using panels of experts is the difficulty to identify who a real expert is. Engaging experts with high level of experience, and encouraging the interaction of experts are also supported by Armstrong and Green (2018).

2.11.5. Scenarios and judgmental forecasting⁷⁹

Scenarios provide exhilarating narratives about conceivable futures that are likely to occur. Through such depictions they broaden the perspectives of decision makers and act as mental stimulants to think about alternatives. Scenarios enhance information

sharing and provide capable tools for communication within organisations. By virtue of these merits, they have been widely used in corporate planning and strategy setting since 1960's (Godet, 1982, Goodwin and Wright, 2010, Schoemaker, 1991, Wright and Goodwin, 1999, Wright and Goodwin, 2009). Even though utilisation of scenarios as decision advice to judgmental forecasting has been proposed earlier (Bunn and Salo, 1993, Schnaars and Topol, 1987), the research within this domain remained limited until recently when the interest in the subject has rekindled (Goodwin et al., 2019b, Önkal et al., 2013, Wicke et al., 2019).

The recent research has used behavioural experimentation to examine various features of scenarios and their interactions with judgmental adjustments (see Section 2.11.2) of model-based forecasts. Önkal et al. (2013) explored the 'content' effects of scenarios where through the narration either a bleak/negative future (a pessimistic scenario) or a bright/positive future (an optimistic scenario) was portrayed. On a demand forecasting context for mobile phones, the participants first received time-series data, model-based forecasts and then asked to generate point and interval forecasts as well as provide a confidence measure. With respect to the existence of scenarios, there were four conditions where the participants may receive: (i) no scenarios, (ii) optimistic scenarios, (iii) pessimistic scenarios, and (iv) both scenarios. Findings indicate that decision makers respond differently to optimistic and pessimistic scenarios. Receiving optimistic scenarios resulted in making larger adjustments to the model-based forecasts. At the same time, led to an increased confidence of the participants in their predictions. On the other hand, participants who received negative scenarios tend to lower their predictions the most among the four groups. An intriguing finding was the balancing effect of scenarios on the interval forecast symmetry. The lower interval bounds were adjusted upwards the most towards the centre-point of the interval (i.e., model-based predictions) when optimistic scenarios were received. Similarly, the upper bounds were adjusted downwards the most towards the centre-point of the interval in the presence of pessimistic scenarios.

The prospects of receiving a single scenario versus multiple scenarios were further explored in Goodwin et al. (2019b). The researchers investigated whether assimilation or contrast effects will occur when decision makers see optimistic (pessimistic) forecasts followed by pessimistic (optimistic) ones compared against receiving a single scenario in solitude. In case of assimilation, a scenario presenting an opposing world view with the

initial one would cause adjustments in the opposite direction creating an offset effect. On the other hand, in case of contrast, the forecasts generated after the initial scenarios would be adjusted to more extremes when an opposing scenario is seen. In two experiments conducted in different contexts the researchers found resilient evidence for contrast effects taking place. Interestingly, seeing an opposing scenario also increased the confidence of the forecasters in their initial predictions.

In terms of the effects of scenario presence on the forecasting performance, however, the experimental evidence indicates the benefits are only circumstantial. [Goodwin, Gönül, and Önköl \(2019a\)](#) found that providing scenarios worsened forecast accuracy and shifted the resultant production order decisions further away from optimality. Despite this performance controversy, the decision makers express their fondness in receiving scenarios and belief in their merits ([Goodwin et al., 2019b](#), [Önköl et al., 2013](#)). Therefore, we need more tailored research on scenarios and judgmental forecasting to reveal the conditions when scenarios can provide significant improvements to the forecasting accuracy.

2.11.6. Trusting model and expert forecasts⁸⁰

Defined as “firm belief in the reliability, truth, and ability of someone/something” (Oxford English Dictionary), trust entails accepting vulnerability and risk ([Rousseau, Sitkin, Burt, & Camerer, 1998](#)). Given that forecasts are altered or even discarded when distrusted by users, examining trust is a central theme for both forecasting theory and practice.

Studies examining individual’s trust in model versus expert forecasts show that individuals often distrust algorithms ([Burton et al., 2020](#), [Meehl, 2013](#)) and place higher trust on human advice ([Diab et al., 2011](#), [Eastwood et al., 2012](#) but also Sections [2.11.2 Judgmental adjustments of computer-based forecasts](#), [2.11.3 Judgmental model selection](#), [3.7.4 Trust in forecasts](#)). We live in an era where we are bombarded with news about how algorithms get it wrong, ranging from COVID-19 forecasts affecting lockdown decisions to algorithmic grade predictions affecting university admissions. Individuals appear to prefer forecasts from humans over those from statistical algorithms even when those forecasts are identical ([Önköl et al., 2009](#)). Furthermore, they lose trust in algorithms quicker when they see forecast errors ([Dietvorst et al., 2015](#), [Prah and Van Swol, 2017](#)). Such ‘algorithm aversion’ and error intolerance is reduced when users have opportunity to adjust the forecasting

outcome, irrespective of the extent of modification allowed (Dietvorst, Simmons, & Massey, 2018). Feedback appears to improve trust, with individuals placing higher trust in algorithms if they can understand them (Seong & Bisantz, 2008). Overuse of technical language may reduce understanding of the forecast/advice, in turn affecting perceptions of expertise and trustworthiness (Joiner, Leveson, & Langfield-Smith, 2002). Explanations can be helpful (Goodwin, Gonul & Onkal, 2013), with their packaging affecting judgments of trustworthiness (Elsbach & Eloffson, 2000). Algorithmic appreciation appears to easily fade with forecasting expertise (Logg, Minson, & Moore, 2019), emphasising the importance of debiasing against overconfidence and anchoring on one's own predictions.

Trusting experts also presents challenges (Hendriks et al., 2015, Hertzum, 2014, Maister et al., 2012). Expert forecasts are typically seen as predisposed to group-based preconceptions (Brennan, 2020, Vermue et al., 2018), along with contextual and motivational biases (Burgman, 2016). Misinformed expectations, distorted exposures to 'forecast failures', and over-reliance on one's own judgments may all contribute to distrusting experts as well as algorithms.

Credibility of forecast source is an important determinant in gaining trust (Önkal, Gönül, & De Baets, 2019). Studies show that the perceived credibility of system forecasts affects expert forecasters' behaviours and trust (Alvarado-Valencia & Barrero, 2014), while providing information on limitations of such algorithmic forecasts may reduce biases (Alvarado-Valencia, Barrero, Önkal, & Dennerlein, 2017). Previous experience with the source appears to be key to assessing credibility (Hertzum, 2002) and trust (Cross & Sproull, 2004). Such 'experienced' credibility appears to be more influential on users' acceptance of given forecasts as opposed to 'presumed' credibility (Önkal, Gönül, Goodwin, Thomson, & Öz, 2017). Source credibility can be revised when forecast (in)accuracy is encountered repetitively (Jiang, Muhanna, & Pick, 1996), with forecaster and user confidence playing key roles (Sah, Moore, & MacCoun, 2013).

Trust is critical for forecasting efforts to be translated into sound decisions (Choi et al., 2020, Özer et al., 2011). Further work on fostering trust in individual/collaborative forecasting will benefit from how trusted experts and models are selected and combined to enhance decision-making.

2.12. Evaluation, validation, and calibration

2.12.1. Benchmarking⁸¹

When a new forecasting model or methodology is proposed, it is common for its performance to be benchmarked according to some measure of forecast accuracy against other forecasting methods using a sub-sample of some particular time series. In this process, there is the risk that either the measures of accuracy, competing forecasting methods or test data, are chosen in a way that exaggerates the benefits of a new method. This possibility is only exacerbated by the phenomenon of publication bias ([Dickersin, 1990](#)).

A rigorous approach to benchmarking new forecasting methods should follow the following principles:

- 1.
New methods should always be compared to a larger number of suitable benchmark methods. These should at a minimum include naïve methods such as a random walk and also popular general purpose forecasting algorithms such as ARIMA models, Exponential Smoothing, Holt Winters and the Theta method (see Section [2.3](#) and references therein).
- 2.
Forecasts should be evaluated using a diverse set of error metrics for point, interval and probabilistic forecasts (see Section [2.12.2](#)). Where the forecasting problem at hand should be tailored to a specific problem, then appropriate measures of forecast accuracy must be used. As an example, the literature on Value at Risk forecasting has developed a number of backtesting measures for evaluating the quality of quantile forecasts (see [Zhang & Nadarajah, 2018](#) and references therein).
- 3.
Testing should be carried out to discern whether differences between forecasting methods are statistically significant. For discussion see Section [2.12.6](#). However, there should also be a greater awareness of the debate around the use of hypothesis testing both in forecasting ([Armstrong, 2007](#)) and more generally in statistics ([Wasserstein & Lazar, 2016](#)).
- 4.

Sample sizes for rolling windows should be chosen with reference to the latest literature on rolling window choice (see [Inoue et al., 2017](#) and references therein).

- 5.
All code used to implement and benchmark new forecasting methods should, where possible, be written in open source programming languages (such as C, Python and R). This is to ensure replicability of results (for more on the replicability crisis in research see [Peng, 2015](#) and references therein)
- 6.
Methods should be applied to appropriate benchmark datasets.

Regarding the last of these points there are some examples in specific fields, of datasets that already play a de facto role as benchmarks. In macroeconomic forecasting, the U.S. dataset of [Stock and Watson \(2012, see §2.7.1\)](#) is often used to evaluate forecasting methods that exploit a large number of predictors, with [Forni, Hallin, Lippi, and Reichlin \(2003\)](#) and [Panagiotelis, Athanasopoulos, Hyndman, Jiang, and Vahid \(2019\)](#) having constructed similar datasets for the EU and Australia respectively. In the field of energy, the GEFCom data ([Hong, Pinson, Fan, Zareipour, Troccoli, & Hyndman, 2016](#)) discussed in Section 3.4.3 and the IEEE 118 Bus Test Case data ([Peña, Martínez-Anido, & Hodge, 2018](#)) are often used as benchmarks. Finally, the success of the M Forecasting competitions ([Makridakis, Spiliotis et al., 2020](#)) provide a benchmark dataset for general forecasting methodologies (see Section 2.12.7 and references therein).

A recent trend that has great future potential is the publication of websites that demonstrate the efficacy of different forecasting methods on real data. The Covid-19 Forecast Hub⁸² and the Business Forecast Lab⁸³ provide notable examples in the fields of epidemiology and macroeconomics and business respectively.

2.12.2. Point, interval, and pHDR forecast error measures⁸⁴

Point forecasts are single number forecasts for an unknown future quantity also given by a single number. *Interval forecasts* take the form of two point forecasts, an upper and a lower limit. Finally, a less common type of forecast would be a *predictive Highest Density Region* (pHDR), i.e., an HDR ([Hyndman, 1996](#)) for the conditional density of the future observable. pHDRs would be interesting for multimodal (possibly implicit)

predictive densities, e.g., in scenario planning. Once we have observed the corresponding realisation, we can evaluate our point, interval and pHDR forecasts. There are many common point forecast error measures (PFEMs), e.g., the mean squared error (MSE), mean absolute error (MAE), mean absolute scaled error (MASE), mean absolute percentage error (MAPE) or many others (see Section 3.4 in [Hyndman & Athanasopoulos, 2018](#)). Which one is most appropriate for our situation, or should we even use multiple different PFEMs?

There are many common point forecast error measures (PFEMs), e.g., the MSE, MAE, MASE, (s)MAPE, the quantile score or pinball loss or many others (e.g., sections 5.8 and 5.9 in [Hyndman & Athanasopoulos, 2021](#)). Assuming historical periods with observations and a forecasting horizon with observations and point forecasts, we have:

Let us take a step back. Assume we have a full density forecast and wish to “condense” it to a point forecast that will minimise some PFEM in expectation. The key observation is that *different PFEMs will be minimised by different point forecasts derived from the same density forecast* ([Kolassa, 2020b](#)).

- •
The MSE is minimised by the expectation.
- •
The MAE and MASE are minimised by the median ([Hanley, Joseph, Platt, Chung, & Belisle, 2001](#)).
- •
The MAPE is minimised by the α -median ([Gneiting, 2011a](#) p. 752 with $\alpha = 0.5$).
- •
The sMAPE is minimised by an unnamed functional that would need to be minimised numerically ([Gonçalves, 2015](#)).
- •
The hinge/tick/pinball loss is minimised by the appropriate α -quantile ([Gneiting, 2011b](#)).
- •
In general, there is no loss function that is minimised by the mode ([Heinrich, 2014](#)).

We note that intermittent demand (see Section 2.8) poses specific challenges. On the one hand, the MAPE is undefined if there are zeros in the actuals. On the other hand, the point forecasts minimising different PFEMs will be very different. For instance, the conditional median (minimising the MAE) may well be a flat zero, while the conditional mean (minimising the MSE) will usually be nonzero. Our forecasting algorithm may not output an explicit density forecast. It is nevertheless imperative to think about which functional of the implicit density we want to elicit (Gneiting, 2011a), and tailor our error measure – and forecasting algorithm! – to it. It usually makes no sense to evaluate a point forecast with *multiple* PFEMs (Kolassa, 2020b).

Interval forecasts can be specified in multiple ways. We can start with a probability coverage and require two appropriate quantiles – e.g., we could require a 2.5% and a 97.5% quantile forecast, yielding a symmetric or equal-tailed 95% interval forecast. Interval forecasts $(\hat{F}_\alpha, \hat{F}_\beta)$ of this form can be evaluated by the interval score (Brehmer and Gneiting, 2020, Winkler, 1972), a proper scoring rule (section 6.2 in Gneiting & Raftery, 2007):(3)

We can also use the hinge loss to evaluate the quantile forecasts separately.

Alternatively, we can require a shortest interval subject to a specified coverage. This interval is not elicitable relative to practically relevant classes of distributions (Brehmer and Gneiting, 2020, Fissler et al., 2020).

Yet another possibility is to maximise the interval forecast's probability coverage, subject to a maximum length l . This modal interval forecast is elicitable by an appropriate -zero-one-loss (Brehmer & Gneiting, 2020):(4)

The pHDR is not elicitable even for unimodal densities (Brehmer & Gneiting, 2020). In the multimodal case, the analysis is likely difficult. Nevertheless, a variation of the Winkler score has been proposed to evaluate pHDRs on an ad hoc basis (Hyndman, 2020). One could also compare the achieved to the nominal coverage, e.g., using a binomial test – which disregards the volume of the pHDR (Kolassa, 2020a).

In conclusion, there is a bewildering array of PFEMs, which require more thought in choosing among than is obvious at first glance. The difficulties involved in evaluating interval and pHDR forecasts motivate a stronger emphasis on full density forecasts (cf. Askanazi, Diebold, Schorfheide, & Shin, 2018 and Section 2.12.4).

2.12.3. Scoring expert forecasts⁸⁵

Evaluating forecasting capabilities can be a difficult task. One prominent way to evaluate an expert's forecast is to score the forecast once the realisation of the uncertainty is known. Scoring forecasts using the outcome's realisations over multiple forecasts offers insights into an individual's expertise. Experts can also use scoring information to identify ways to improve future forecasts. In addition, scoring rules and evaluation measures can be designed to match decision-making problems, incentivising forecasts that are most useful in a specific situation ([Winkler, Grushka-Cockayne, Lichtendahl, & Jose, 2019](#)).

Scoring rules were first suggested for evaluate meteorological forecasts in work by [Brier \(1950\)](#). Scoring rules have since been used in a wide variety of settings, such as business and other applications. When forecasting a discrete uncertainty with only two possible outcomes (e.g., a loan will be defaulted on or not, a customer will click on an ad or not), the Brier score assigns a score of $2(p - o)^2$, where p is the probability forecast reported that the event will occur. The greater the probability reported for an event that occurs, the higher the score the forecast receives. Over multiple forecasts, better forecasters will tend to have higher average Brier scores. For discrete events with more than two outcomes, a logarithmic scoring rule can be used.

The scoring rules are attractive to managers in practice since they are considered proper. Proper scoring rules (see also [Section 2.12.4](#)) incentivise honest forecasts from the experts, even prior to knowing the realisation of an uncertainty, since ex ante the expected score is maximised only when reported probabilities equals true beliefs ([Bickel, 2007](#), [Gneiting and Raftery, 2007](#), [Merkle and Steyvers, 2013](#), [O'Hagan et al., 2006](#), [Winkler et al., 1996](#)). Examples of a scoring rule that is not proper yet still commonly used are the linear score, which simply equals the reported probability or density for the actual outcome, or the skill score, which is the percentage improvement of the Brier score for the forecast relative to the Brier score of some base line naive forecast ([Winkler et al., 2019](#)).

For forecasting continuous quantities, forecasts could be elicited by asking for an expert's quantile (or fractile) forecast rather than a probability forecast. For instance, the 0.05, 0.25, 0.50, 0.75 and 0.95 quantiles are often elicited in practice, and in some cases every 0.01 quantile, between 0–1 are elicited (e.g., the 2014 Global Energy Forecasting Competition, [Hong et al., 2016](#)). Proper scoring rules for quantiles are developed in [Jose and Winkler \(2009\)](#).

When forecasts are used for decision-making, it is beneficial if the scoring rule used relates in some manner to the decision problem itself. In certain settings, the connection of the scoring rule to the decision context is straight forward. For example, [Jose, Nau, and Winkler \(2008\)](#) develop scoring rules that can be mapped to decision problems based on the decision maker's utility function. [Johnstone, Jose, and Winkler \(2011\)](#) develop tailored scoring rules aligning the interest of the forecaster and the decision maker. [Grushka-Cockayne, Lichtendahl, Jose and Winkler \(2017\)](#) link quantile scoring rules to business profit-sharing situations.

2.12.4. Evaluating probabilistic forecasts⁸⁶

Probabilistic forecasting is a term that is not strictly defined, but usually refers to everything beyond point forecasting ([Gneiting, 2011a](#)). However, in this section we consider only the evaluation of full predictive distributions or equivalent characterisations. For the evaluation of prediction of quantiles, intervals and related objects, see Section [2.12.2](#).

One crucial point for evaluating probabilistic forecasts is the reporting, which is highly influenced from meteorologic communities. From the theoretical point of view, we should always report the predicted cumulative distribution function of our prediction target. Alternatively for continuous data, reporting the probability density function is a popular choice. For univariate prediction problems a common alternative is to report quantile forecast on a dense grid of probabilities, as it approximates the full distribution ([Hong et al., 2016](#)). For multivariate forecasts, it seems to become standard to report a large ensemble (a set of simulated trajectories/paths) of the full predictive distribution. The reason is that the reporting of a multivariate distribution (or an equivalent characterisation) of sophisticated prediction models is often not feasible or practicable, especially for non-parametric or copula-based forecasting methods. In general, suitable tools for forecasting evaluation are proper scoring rules as they address calibration and sharpness simultaneously ([Gneiting and Katzfuss, 2014](#), [Gneiting and Raftery, 2007](#)). Preferably, we consider strictly proper scoring rules which can identify the true predicted distribution among a set of forecast candidates that contains the true model.

In the univariate case the theory is pretty much settled and there is quite some consensus about the evaluation of probabilistic forecasts ([Gneiting & Katzfuss, 2014](#)). The continuous ranked probability score (CRPS) and logarithmic scores (log-score) are

popular strictly proper scoring rules, while the quadratic and pseudospherical score remain strictly proper alternatives. The CRPS can be well approximated by averaging across quantile forecasts on an equidistant grid of probabilities (Nowotarski & Weron, 2018).

For multivariate forecast evaluation the situation is more complicated and many questions remain open (Gneiting and Raftery, 2007, Meng et al., 2020). The multivariate version of the log-score is a strictly proper scoring rule, but it requires the availability of a multivariate density forecast. This makes it impracticable for many applications. Gneiting and Raftery (2007) discuss the energy score, a multivariate generalisation of the CRPS, that is strictly proper. Still, it took the energy score more than a decade to increase its popularity in forecasting. A potential reason is the limited simulation study of Pinson and Tastu (2013) that concludes that the energy score can not discriminate well differences in the dependency structure. In consequence other scoring rules were proposed in literature, e.g., the variogram score (Scheuerer & Hamill, 2015) which is not strictly proper. Ziel and Berk (2019) consider a strictly proper scoring method for continuous variables using copula techniques. In contrast to Pinson and Tastu (2013), recent studies (Lerch et al., 2020, Ziel and Berk, 2019) show that the energy score discriminates well when used together with significance tests like the Diebold–Mariano (DM) test. In general, we recommended scoring be applied with reliability evaluation (see Section 2.12.5) and significance tests (see Section 2.12.6). Additionally, if we want to learn about the performance of our forecasts it is highly recommended to consider multiple scoring rules and evaluate on lower-dimensional subspaces. For multivariate problems, this holds particularly for the evaluation of univariate and bivariate marginal distributions.

2.12.5. Assessing the reliability of probabilistic forecasts⁸⁷

Probabilistic forecasts in the form of predictive distributions are central in risk-based decision making where reliability, or calibration, is a necessary condition for the optimal use and value of the forecast. A probabilistic forecast is calibrated if the observation cannot be distinguished from a random draw from the predictive distribution or, in the case of ensemble forecasts, if the observation and the ensemble members look like random draws from the same distribution. Additionally, to ensure their utility in decision making, forecasts should be sharp, or specific, see Sections 2.12.2 Point,

interval, and pHDR forecast error measures, 2.12.4 Evaluating probabilistic forecasts as well as (Gneiting et al., 2007).

In the univariate setting, several alternative notions of calibration exist for both a single forecast (Gneiting et al., 2007, Tsyplov, 2013) and a group of forecasts (Strähl & Ziegel, 2017). The notion most commonly used in applications is probabilistic calibration (Dawid, 1984); the forecast system is probabilistically calibrated if the probability integral transform (PIT) of a random observation, that is, the value of the predictive cumulative distribution function in the observation, is uniformly distributed. If the predictive distribution has a discrete component, a randomised version of the PIT should be used (Gneiting & Ranjan, 2013).

Probabilistic calibration is assessed visually by plotting the histogram of the PIT values over a test set. A calibrated forecast system will return a uniform histogram, a U-shape indicates overdispersion and a J-shape indicates underdispersion, while a systematic bias results in a biased histogram (e.g. Thorarinsdottir & Schuhen, 2018). The discrete equivalent of the PIT histogram, which applies to ensemble forecasts, is the verification rank histogram (Anderson, 1996, Hamill and Colucci, 1997). It shows the distribution of the ranks of the observations within the corresponding ensembles and has the same interpretation as the PIT histogram.

For small test sets, the bin number of a PIT/rank histogram must be chosen with care. With very few bins, the plot may obscure miscalibration while with many bins, even perfectly calibrated forecasts can yield non-uniformly appearing histograms (Heinrich, 2020, Thorarinsdottir and Schuhen, 2018). The bin number should be chosen based on the size of the test set, with the bin number increasing linearly with the size of the test set (Heinrich, 2020). More specifically, the uniformity of PIT/rank values can be assessed with statistical tests (Delle Monache et al., 2006, Taillardat et al., 2016, Wilks, 2019), where the test statistics can be interpreted as a distance between the observed and a flat histogram (Heinrich, 2020, Wilks, 2019). Testing predictive performance is further discussed in Section 2.12.6.

Calibration assessment of multivariate forecasts is complicated by the lack of a unique ordering in higher dimensions and the many ways in which the forecasts can be miscalibrated (Wilks, 2019). Gneiting, Stanberry, Gneiting, Held, and Johnson (2008) propose a general two-step approach where an ensemble forecast and the corresponding observation are first mapped to a single value by a pre-rank function.

Subsequently, the pre-rank function values are ranked in a standard manner. The challenge here is to find a pre-rank function that yields informative and discriminative ranking (Gneiting et al., 2008, Thorarinsdottir et al., 2016, Wilks, 2004); see Thorarinsdottir et al. (2016) and Wilks (2019) for comparative studies. Alternatively, Ziegel and Gneiting (2014) propose a direct multivariate extension of the univariate setting based on copulas.

2.12.6. Statistical tests of forecast performance⁸⁸

A natural consequence of growth in forecasting methodologies was the development of statistical tests for predictive ability in the last thirty years. These tests provided forecasters some formal reassurance that the predictive superiority of a leading forecast is statistically significant and is not merely due to random chance.

One of the early papers that undoubtedly sparked growth in this field was Diebold and Mariano (1995, DM hereafter). In their seminal paper, DM provided a simple yet general approach for testing equal predictive ability, i.e., if two forecasting sources (and ,) are equally accurate on average. Mathematically, if we denote the error for , and , the hypotheses for this DM test is : for all versus : under a loss function . Their population-level predictive ability test has very few assumptions (e.g., covariance stationary loss differential) and is applicable to a wide range of loss functions, multi-period settings, and wide class of forecast errors (e.g., non-Gaussian, serially and/or contemporaneously correlated). This test though not originally intended for models has been widely used by others to test forecasting models' accuracy (Diebold, 2015).

Modifications were later introduced by Harvey, Leybourne, and Newbold (1998) to improve small sample properties of the test. Generalisations and extensions have emerged to address issues that DM tests encountered in practice such as nested models (Clark and McCracken, 2001, Clark and McCracken, 2009), parameter estimation error (West, 1996), cointegrated variables (Corradi, Swanson, & Olivetti, 2001), high persistence (Rossi, 2005), and panel data (Timmermann & Zhu, 2019). Finite-sample predictive ability tests also emerged from the observation that models may have equal predictive ability in finite samples, which generated a class called conditional predictive accuracy tests (Clark and McCracken, 2013, Giacomini and White, 2006).

An alternative approach to comparing forecast accuracy is through the notion of forecast encompassing, which examines if a forecast encompasses all useful information from another with respect to predictions (Chong and Hendry, 1986, Clark and McCracken, 2001, Harvey et al., 1998). Though it has a few more assumptions, forecast encompassing tests in certain contexts might be preferable to the mean square prediction error tests à la Diebold–Mariano (Buseti & Marcucci, 2013). Another stream of available statistical tests looks at multiple forecasts simultaneously instead of pairs. Addressing a need for a reality check on “data snooping”, White (2000) later modified by Hansen (2005) developed a multiple model test that uses a null hypothesis of “superior predictive ability” instead of the equal predictive ability used in DM tests. These have also been generalised to deal with issues such as cointegrated variables (Corradi et al., 2001) and multi-horizon forecasts (Quaedvlieg, 2019). Recently, Li, Liao and Quaedvlieg (2020) proposed a conditional superior predictive ability test similar to Giacomini and White (2006)’s innovation to the DM test. A different approach for studying performance of multiple forecasting models is through the use of multiple comparison tests such as multiple comparison with a control and multiple comparison with the best (Edwards and Hsu, 1983, Horrace and Schmidt, 2000, Hsu, 1981). These tests often are based on jointly estimated confidence intervals that measure the difference between two parameters of interest such as the forecast accuracies of a model and a benchmark. Koning, Franses, Hibon, and Stekler (2005) illustrates how they can be ex post used to analyse forecasting performance in the M3 forecasting competition (Makridakis & Hibon, 2000) using model ranking instead of forecast accuracy scores as its primitives. The multiple comparison of the best was used in the analysis of the subsequent M4 and M5 Competitions (Makridakis, Spiliotis et al., 2020, Makridakis, Spiliotis, Assimakopoulos, 2021 and Section 2.12.7).

2.12.7. Forecasting competitions⁸⁹

Forecasting competitions provide a “playground” for academics, data scientists, students, practitioners, and software developers to compare the forecasting performance of their methods and approaches against others. Organisers of forecasting competitions test the performance of the participants’ submissions against some hidden data, usually the last window of observations for each series. The benefits from forecasting competitions are multifold. Forecasting competitions (i) motivate the development of innovative forecasting solutions, (ii) provide a deeper understanding of

the conditions that some methods work and others fail, (iii) promote knowledge dissemination, (iv) provide a much-needed, explicit link between theory and practice, and (v) leave as a legacy usable and well-defined data sets. Participation in forecasting competitions is sometimes incentivised by monetary prizes. However, the stakes are usually much higher, including reputational benefits.

The most famous forecasting competitions are the ones organised by Spyros Makridakis. Initially, the research question focused on the relative performance of simple versus complex forecast. M and M3 competitions (Makridakis et al., 1982, Makridakis and Hibon, 2000) empirically showed that simple methods (such as exponential smoothing; see Section 2.3.1) are equally good compared to other more complex methods and models (such as ARIMA and neural networks; see Section 2.3.4 and Section 2.7.8 respectively) in point-forecast accuracy – if not better. Moreover, the early Makridakis competitions showed the importance of forecast combinations in increasing predictive accuracy. For example, the winner of the M3 competition was the Theta method (see Section 2.3.3), a simple statistical method that involved the combination of linear regression and simple exponential smoothing forecasts (Assimakopoulos & Nikolopoulos, 2000).

The M4 competition (Makridakis, Spiliotis et al., 2020) challenged researchers and practitioners alike with a task of producing point forecasts and prediction intervals for 100 thousand time series of varied frequencies. This time, the main hypothesis focused on the ability of machine learning and neural network approaches in the task of time series forecasting. Machine learning approaches (see Section 2.7.10) that focused on each series independently performed poorly against statistical benchmarks, such as Theta, Damped exponential smoothing or simple averages of exponential smoothing models. However, the best two performing submissions in the M4 competition (Montero-Manso et al., 2020, Smyl, 2020) used neural network and machine learning algorithms towards utilising cross-learning. So, the main learning outcome from the M4 competition is that, if utilised properly, machine learning can increase the forecasting performance. Similarly to previous competitions, M4 demonstrated again the usefulness of combining across forecasts, with five out of the top six submissions offering a different implementation of forecast combinations. Several other forecasting competitions focused on specific contexts and applications. For example, M2 competition (Makridakis et al., 1993) suggested that the benefits from

additional information (domain expertise) are limited; see also Section 2.11.4. The tourism forecasting competition (Athanasopoulos, Hyndman, Song, & Wu, 2011) also showed that exogenous variables do not add value, while naive forecasts perform very well on a yearly frequency (for a discussion on tourism forecasting applications, see Section 3.8.1). The NN3 competition (Crone, Hibon, & Nikolopoulos, 2011) confirmed the superior performance of statistical methods, but noted that neural network approaches are closing the distance. Tao Hong's series of energy competitions (Hong et al., 2014, Hong et al., 2016, Hong et al., 2019) demonstrated best practices for load, price, solar, and wind forecasting, with extensions to probabilistic and hierarchical forecasts (for energy forecasting applications, see Section 3.4). Finally, many companies have hosted forecasting challenges through the Kaggle platform. Bojer and Meldgaard (2020) reviewed the Kaggle competitions over the last five years, and concluded that access to hierarchical information, cross-learning, feature engineering, and combinations (ensembles) can lead to increased forecasting performance, outperforming traditional statistical methods. These insights were a forerunner to the results of the M5 competition, which focused on hierarchically organised retail data (Makridakis, Spiliotis, Assimakopoulos, 2021, Makridakis, Spiliotis, Assimakopoulos, Chen et al., 2021).

Makridakis, Fry, Petropoulos and Spiliotis (2021) provide a list of design attributes for forecasting competitions and propose principles for future competitions.

2.13. The future of forecasting theory⁹⁰

The theory of forecasting appears mature today, based on dedicated developments at the interface among a number of disciplines, e.g., mathematics and statistics, computer sciences, psychology, etc. A wealth of these theoretical developments have originated from specific needs and challenges in different application areas, e.g., in economics, meteorology and climate sciences, as well as management science among others. In this section, many aspects of the theory of forecasting were covered, with aspects related to data, modelling and reasoning, forecast verification. Now, the fact that forecasting is mature does not mean that all has been done – we aim here at giving a few pointers at current and future challenges.

First of all, it is of utmost importance to remember that forecasting is a process that involves both quantitative aspects (based on data and models) and humans, at various levels, i.e., from the generation of forecasts to their use in decision-making. A first

consequence is that we always need to find, depending on the problem at hand, an optimal trade-off between data-driven approaches and the use of expert judgment. In parallel, forecasting is to be thought of in a probabilistic framework in a systematic manner (Gneiting & Katzfuss, 2014). This allows us to naturally convey uncertainty about the future, while providing the right basis to make optimal decisions in view of the characteristics of the decision problem, as well as the loss (or utility) function and risk aversion of the decision maker. Another consequence is that using forecasts as input to decision-making often affects the outcome to be predicted itself – a problem known as self-negating forecasts (possibly also self-fulfilling) or the prophet dilemma. With advances in the science of dynamic systems and game theory, we should invest in modelling those systems as a whole (i.e., forecasting and decision-making) in order to predict the full range of possible outcomes, based on the decisions that could be made. In parallel, it is clear that today, the amount of data being collected and possibly available for forecasting is growing at an astounding pace. This requires re-thinking our approaches to forecasting towards high-dimensional models, online learning, etc. Importantly, the data being collected is distributed in terms of ownership. And, due to privacy concerns and competitive interests, some may not be ready to share their data. Novel frameworks to learning and forecasting ought to be developed with that context in mind, for instance focusing on distributed and privacy-preserving learning – an example among many others is that of Google pushing forward federated learning (Abadi et al., 2016), an approach to deep learning where the learning process is distributed and with a privacy layer. Eventually the access and use of data, as well as the contribution to distributed learning (and collaborative analytics, more generally), may be monetised, bringing a mechanism design component to the future theory of forecasting. A simple and pragmatic example is that of forecast reconciliation: if asking various agents to modify their forecasts to make them coherent within a hierarchy, such modifications could be monetised to compensate for accuracy loss. A large part of today's modelling and forecasting approaches uses a wealth of data to identify and fit models, to be eventually used to forecast based on new data and under new conditions. Different approaches have been proposed to maximise the generalisation ability of those models, to somewhat maximise chances to do well out-of-sample. At the root of this problem is the effort to go beyond correlation only, and to identify causality (see, e.g., Pearl (2009) for a recent extensive coverage). While

causality has been a key topic of interest to forecasters for a long time already, new approaches and concepts are being pushed forward for identification of and inference in causal models (Peters, Janzing, & Schölkopf, 2017), which may have a significant impact on the theory of forecasting.

Eventually, the key question of *what a good forecast is* will continue to steer new developments in the theory of forecasting in the foreseeable future. The nature of goodness of forecasts (seen from the meteorological application angle) was theorised a few decades ago already (Murphy, 1993), based on consistency, quality and value. We still see the need to work further on that question – possibly considering these 3 pillars, but possibly also finding other ways to define desirable properties of forecasts. This will, in all cases, translate to further developing frameworks for forecast verification, focusing on the interplay between forecast quality and value, but also better linking to psychology and behavioural economics. In terms of forecast verification, some of the most pressing areas most likely relate to (multivariate) probabilistic forecasting and to the forecasting of extreme events. When it comes to forecast quality and value, we need to go beyond the simple plugging of forecasts into decision problems to assess whether this yields better decisions, or not. Instead, we ought to propose suitable theoretical frameworks that allow assessing whether certain forecasts are fundamentally better (than others) for given classes of decision problems. Finally, the link to psychology and behavioural economics should ensure a better appraisal of how forecasts are to be communicated, how they are perceived and acted upon.

Most of the advances in the science of forecasting have come from the complementarity between theoretical developments and applications. We can then only be optimistic for the future since more and more application areas are relying heavily on forecasting. Their specific needs and challenges will continue fuelling upcoming developments in the theory of forecasting.

3. Practice

3.1. Introduction to forecasting practice⁹¹

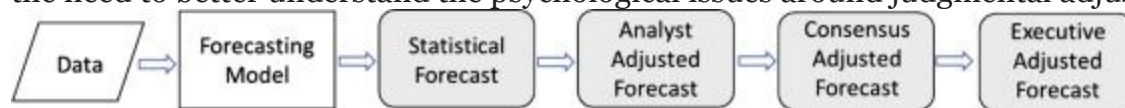
The purpose of forecasting is to improve decision making in the face of uncertainty. To achieve this, forecasts should provide an unbiased guess at what is most likely to happen

(the point forecast), along with a measure of uncertainty, such as a prediction interval (PI). Such information will facilitate appropriate decisions and actions.

Forecasting should be an objective, dispassionate exercise, one that is built upon facts, sound reasoning, and sound methods. But since forecasts are created in social settings, they are influenced by organisational politics and personal agendas. As a consequence, forecasts will often reflect aspirations rather than unbiased projections.

In organisations, forecasts are created through processes that can involve multiple steps and participants. The process can be as simple as executive fiat (also known as evangelical forecasting), unencumbered by what the data show. More commonly, the process begins with a statistical forecast (generated by forecasting software), which is then subject to review and adjustment, as illustrated in Fig. 5.

In concept, such an elaborate multi-stage process allows “management intelligence” to improve forecast quality, incorporating information not accounted for in the statistical model. In reality, however, benefits are not assured. Lawrence, Goodwin, O’Connor, and Önkal (2006) reviewed more than 200 studies, concluding that human judgment can be of significant benefit but is also subject to significant biases. Among the many papers on this subject, there is general agreement on the need to track and review overrides, and the need to better understand the psychological issues around judgmental adjustments.



1. [Download: Download high-res image \(86KB\)](#)
2. [Download: Download full-size image](#)

Fig. 5. Multi-stage forecasting process.

The underlying problem is that each human touch point subjects the forecast to the interests of the reviewers – and these interests may not align with creating an accurate, unbiased forecast. To identify where such problems are occurring, Forecast Value Added (FVA) analysis is an increasingly popular approach among practitioners.

FVA is defined as the change in a forecasting performance metric that can be attributed to a particular step or participant in the forecasting process (Gilliland, 2002). Any activity that fails to deliver positive FVA (i.e., fails to improve forecast quality) is considered process waste.

Starting with a naive forecast, FVA analysis seeks to determine whether each subsequent step in the process improves upon the prior steps. The “stairstep report” of Table 1 is a

familiar way of summarising results, as in this example from Newell Rubbermaid (Schubert & Rickard, 2011).

Here, averaged across all products, naive (random walk) achieved forecast accuracy of 60%. The company’s statistical forecast delivered five percentage points of improvement, but management review and adjustment delivered negative value. Such findings – not uncommon – urge further investigation into causes and possible process corrections (such as training reviewers or limiting adjustments). Alternatively, the management review step could be eliminated, providing the dual benefits of freeing up management time spent on forecasting and, on average, more accurate forecasts.

Table 1. Stairstep report showing FVA results.

Process step	Forecast accuracy (100%–MAPE)	FVA vs. Naive	FVA vs. statistical
Naive forecast	60%		
Statistical forecast	65%	5%	
Adjusted forecast	62%	2%	–3%

Morlidge (2014c) expanded upon FVA analysis to present a strategy for prioritising judgmental adjustments, finding the greatest opportunity for error reduction in products with high volume and high relative absolute error. Chase (2021) described a machine learning (ML) method to guide forecast review, identifying which forecasts are most likely to benefit from adjustment along with a suggested adjustment range. Baker (2021) used ML classification models to identify characteristics of non-value adding overrides, proposing the behavioural economics notion of a “nudge” to prompt desired forecaster behaviour. Further, Goodwin, Petropoulos, and Hyndman (2017) derived upper bounds for FVA relative to naive forecasts. And de Kok (2017) created a Stochastic Value Added (SVA) metric to assess the difference between actual and forecasted distributions, knowledge of which is valuable for inventory management. Including an indication of uncertainty around the point forecast remains an uncommon practice. Prediction intervals in software generally underestimate uncertainty, often dramatically, leading to unrealistic confidence in the forecast. And even when provided, PIs largely go unused by practitioners. Goodwin (2014) summarised the psychological issues, noting that the generally poor calibration of the PIs may not explain the reluctance to utilise them. Rather, “an interval forecast may accurately reflect the

uncertainty, but it is likely to be spurned by decision makers if it is too wide and judged to be uninformative” (Goodwin, 2014 page 5).

It has long been recognised (Chatfield, 1986, Lawrence, 2000) that the practice of forecasting falls well short of the potential exhibited in academic research, and revealed by the M forecasting competitions. In the M4, a simple benchmark combination method (the average of Single, Holt, and Damped exponential smoothing) reduced the overall weighted average (OWA) error by 17.9% compared to naive. The top six performing methods in M4 further reduced OWA by over 5% compared to the combination benchmark (Makridakis, Spiliotis et al., 2020). But in forecasting practice, just bettering the accuracy of naive has proven to be a surprising challenge. Morlidge’s (2014b) study of eight consumer and industrial businesses found 52% of their forecasts failed to do so. And, as shown, Newel Rubbermaid beat naive by just two percentage points after management adjustments.

Ultimately, forecast accuracy is limited by the nature of the behaviour being forecast. But even a highly accurate forecast is of little consequence if overridden by management and not used to enhance decision making and improve organisational performance. Practitioners need to recognise limits to forecastability and be willing to consider alternative (non-forecasting) approaches when the desired level of accuracy is not achievable (Gilliland, 2010). Alternatives include supply chain re-engineering – to better react to unforeseen variations in demand, and demand smoothing – leveraging pricing and promotional practices to shape more favourable demand patterns.

Despite measurable advances in our statistical forecasting capabilities (Makridakis, Hyndman & Petropoulos, 2020), it is questionable whether forecasting practice has similarly progressed. The solution, perhaps, is what Morlidge (2014a, page 39) suggests that “users should focus less on trying to optimise their forecasting process than on detecting where their process is severely suboptimal and taking measures to redress the problem”. This is where FVA can help.

For now, the challenge for researchers remains: To prompt practitioners to adopt sound methods based on the objective assessment of available information, and avoid the “worst practices” that squander resources and fail to improve the forecast.

3.2. Operations and supply chain management

3.2.1. Demand management⁹²

Demand management is one of the dominant components of supply chain management (Fildes, Goodwin, & Lawrence, 2006). Accurate demand estimate of the present and future is a first vital step for almost all aspects of supply chain optimisation, such as inventory management, vehicle scheduling, workforce planning, and distribution and marketing strategies (Kolassa & Siemsen, 2016). Simply speaking, better demand forecasts can yield significantly better supply chain management, including improved inventory management and increased service levels. Classic demand forecasts mainly rely on qualitative techniques, based on expert judgment and past experience (e.g., Weaver, 1971), and quantitative techniques, based on statistical and machine learning modelling (e.g., Bacha and Meyer, 1992, Taylor, 2003b). A combination of qualitative and quantitative methods is also popular and proven to be beneficial in practice by, e.g., judgmental adjustments (Önkal and Gönül, 2005, Syntetos, Kholidasari and Naim, 2016, Turner, 1990 and Section 2.11.2), judgmental forecast model selection (Han et al., 2019, Petropoulos et al., 2018b and Section 2.11.3), and other advanced forecasting support systems (Arvan et al., 2019, Baecke et al., 2017 see also Section 3.7.1).

The key challenges that demand forecasting faces vary from domain to domain. They include:

- 1.
The existence of intermittent demands, e.g., irregular demand patterns of fashion products. According to Nikolopoulos (2020), limited literature has focused on intermittent demand. The seminal work by Croston (1972) was followed by other representative methods such as the SBA method by Syntetos and Boylan (2001), the aggregate–disaggregate intermittent demand approach (ADIDA) by Nikolopoulos et al. (2011), the multiple temporal aggregation by Petropoulos and Kourentzes (2015), and the nearest neighbour (NN) based approach by Nikolopoulos et al. (2016). See Section 2.8 for more details on intermittent demand forecasting and Section 2.10.2 for a discussion on temporal aggregation.
- 2.
The emergence of new products. Recent studies on new product demand forecasting are based on finding analogies (Hu et al., 2019, Wright and Stern, 2015), leveraging comparable products (Baardman, Levin, Perakis, & Singhvi,

2018), and using external information like web search trends (Kulkarni, Kannan, & Moe, 2012). See Section 3.2.6 for more details on new product demand forecasting.

- 3. The existence of short-life-cycle products, e.g., smartphone demand (e.g., Chung et al., 2012, Shi et al., 2020, Szozda, 2010).
- 4. The hierarchical structure of the data such as the electricity demand mapped to a geographical hierarchy (e.g., Athanasopoulos et al., 2009, Hong et al., 2019, Hyndman et al., 2011 but also Section 2.10.1).

With the advent of the big data era, a couple of coexisting new challenges have drawn the attention of researchers and practitioners in the forecasting community: the need to forecast a large volume of related time series (e.g., thousands or millions of products from one large retailer: Salinas, Michael et al., 2019), and the increasing number of external variables that have significant influence on future demand (e.g., massive amounts of keyword search indices that could impact future tourism demand (Law, Li, Fong, & Han, 2019)). Recently, to deal with these new challenges, numerous empirical studies have identified the potentials of deep learning based global models, in both point and probabilistic demand forecasting (e.g., Bandara et al., 2020b, Rangapuram et al., 2018, Salinas, Michael et al., 2019, Wen et al., 2017). With the merits of cross-learning, global models have been shown to be able to learn long memory patterns and related effects (Montero-Manso & Hyndman, 2020), latent correlation across multiple series (Smyl, 2020), handle complex real-world forecasting situations such as data sparsity and cold-starts (Chen, Kang, Chen, & Wang, 2020), include exogenous covariates such as promotional information and keyword search indices (Law et al., 2019), and allow for different choices of distributional assumptions (Salinas, Michael et al., 2019).

3.2.2. Forecasting in the supply chain⁹³

A supply chain is ‘a network of stakeholders (e.g., retailers, manufacturers, suppliers) who collaborate to satisfy customer demand’ (Perera et al., 2019). Forecasts inform many supply chain decisions, including those relating to inventory control, production planning, cash flow management, logistics and human resources (also see Section 3.2.1). Typically, forecasts are based on an amalgam of statistical methods and management

judgment (Fildes & Goodwin, 2007). Hofmann and Rutschmann (2018) have investigated the potential for using big data analytics in supply chain forecasting but indicate more research is needed to establish its usefulness.

In many organisations forecasts are a crucial element of Sales and Operations Planning (S&OP), a tool that brings together different business plans, such as those relating to sales, marketing, manufacturing and finance, into one integrated set of plans (Thomé, Scavarda, Fernandez, & Scavarda, 2012). The purposes of S&OP are to balance supply and demand and to link an organisation's operational and strategic plans. This requires collaboration between individuals and functional areas at different levels because it involves data sharing and achieving a consensus on forecasts and common objectives (Mello, 2010). Successful implementations of S&OP are therefore associated with forecasts that are both aligned with an organisation's needs and able to draw on information from across the organisation. This can be contrasted with the 'silo culture' identified in a survey of companies by Moon, Mentzer, and Smith (2003) where separate forecasts were prepared by different departments in 'islands of analysis'. Methods for reconciling forecasts at different levels in both cross-sectional hierarchies (e.g., national, regional and local forecasts) and temporal hierarchies (e.g., annual, monthly and daily forecasts) are also emerging as an approach to break through information silos in organisations (see Section 2.10.1, Section 2.10.2, and Section 2.10.3). Cross-temporal reconciliation provides a data-driven approach that allows information to be drawn from different sources and levels of the hierarchy and enables this to be blended into coherent forecasts (Kourentzes & Athanasopoulos, 2019).

In some supply chains, companies have agreed to share data and jointly manage planning processes in an initiative known as Collaborative Planning, Forecasting, and Replenishment (CPFR) (Seifert, 2003 also see Section 3.2.3). CPFR involves pooling information on inventory levels and on forthcoming events, like sales promotions. Demand forecasts can be shared, in real time via the Internet, and discrepancies between them reconciled. In theory, information sharing should reduce forecast errors. This should mitigate the 'bullwhip effect' where forecast errors at the retail-end of supply chains cause upstream suppliers to experience increasingly volatile demand, forcing them to hold high safety stock levels (Lee et al., 2007). Much research demonstrating the benefits of collaboration has involved simulated supply chains (Fildes, 2017). Studies of real companies have also found improved performance

through collaboration (e.g., [Boone and Ganeshan, 2008](#), [Eksoz et al., 2019](#), [Hill et al., 2018](#)), but case study evidence is still scarce ([Syntetos, Babai, Boylan, Kolassa, & Nikolopoulos, 2016](#)). The implementation of collaborative schemes has been slow with many not progressing beyond the pilot stage ([Galbreth et al., 2015](#), [Panahifar et al., 2015](#)). Barriers to successful implementation include a lack of trust between organisations, reward systems that foster a silo mentality, fragmented forecasting systems within companies, incompatible systems, a lack of relevant training and the absence of top management support ([Fliedner, 2003](#), [Thomé et al., 2014](#)). Initiatives to improve supply chain forecasting can be undermined by political manipulation of forecasts and gaming. Examples include ‘enforcing’: requiring inflated forecasts to align them with sales or financial goals, ‘sandbagging’: underestimating sales so staff are rewarded for exceeding forecasts, and ‘spinning’: manipulating forecasts to garner favourable reactions from colleagues ([Mello, 2009](#)). [Pennings, van Dalen, and Rook \(2019\)](#) discuss schemes for correcting such intentional biases. For a discussion of the forecasting of returned items in supply chains, see Section [3.2.9](#), while Section [3.9](#) offers a discussion of possible future developments in supply chain forecasting.

3.2.3. Forecasting for inventories⁹⁴

Three aspects of the interaction between forecasting and inventory management have been studied in some depth and are the subject of this review: the bullwhip effect, forecast aggregation, and performance measurement.

The ‘bullwhip effect’ occurs whenever there is amplification of demand variability through the supply chain ([Lee, Padmanabhan, & Whang, 2004](#)), leading to excess inventories. This can be addressed by supply chain members sharing downstream demand information, at stock keeping unit level, to take advantage of less noisy data. Analytical results on the translation of ARIMA (see Section [2.3.4](#)) demand processes have been established for order-up-to inventory systems ([Gilbert, 2005](#)). There would be no value in information sharing if the wholesaler can use such relationships to deduce the retailer’s demand process from their orders (see, for example, [Graves, 1999](#)). Such deductions assume that the retailer’s demand process and demand parameters are common knowledge to supply chain members. [Ali and Boylan \(2011\)](#) showed that, if such common knowledge is lacking, there is value in sharing the demand data itself and [Ali, Boylan, and Syntetos \(2012\)](#) established relationships between accuracy gains

and inventory savings. Analytical research has tended to assume that demand parameters are known. [Pastore, Alfieri, Zotteri, and Boylan \(2020\)](#) investigated the impact of demand parameter uncertainty, showing how it exacerbates the bullwhip effect.

Forecasting approaches have been developed that are particularly suitable in an inventory context, even if not originally proposed to support inventory decisions. For example, [Nikolopoulos et al. \(2011\)](#) proposed that forecasts could be improved by aggregating higher frequency data into lower frequency data (see also Section [2.10.2](#); other approaches are reviewed in Section [3.2.1](#)). Following this approach, forecasts are generated at the lower frequency level and then disaggregated, if required, to the higher frequency level. For inventory replenishment decisions, the level of aggregation may conveniently be chosen to be the lead time, thereby taking advantage of the greater stability of data at the lower frequency level, with no need for disaggregation.

The variance of forecast errors over lead time is required to determine safety stock requirements for continuous review systems. The conventional approach is to take the variance of one-step-ahead errors and multiply it by the lead time. However, this estimator is unsound, even if demand is independent and identically distributed, as explained by [Prak, Teunter, and Syntetos \(2017\)](#). A more direct approach is to smooth the mean square errors over the lead time ([Syntetos & Boylan, 2006](#)).

[Strijbosch and Moors \(2005\)](#) showed that unbiased forecasts will not necessarily lead to achievement, on average, of target cycle service levels or fill rates. [Wallström and Segerstedt \(2010\)](#) proposed a 'Periods in Stock' measure, which may be interpreted, based on a 'fictitious stock', as the number of periods a unit of the forecasted item has been in stock or out of stock. Such measures may be complemented by a detailed examination of error-implication metrics ([Boylan & Syntetos, 2006](#)). For inventory management, these metrics will typically include inventory holdings and service level implications (e.g., cycle service level, fill rate). Comparisons may be based on total costs or via 'exchange curves', showing the trade-offs between service and inventory holding costs. Comparisons such as these are now regarded as standard in the literature on forecasting for inventories and align well with practice in industry.

3.2.4. Forecasting in retail⁹⁵

Retail companies depend crucially on accurate demand forecasting to manage their supply chain and make decisions concerning planning, marketing, purchasing,

distribution and labour force. Inaccurate forecasts lead to unnecessary costs and poor customer satisfaction. Inventories should be neither too high (to avoid waste and extra costs of storage and labour force), nor too low (to prevent stock-outs and lost sales, [Ma & Fildes, 2017](#)).

Forecasting retail demand happens in a three-dimensional space ([Syntetos et al., 2016](#)): the position in the supply chain hierarchy (store, distribution centre, or chain), the level in the product hierarchy (SKU, brand, category, or total) and the time granularity (day, week, month, quarter, or year). In general, the higher is the position in the supply chain, the lower is the time granularity required, e.g., retailers need daily forecasts for store replenishment and weekly forecasts for DC distribution/logistics activities at the SKU level ([Fildes, Ma, & Kolassa, 2019b](#)). Hierarchical forecasting (see Section [2.10.1](#)) is a promising tool to generate coherent demand forecasts on multiple levels over different dimensions ([Oliveira & Ramos, 2019](#)).

Several factors affect retail sales, which often increase substantially during holidays, festivals, and other special events. Price reductions and promotions on own and competitors' products, as well as weather conditions or pandemics, can also change sales considerably ([Huang, Fildes, & Soopramanien, 2019](#)).

Zero sales due to stock-outs or low demand occur very often at the SKU store level, both at weekly and daily granularity. The most appropriate forecasting approaches for intermittent demand are Croston's method ([Croston, 1972](#)), the Syntetos-Boylan approximation (SBA; [Syntetos & Boylan, 2005](#)), and the TSB method ([Teunter, Syntetos, & Zied Babai, 2011](#)), all introduced in Section [2.8.1](#). These methods have been used to forecast sales of spare parts in automotive and aerospace industries but have not yet been evaluated in the retail context.

Univariate forecasting models are the most basic methods retailers may use to forecast demand. They range from simple methods such as simple moving averages or exponential smoothing to ARIMA and ETS models (discussed in Section [2.3](#)). These are particularly appropriate to forecast demand at higher aggregation levels ([Ramos and Oliveira, 2016](#), [Ramos et al., 2015](#)). The main advantage of linear causal methods such as multiple linear regression is to allow the inclusion of external effects discussed above. There is no clear evidence yet that nonlinear models and novel machine learning methods can improve forecast accuracy ([Fildes et al., 2019b](#)).

To be effective, point estimates should be combined with quantile predictions or prediction intervals for determining safety stock amounts needed for replenishment. However, to the best of our knowledge this is an under-investigated aspect of retail forecasting (Kolassa, 2016, Taylor, 2007).

The online channel accounts for an ever-increasing proportion of retail sales and poses unique challenges to forecasting, beyond the characteristics of brick and mortar (B&M) retail stores. First, there are multiple *drivers or predictors* of demand that could be leveraged in online retail, but not in B&M:

- •
Online retailers can fine-tune customer interactions, e.g., through the landing page, product recommendations, or personalised promotions, leveraging the customer's purchasing, browsing or returns history, current shopping cart contents, or the retailer's stock position, in order to tailor the message to one specific customer in a way that is impossible in B&M.
- •
Conversely, product reviews are a type of interaction between the customer and the retailer and other customers which drives future demand.

Next, there are differences in forecast *use*:

- •
Forecast use strongly depends on the retailer's omnichannel strategy (Armstrong, 2017, Melacini et al., 2018, Sopadjieva et al., 2017): e.g., for "order online, pick up in store" or "ship from store" fulfillment, we need separate but related forecasts for both total online demand and for the demand fulfilled at each separate store.
- •
Online retailers, especially in fashion, have a much bigger problem with product returns. They may need to forecast how many products are returned overall (e.g., Shang, McKie, Ferguson, & Galbreth, 2020), or whether a *specific* customer will return a *specific* product.

Finally, there are differences in the forecasting *process*:

- •
B&M retailers decouple pricing/promotion decisions and optimisation from the customer interaction, and therefore from forecasting. Online, this is not possible, because the customer has total transparency to competitors' offerings. Thus,

online pricing needs to react much more quickly to competitive pressures – faster than the forecasting cycle.

-
-

Thus, the specific value of predictors is often not known at the time of forecasting: we don't know yet which customer will log on, so we don't know yet how many people will see a particular product displayed on their personalised landing page. (Nor do we know today what remaining stock will be displayed.) Thus, changes in drivers need to be “baked into” the forecasting algorithm.

-
-

Feedback loops between forecasting and other processes are thus even more important online: yesterday's forecasts drive today's stock position, driving today's personalised recommendations, driving demand, driving today's forecasts for tomorrow. Overall, online retail forecasting needs to be more agile and responsive to the latest interactional decisions taken in the web store, and more tightly integrated into the retailer's interactional tactics and omnichannel strategy.

Systematic research on demand forecasting in an online or omnichannel context is only starting to appear (e.g. [Omar, Klibi, Babai, & Ducq, 2021](#) who use basket data from online sales to improve omnichannel retail forecasts).

3.2.5. Promotional forecasting⁹⁶

Promotional forecasting is central for retailing (see Section [3.2.4](#)), but also relevant for many manufacturers, particularly of Fast Moving Consumer Goods (FMCG). In principle, the objective is to forecast sales, as in most business forecasting cases. However, what sets promotional forecasting apart is that we also make use of information about promotional plans, pricing, and sales of complementary and substitute products ([Bandyopadhyay, 2009](#), [Zhang et al., 2008](#)). Other relevant variables may include store location and format, variables that capture the presentation and location of a product in a store, proxies that characterise the competition, and so on ([Andrews et al., 2008](#), [Van Heerde et al., 2002](#)).

Three modelling considerations guide us in the choice of models. First, promotional (and associated effects) are proportional. For instance, we do not want to model the increase in sales as an absolute number of units, but instead, as a percentage uplift. We do this to not only make the model applicable to both smaller and larger applications,

for example, small and large stores in a retailing chain, but also to gain a clearer insight into the behaviour of our customers. Second, it is common that there are synergy effects. For example, a promotion for a product may be offset by promotions for substitute products. Both these considerations are easily resolved if we use multiplicative regression models. However, instead of working with the multiplicative models, we rely on the logarithmic transformation of the data (see Section 2.2.1) and proceed to construct the promotional model using the less cumbersome additive formulation (see Section 2.3.2). Third, the objective of promotional models does not end with providing accurate predictions. We are also interested in the effect of the various predictors: their elasticity. This can in turn provide the users with valuable information about the customers, but also be an input for constructing optimal promotional and pricing strategies (Zhang et al., 2008).

Promotional models have been widely used on brand-level data (for example, Divakar, Ratchford, & Shankar, 2005). However, they are increasingly used on Stock Keeping Unit (SKU) level data (Ma et al., 2016, Trapero et al., 2015), given advances in modelling techniques. Especially at that level, limited sales history and potentially non-existing examples of past promotions can be a challenge. Trapero et al. (2015) consider this problem and propose using a promotional model that has two parts that are jointly estimated. The first part focuses on the time series dynamics and is modelled locally for each SKU. The second part tackles the promotional part, which pools examples of promotions across SKUs to enable providing reasonable estimates of uplifts even for new SKUs. To ensure the expected heterogeneity in the promotional effects, the model is provided with product group information. Another recent innovation is looking at modelling promotional effects both at the aggregate brand or total sales level, and disaggregate SKU level, relying on temporal aggregation (Kourentzes & Petropoulos, 2016 and Section 2.10.2). Ma et al. (2016) concern themselves with the intra-and inter-category promotional information. The challenge now is the number of variables to be considered for the promotional model, which they address by using sequential LASSO (see also Section 2.5.3). Although the aforementioned models have shown very promising results, one has to recognise that in practice promotions are often forecasted using judgmental adjustments, with inconsistent performance (Trapero, Pedregal, Fildes, & Kourentzes, 2013); see also Section 2.11.2 and Section 3.7.3.

3.2.6. New product forecasting⁹⁷

Forecasting the demand for a new product accurately has even more consequence with regards to well-being of the companies than that for a product already in the market. However, this task is one of the most difficult tasks managers must deal with simply because of non-availability of past data (Wind, 1981). Much work has been going on for the last five decades in this field. Despite his Herculean attempt to collate the methods reported, Assmus (1984) could not list all even at that time. The methods used before and since could be categorised into three broad approaches (Goodwin, Dyussekeneva & Meeran, 2013) namely management judgment, consumer judgment and diffusion/formal mathematical models. In general, the hybrid methods combining different approaches have been found to be more useful (Hyndman and Athanasopoulos, 2018, Peres et al., 2010). Most of the attempts in New product Forecasting (NPF) have been about forecasting 'adoption' (i.e., enumerating the customers who bought at least one time) rather than 'sales', which accounts for repeat purchases also. In general, these attempts dealt with point forecast although there have been some attempts in interval and density forecasting (Meade & Islam, 2001). Out of the three approaches in NPF, management judgment is the most used approach (Gartner and Thomas, 1993, Kahn, 2002, Lynn et al., 1999) which is reported to have been carried out by either individual managers or group of them. Ozer (2011) and Surowiecki (2005) articulated their contrasting benefits and deficits. The Delphi method (see Section 2.11.4) has combined the benefits of these two modes of operation (Rowe & Wright, 1999) which has been effective in NPF. Prediction markets in the recent past offered an alternative way to aggregate forecasts from a group of Managers (Meeran et al., 2013, Wolfers and Zitzewitz, 2004) and some successful application of prediction markets for NPF have been reported by Karniouchina (2011) and Plott and Chen (2002).

In the second category, customer surveys, among other methods, are used to directly ask the customers the likelihood of them purchasing the product. Such surveys are found to be not very reliable (Morwitz, 1997). An alternative method to avoid implicit bias associated with such surveys in extracting inherent customer preference is conjoint analysis, which makes implicit trade off customers make between features explicit by analysing the customers' preference for different variants of the product. One analysis technique that attempts to mirror real life experience more is Choice Based Conjoint analysis (CBC) in which customers choose the most preferred product among available

choices. Such CBC models used together with the analysis tools such as Logit (McFadden, 1977) have been successful in different NPF applications (Meeran, Jahanbin, Goodwin, & Quariguasi Frota Neto, 2017).

In the third approach, mathematical/formal models known as growth or diffusion curves (see Section 2.3.18 and Section 2.3.19) have been used successfully to do NPF (Hu et al., 2019). The non-availability of past data is mitigated by growth curves by capturing the generic pattern of the demand growth of a class of products, which could be defined by a limited number of parameters such as saturation level, inflexion point, etc. For a new product a growth curve can be constituted from well-estimated parameters using analogous products, market intelligence or regression methods. Most extensively used family of growth curves for NPF has started with Bass model (Bass, 1969) that has been extended extensively (Bass et al., 2001, Easingwood et al., 1983, Islam and Meade, 2000, Peres et al., 2010, Simon and Sebastian, 1987). A recent applications of NPF focused on consumer electronic goods using analogous products (Goodwin, Dyussekeneva et al., 2013).

3.2.7. Spare parts forecasting⁹⁸

Spare parts are ubiquitous in modern societies. Their demand arises whenever a component fails or requires replacement. Demand for spare parts is typically intermittent, which means that it can be forecasted using the plethora of parametric and non-parametric methods presented in Section 2.8. In addition to the intermittence of demand, spare parts have two additional characteristics that make them different from Work-In-Progress and final products, namely: (i) they are generated by maintenance policies and part breakdowns, and (ii) they are subject to obsolescence (Bacchetti and Saccani, 2012, Kennedy et al., 2002).

The majority of forecasting methods do not link the demand to the generating factors, which are often related to maintenance activities. The demand for spare parts originates from the replacement of parts in the installed base of machines (i.e., the location and number of products in use), either preventively or upon breakdown of the part (Kim, Dekker, & Heij, 2017). Fortuin (1984) claims that using installed base information to forecast the spare part demand can lead to stock reductions of up to 25%. An overview of the literature that deals with spare parts forecasting with installed base information is given by Van der Auweraer and Boute (2019). Spare parts demand can be driven by the result of maintenance inspections and, in this case, a maintenance-based forecasting

model should then be considered to deal with this issue. Such forecasting models include the Delay Time (DT) model analysed in [Wang and Syntetos \(2011\)](#). Using the fitted values of the distribution parameters of a data set related to a hospital pumps, [Wang and Syntetos \(2011\)](#) have shown that when the failure and fault arriving characteristics of the items can be captured, it is recommended to use the DT model to forecast the spare part demand with a higher forecast accuracy. However, when such information is not available, then time series forecasting methods, such as those presented in Section 2.8.1, are recommended. The maintenance based forecasting is further discussed in Section 3.2.8.

Given the life cycle of products, spare parts are associated with a risk of obsolescence. [Molenaers, Baets, Pintelon, and Waeyenbergh \(2012\)](#) discussed a case study where 54% of the parts stocked at a large petrochemical company had seen no demand for the last 5 years. [Hinton \(1999\)](#) reported that the US Department of Defence was holding 60% excess of spare parts, with 18% of the parts (with a total value of \$1.5 billion) having no demand at all. To take into account the issue of obsolescence in spare parts demand forecasting, [Teunter et al. \(2011\)](#) have proposed the TSB method, which deals with linearly decreasing demand and sudden obsolescence cases. By means of an empirical investigation based on the individual demand histories of 8000 spare parts SKUs from the automotive industry and the Royal Air Force (RAF, UK), [Babai, Syntetos, and Teunter \(2014\)](#) have demonstrated the high forecast accuracy and inventory performance of the TSB method. Other variants of the Croston's method developed to deal with the risk of obsolescence in forecasting spare parts demand include the Hyperbolic-Exponential Smoothing method proposed by [Prestwich, Tarim, Rossi, and Hnich \(2014\)](#) and the modified Croston's method developed by [Babai, Dallery, Boubaker, and Kalai \(2019\)](#).

3.2.8. Predictive maintenance⁹⁹

A common classification of industrial maintenance includes three types of maintenance ([Montero Jimenez, Schwartz, Vingerhoeds, Grabot, & Salaün, 2020](#)). Corrective maintenance refers to maintenance actions that occur after the failure of a component. Preventive maintenance consists of maintenance actions that are triggered after a scheduled number of units as cycles, kilometers, flights, etc. To schedule the fixed time between two preventive maintenance actions, the Weibull distribution is commonly used ([Baptista, Sankararaman, de Medeiros, Nascimento, Prendinger, &](#)

Henriques, 2018). The drawbacks of preventive maintenance are related to the replacement of components that still have a remaining useful life; therefore, early interventions imply a waste of resources and too late actions could imply catastrophic failures. Additionally, the preventive intervention itself could be a source of failures too. Finally, predictive maintenance (PdM) complements the previous ones and, essentially, uses predictive tools to determine when actions are necessary (Carvalho, Soares, Vita, Francisco, Basto, & Alcalá, 2019). Within this predictive maintenance group, other terms are usually found in the literature as Condition-Based Maintenance and Prognostic and Health Management, (Montero Jimenez et al., 2020).

The role of forecasting in industrial maintenance is of paramount importance. One application is to forecast spare parts (see Section 3.2.7), whose demands are typically intermittent, usually required to carry out corrective and preventive maintenances (Van der Auweraer and Boute, 2019, Wang and Syntetos, 2011). On the other hand, it is crucial for PdM the forecast of the remaining useful time, which is the useful life left on an asset at a particular time of operation (Si, Wang, Hu, & Zhou, 2011). This work will be focused on the latter, which is usually found under the prognostic stage (Jardine, Lin, & Banjevic, 2006).

The typology of forecasting techniques employed is very ample. Montero Jimenez et al. (2020) classify them in three groups: physics-based models, knowledge-based models, and data-driven models. Physics-based models require high skills on the underlying physics of the application. Knowledge-based models are based on facts or cases collected over the years of operation and maintenance. Although, they are useful for diagnostics and provide explicative results, its performance on prognostics is more limited. In this sense, data-driven models are gaining popularity for the development of computational power, data acquisition, and big data platforms. In this case, data coming from vibration analysis, lubricant analysis, thermography, ultrasound, etc. are usually employed. Here, well-known forecasting models as VARIMAX/GARCH (see also Section 2.3) are successfully used (Baptista et al., 2018, Cheng et al., 2012, García et al., 2010, Gomez Munoz et al., 2014). State Space models based on the Kalman Filter are also employed (Pedregal and Carmen Carnero, 2006, Pedregal et al., 2009 and Section 2.3.6). Recently, given the irruption of the Industry 4.0, physical and digital systems are getting more integrated and Machine Learning/Artificial Intelligence are drawing the attention of practitioners and academics alike (Carvalho et al., 2019). In

that same reference, it is found that the most frequently used Machine Learning methods in PdM applications were Random Forest, Artificial Neural Networks, Support Vector Machines and K-means.

3.2.9. Reverse logistics¹⁰⁰

As logistics and supply chain operations rely upon accurate demand forecasts (see also Section 3.2.2), reverse logistics and closed loop supply chain operations rely upon accurate forecasts of returned items. Such items (usually referred as cores) can be anything from reusable shipping or product containers to used laptops, mobile phones or car engines. If some (re)manufacturing activity is involved in supply chains, it is both demand and returned items forecasts that are needed since it is net demand requirements (demand – returns) that drive remanufacturing operations.

Forecasting methods that are known to work well when applied to demand forecasting, such as SES for example (see Section 2.3.1), do not perform well when applied to time-series of returns because they assume returns to be a process independent of sales.

There are some cases when this independence might hold, such as when a recycler receives items sold by various companies and supply chains (Goltsos & Syntetos, 2020). In these cases, simple methods like SES applied on the time series of returns might prove sufficient. Typically though, returns are strongly correlated with past sales and the installed base (number of products with customers). After all, there cannot be a product return if a product has not first been sold. This lagged relationship between sales and returns is key to the effective characterisation of the returns process.

Despite the increasing importance of circular economy and research on closed loop supply chains, returns forecasting has not received sufficient attention in the academic literature (notable contributions in this area include Clottey et al., 2012, de Brito and van der Laan, 2009, Goh and Varaprasad, 1986, Toktay, 2003, Toktay et al., 2000). The seminal work by Kelle and Silver (1989) offers a useful framework to forecasting that is based on the degree of available information about the relationship between demand and returns. Product level (PL) information consists of the time series of sales and returns, alongside information on the time each product spends with a customer. The question then is how to derive this time to return distribution. This can be done through managerial experience, by investigating the correlation of the demand and the returns time series, or by serialising and tracking a subset (sample) of items. Past sales can then be used in conjunction with this distribution to create forecasts of returns. Serial

number level (SL) information, is more detailed and consists of the time matching of an individual unit item's issues and returns and thus exactly the time each individual unit, on a serial number basis, spent with the customer. Serialisation allows for a complete characterisation of the time to return distribution. Very importantly, it also enables tracking exactly how many items previously sold remain with customers, providing time series of unreturned past sales. Unreturned past sales can then be extrapolated — along with a time to return distribution — to create forecasts of returns.

[Goltsos, Syntetos, and van der Laan \(2019\)](#) offered empirical evidence in the area of returns forecasting by analysing a serialised data set from a remanufacturing company in North Wales. They found the Beta probability distribution to best fit times-to-return. Their research suggests that serialisation is something worthwhile pursuing for low volume products, especially if they are expensive. This makes a lot of sense from an investment perspective, since the relevant serial numbers are very few. However, they also provided evidence that such benefits expand in the case of high volume items. Importantly, the benefits of serialisation not only enable the implementation of the more complex SL method, but also the accurate characterisation of the returns process, thus also benefiting the PL method (which has been shown to be very robust).

3.3. Economics and finance

3.3.1. Macroeconomic survey expectations¹⁰¹

Macroeconomic survey expectations allow tests of theories of how agents form their expectations. Expectations play a central role in modern macroeconomic research ([Gali, 2008](#)). Survey expectations have been used to test theories of expectations formation for the last 50 years. Initially the Livingston survey data on inflationary expectations was used to test extrapolative or adaptive hypothesis, but the focus soon turned to testing whether expectations are formed rationally (see [Turnovsky & Wachter, 1972](#), for an early contribution). According to [Muth \(1961, p. 316\)](#), rational expectations is the hypothesis that: 'expectations, since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory.' This assumes all agents have access to all relevant information. Instead, one can test whether agents make efficient use of the information they possess. This is the notion of forecast efficiency ([Mincer & Zarnowitz, 1969](#)), and can be tested by regressing the outturns on a

constant and the forecasts of those outturns. Under forecast efficiency, the constant should be zero and the coefficient on the forecasts should be one. When the slope coefficient is not equal to one, the forecast errors will be systematically related to information available at the forecast origin, namely, the forecasts, and cannot be optimal. The exchange between [Figlewski and Wachtel, 1981](#), [Figlewski and Wachtel, 1983](#) and [Dietrich and Joines \(1983\)](#) clarifies the role of partial information in testing forecast efficiency (that is, full information is not necessary), and shows that the use of the aggregate or consensus forecast in the individual realisation-forecast regression outlined above will give rise to a slope parameter less than one when forecasters are efficient but possess partial information. [Zarnowitz \(1985\)](#), [Keane and Runkle \(1990\)](#) and [Bonham and Cohen \(2001\)](#) consider pooling across individuals in the realisation-forecast regression, and the role of correlated shocks across individuals. Recently, researchers considered why forecasters might not possess full-information, stressing informational rigidities: sticky information (see, *inter alia*, [Mankiw and Reis, 2002](#), [Mankiw et al., 2003](#)), and noisy information (see, *inter alia*, [Sims, 2003](#), [Woodford, 2002](#)). [Coibion and Gorodnichenko, 2012](#), [Coibion and Gorodnichenko, 2015](#) test these models using aggregate quantities, such as mean errors and revisions.

Forecaster behaviour can be characterised by the response to new information (see also Section 2.11.1). Over or under-reaction would constitute inefficiency. [Broer and Kohlhas \(2018\)](#) and [Bordalo, Gennaioli, Ma, and Shleifer \(2018\)](#) find that agents over-react, generating a negative correlation between their forecast revision and error. The forecast is revised by more than is warranted by the new information (over-confidence regarding the value of the new information). [Bordalo et al. \(2018\)](#) explain the over-reaction with a model of ‘diagnostic’ expectations, whereas [Fuhrer \(2018\)](#) finds ‘intrinsic inflation persistence’: individuals under-react to new information, smoothing their responses to news.

The empirical evidence is often equivocal, and might reflect: the vintage of data assumed for the outturns; whether allowance is made for ‘instabilities’ such as alternating over- and under-prediction ([Rossi & Sekhposyan, 2016](#)) and the assumption of squared-error loss (see, for example, [Clements, 2014b](#), [Patton and Timmermann, 2007](#)).

Research has also focused on the histogram forecasts produced by a number of macro-surveys. Density forecast evaluation techniques such as the probability integral transform¹⁰² have been applied to histogram forecasts, and survey histograms have been compared to benchmark forecasts (see, for example, [Bao et al., 2007](#), [Clements, 2018](#), [Hall and Mitchell, 2009](#)). Research has also considered uncertainty measures based on the histograms ([Clements, 2014a](#)). Sections [2.12.4 Evaluating probabilistic forecasts](#), [2.12.5 Assessing the reliability of probabilistic forecasts](#) also discuss the evaluation and reliability of probabilistic forecasts.

[Clements, 2009](#), [Clements, 2010](#) and [Engelberg, Manski, and Williams \(2009\)](#) considered the consistency between the point predictions and histogram forecasts. Reporting practices such as ‘rounding’ have also been considered ([Binder, 2017](#), [Clements, 2011](#), [Manski and Molinari, 2010](#)).

[Clements \(2019\)](#) reviews macroeconomic survey expectations.

3.3.2. Forecasting GDP and inflation¹⁰³

As soon as Bayesian estimation of DSGEs became popular, these models have been employed in forecasting horseraces to predict the key macro variables, for example, Gross Domestic Product (GDP) and inflation, as discussed in [Del Negro and Schorfheide \(2013\)](#). The forecasting performance is evaluated using rolling or recursive (expanded) prediction windows (for a discussion, see [Cardani, Paccagnini, & Villa, 2015](#)). DSGEs are usually estimated using revised data, but several studies propose better results estimating the models using real-time data (see, for example, [Cardani et al., 2019](#), [Del Negro and Schorfheide, 2013](#), [Kolasa and Rubaszek, 2015b](#), [Wolters, 2015](#)).

The current DSGE model forecasting compares DSGE models to competitors (see Section [2.3.15](#) for an introduction to DSGE models). Among them, we can include the Random Walk (the naive model which assumes a stochastic trend), the Bayesian VAR models (Minnesota Prior à la [Doan, Litterman, & Sims, 1984](#); and Large Bayesian VAR à la [Bańbura, Giannone, & Reichlin, 2010](#)), the Hybrid-Models (the DSGE-VAR à la [Del Negro & Schorfheide, 2004](#); and the DSGE-Factor Augmented VAR à la [Consolo, Favero, & Paccagnini, 2009](#)), and the institutional forecasts (Greenbook, Survey Professional Forecasts, and the Blue Chip, as illustrated in [Edge & Gürkaynak, 2010](#)).

[Table 2](#) summarises the current DSGE forecasting literature mainly for the US and Euro Area and provided by estimating medium-scale models. As general findings, DSGEs can outperform other competitors, with the exception for the Hybrid-Models, in the medium

and long-run to forecast GDP and inflation. In particular, [Smets and Wouters \(2007\)](#) was the first empirical evidence of how DSGEs can be competitive with forecasts from Bayesian VARs, convincing researchers and policymakers in adopting DSGEs for prediction evaluations. As discussed in [Del Negro and Schorfheide \(2013\)](#), the accuracy of DSGE forecasts depends on how the model is able to capture low-frequency trends in the data. To explain the macro-finance linkages during the Great Recession, the Smets and Wouters model was also compared to other DSGE specifications including the financial sector. For example, [Cardani et al., 2019](#), [Del Negro and Schorfheide, 2013](#), [Galvão et al., 2016](#), and [Kolasa and Rubaszek \(2015a\)](#) provide forecasting performance for DSGEs with financial frictions. This strand of the literature shows how this feature can improve the baseline Smets and Wouters predictions for the business cycle, in particular during the recent Great Recession.

However, the Hybrid-Models always outperform the DSGEs thanks to the combination of the theory-based model (DSGE) and the statistical representation (VAR or Factor Augmented VAR), as illustrated by [Del Negro and Schorfheide \(2004\)](#) and [Consolo et al. \(2009\)](#).

Moreover, several studies discuss how prediction performance could depend on the parameters' estimation. [Kolasa and Rubaszek \(2015b\)](#) suggest that updating DSGE model parameters only once a year is enough to have accurate and efficient predictions about the main macro variables.

Table 2. Alternative competitors to DSGE models.

Competitor	Reference
Hybrid Models	US: Del Negro and Schorfheide (2004) , Consolo et al. (2009)
Random Walk	US: Gürkaynak, Kısacıkoglu, and Rossi (2013) , Euro Area: Warne et al. (2010) , Smets, Warne, and Wouters (2014)
Bayesian VAR	US: Smets and Wouters (2007) , Gürkaynak et al. (2013) , Wolters (2015) , Bekiros and Paccagnini (2014) , Bekiros and Paccagnini (2015a) , Bekiros and Paccagnini (2015b) , Euro Area: (Warne et al., 2010)
Time-Varying VAR and Markov-Switching	US: Bekiros, Cardani, Paccagnini, and Villa (2016) , Euro Area: Bekiros and Paccagnini (2016)
Institutional Forecasts	US: Edge and Gürkaynak (2010) , Kolasa et al. (2012) , Del Negro and Schorfheide (2013) , Wolters (2015)

3.3.3. Forecasting unemployment¹⁰⁴

Unemployment has significant implications at both the micro and macro levels, influencing individual living standards, health and well-being, as well as imposing direct costs on the economy. Given its importance, policy-makers put unemployment at the heart of their economic plans, and as such require accurate forecasts to feed into economic policy decisions. Unemployment is described as a lagging indicator of the economy, with characteristics including business cycles and persistence. Despite this, forecasting the unemployment rate is difficult, because the data are highly non-stationary with abrupt distributional shifts, but persistence within regimes. In this section we focus on methods used to forecast the aggregate unemployment rate.

Unemployment is the outcome of supply and demand for labour, aggregated across all prospective workers, with labour demand derived from demand for goods and services. This implies a highly complex data generating process. Empirical forecasting models tend to simplify this relationship, with two approaches dominating the literature. The first is based on the [Phillips \(1958\)](#) curve capturing a non-linear relationship between nominal wage inflation and the unemployment rate, or the relation between unemployment and output described as Okun's [1962 Law](#). The second uses the time-series properties of the data to produce statistical forecasts, such as univariate linear models (for example, ARIMA or unobserved component models; see Sections [2.3.4 Autoregressive integrated moving average \(ARIMA\) models](#), [2.3.6 State-space models](#)), multivariate linear models (for example, VARMA or CVAR; see Section [2.3.9](#)), various threshold autoregressive models (see Section [2.3.13](#)), Markov Switching models (see Section [2.3.12](#)) and Artificial Neural Networks (see Section [2.7.8](#)).

The empirical literature is inconclusive as to the 'best' forecasting models for unemployment, which varies by country, time period and forecast horizon. There is some evidence that non-linear statistical models tend to outperform within business cycle contractions or expansions, but perform worse across business cycles (see, for example, [Koop and Potter, 1999](#), [Montgomery et al., 1998](#), [Rothman, 1998](#)), whereas [Proietti \(2003\)](#) finds that linear models characterised by higher persistence perform significantly better. Evidence of non-linearities is found by [Johnes, 1999](#), [Milas and Rothman, 2008](#) and [Peel and Speight \(2000\)](#), and [Gil-Alana \(2001\)](#) finds evidence of long-memory. [Barnichon and Garda \(2016\)](#) applies a flow approach to unemployment forecasting and finds improvements, as does [Smith \(2011\)](#).

One approach that does yield accurate forecasts is to use a measure of profitability as the explanatory variable, assuming that unemployment will fall when hiring is profitable. [Hendry \(2001\)](#) proxies profitability () by the gap between the real interest rate (reflecting costs) and the real growth rate (reflecting the demand side), such that the unemployment rate rises when the real interest rate exceeds the real growth rate, and vice versa: where r is the long-term interest rate, π is a measure of inflation and y is a measure of output growth. This is then embedded within a dynamic equilibrium correction model, using impulse indicator saturation (IIS: [Hendry et al., 2008b](#), [Johansen and Nielsen, 2009](#)) and step indicator saturation (SIS: [Castle et al., 2015c](#)) to capture outliers, breaks and regime shifts, as well as allowing for any non-linearities using Taylor expansions for the regressors. The resulting forecasts perform well over the business cycle relative to alternative statistical models (also see [Hendry, 2015](#), and [Castle, Hendry, & Martinez, 2020c](#)).

Forecasts from models of unemployment could be improved with either better economic theories of aggregate unemployment,¹⁰⁵ or more general empirical models that tackle stochastic trends, breaks, dynamics, non-linearities and interdependence,¹⁰⁶ or better still, both. The COVID-19 pandemic and subsequent lockdown policies highlight just how important forecasts of unemployment are ([Castle, Doornik, & Hendry, 2021](#)).

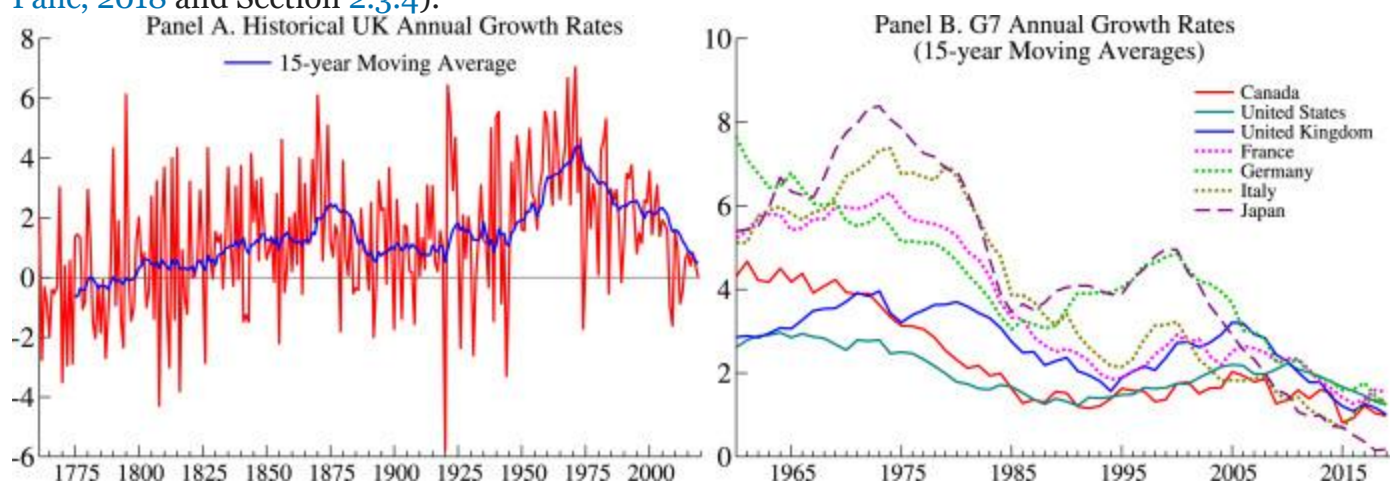
3.3.4. Forecasting productivity¹⁰⁷

The growth of labour productivity, measured by the percent change in output per hours worked, has varied dramatically over the last 260 years. In the UK it ranged from -5.8% at the onset of the 1920 Depression to just over 7% in 1971; see panel A in [Fig. 6](#).

Productivity growth is very volatile and has undergone large historical shifts with productivity growth averaging around 1% between 1800–1950 followed by an increase in the average annual growth to 3% between 1950–1975. Since the mid-1970's productivity growth has gradually declined in many developed economies; see panel B of [Fig. 6](#). In the decade since 2009, 2% annual productivity growth was an upper bound for most G7 countries.

The most common approach for forecasting productivity is to estimate the trend growth in productivity using aggregate data. For example, [Gordon \(2003\)](#) considers three separate approaches for calculating trend labor productivity in the United States based on (i) average historical growth rates outside of the business cycle, (ii) filtering the data using the HP filter ([Hodrick & Prescott, 1997](#)), and (iii) filtering the data using the

Kalman filter (see [Kalman, 1960](#)). The Office for Budget Responsibility (OBR) in the UK and the Congressional Budget Office (CBO) in the US follow similar approaches for generating its forecasts of productivity based on average historical growth rates as well as judgments about factors that may cause productivity to deviate from its historical trend in the short-term.¹⁰⁸ Alternative approaches include forecasting aggregate productivity using disaggregated firm-level data (see [Bartelsman et al., 2011](#), [Bartelsman and Wolf, 2014](#) and Section 2.10.1) and using time-series models (see [Žmuk, Dumičić, & Palić, 2018](#) and Section 2.3.4).



1. [Download: Download high-res image \(682KB\)](#)
2. [Download: Download full-size image](#)

Fig. 6. Productivity Growth (Output per total hours worked).

Bank of England and Penn World Table Version 10.0.

In the last few decades there have been several attempts to test for time-varying trends in productivity and to allow for them. However, the focus of these approaches has been primarily on the United States ([Hansen, 2001](#), [Roberts, 2001](#)), which saw a sharp rise in productivity growth in the 1990's that was not mirrored in other countries ([Basu, Fernald, Oulton, & Srinivasan, 2003](#)). Test for shifts in productivity growth rates in other advanced economies did not find evidence of a changes in productivity growth until well after the financial crisis in 2007 ([Benati, 2007](#), [Glocker and Wegmüller, 2018](#), [Turner and Boulhol, 2011](#)).

A more recent approach by [Martinez et al. \(2021\)](#) allows for a time-varying long-run trend in UK productivity. They show that are able to broadly replicate the OBR's forecasts using a quasi-transformed autoregressive model with one lag, a constant, and a trend. The estimated long-run trend is just over 2% per year through 2007 Q4 which is

consistent with the OBR's assumptions about the long-run growth rate of productivity (OBR, 2019). However, it is possible to dramatically improve upon OBR's forecasts in real-time by allowing for the long-term trend forecast to adjust based on more recent historical patterns. By taking a local average of the last four years of growth rates, Martinez et al. (2021) generate productivity forecasts whose RMSE is on average more than 75% smaller than OBR's forecasts extending five-years-ahead and is 84% smaller at the longest forecast horizon.

3.3.5. Fiscal forecasting for government budget surveillance¹⁰⁹

Recent economic recessions have led to a renewed interest in fiscal forecasting, mainly for deficit and debt surveillance. This was certainly true in the case of the 2008 recession, and looks to become even more important in the current economic crisis brought on by the COVID-19 pandemic. This is particularly important in Europe, where countries are subject to strong fiscal monitoring mechanisms. Two main themes can be detected in the fiscal forecasting literature (Leal, Pérez, Tujula, & Vidal, 2008). First, investigate the properties of forecasts in terms of bias, efficiency and accuracy. Second, check the adequacy of forecasting procedures.

The first topic has its own interest for long, mainly restricted to international institutions (Artis & Marcellino, 2001). Part of the literature, however, argue that fiscal forecasts are politically biased, mainly because there is usually no clear distinction between political targets and rigorous forecasts (Frankel and Schreger, 2013, Strauch et al., 2004). In this sense, the availability of forecasts from independent sources is of great value (Jonung & Larch, 2006). But it is not as easy as saying that independent forecasters would improve forecasts due to the absence of political bias, because forecasting accuracy is compromised by complexities of data, country-specific factors, outliers, changes in the definition of fiscal variables, etc. Very often some of these issues are known by the staff of organisations in charge of making the official statistics and forecasts long before the general public, and some information never leaves such institutions. So this insider information is actually a valuable asset to improve forecasting accuracy (Leal et al., 2008).

As for the second issue, namely the accuracy of forecasting methods, the literature can be divided into two parts, one based on macroeconomic models with specific fiscal modules that allows to analyse the effects of fiscal policy on macro variables and vice versa (see Favero and Marcellino (2005) and references therein), and the other based on

pure forecasting methods and comparisons among them. This last stream of research basically resembles closely what is seen in other forecasting areas: (i) there is no single method outperforming the rest generally, (ii) judgmental forecasting is especially important due to data problems (see Section 2.11), and (iii) combination of methods tends to outperform individual ones, see Leal et al. (2008) and Section 2.6.

Part of the recent literature focused on the generation of very short-term public finance monitoring systems using models that combine annual information with intra-annual fiscal data (Pedregal & Pérez, 2010) by time aggregation techniques (see Section 2.10.2), often set up in a SS framework (see Section 2.3.6). The idea is to produce global annual end-of-year forecasts of budgetary variables based on the most frequently available fiscal indicators, so that changes throughout the year in the indicators can be used as early warnings to infer the changes in the annual forecasts and deviations from fiscal targets (Pedregal, Pérez, & Sánchez, 2014).

The level of disaggregation of the indicator variables are established according to the information available and the particular objectives. The simplest options are the accrual National Accounts annual or quarterly fiscal balances running on their cash monthly counterparts. A somewhat more complex version is the previous one with all the variables broken down into revenues and expenditures. Other disaggregation schemes have been applied, namely by region, by administrative level (regional, municipal, social security, etc.), or by items within revenue and/or expenditure (VAT, income taxes, etc. Asimakopoulos et al., 2020, Paredes et al., 2014).

Unfortunately, what is missing is a comprehensive and transparent forecasting system, independent of Member States, capable of producing consistent forecasts over time and across countries. This is certainly a challenge that no one has yet dared to take up.

3.3.6. Interest rate prediction¹¹⁰

The (spot) rate on a (riskless) bond represents the ex-ante return (yield) to maturity which equates its market price to a theoretical valuation. Modelling and predicting default-free, short-term interest rates are crucial tasks in asset pricing and risk management. Indeed, the value of interest rate-sensitive securities depends on the value of the riskless rate. Besides, the short interest rate is a fundamental ingredient in the formulation and transmission of the monetary policy (see, for example, Section 2.3.15). However, many popular models of the short rate (for instance, continuous time, diffusion models) fail to deliver accurate out-of-sample forecasts. Their

poor predictive performance may depend on the fact that the stochastic behaviour of short interest rates may be time-varying (for instance, it may depend on the business cycle and on the stance of monetary policy).

Notably, the presence of nonlinearities in the conditional mean and variance of the short-term yield influences the behaviour of the entire term structure of spot rates implicit in riskless bond prices. For instance, the level of the short-term rate directly affects the slope of the yield curve. More generally, nonlinear rate dynamics imply a nonlinear equilibrium relationship between short and long-term yields. Accordingly, recent research has reported that dynamic econometric models with regime shifts in parameters, such as Markov switching (MS; see Section 2.3.12) and threshold models (see Section 2.3.13), are useful at forecasting rates.

The usefulness of MS VAR models with term structure data had been established since [Hamilton \(1988\)](#) and [Garcia and Perron \(1996\)](#): single-state, VARMA models are overwhelmingly rejected in favour of multi-state models. Subsequently, a literature has emerged that has documented that MS models are required to successfully forecast the yield curve. [Lanne and Saikkonen \(2003\)](#) showed that a mixture of autoregressions with two regimes improves the predictions of US T-bill rates. [Ang and Bekaert \(2002\)](#) found support for MS dynamics in the short-term rates for the US, the UK, and Germany. [Cai \(1994\)](#) developed a MS ARCH model to examine volatility persistence, reflecting a concern that it may be inflated by regimes. [Gray \(1996\)](#) generalised this attempt to MS GARCH and reported improvements in pseudo out-of-sample predictions. Further advances in the methods and applications of MS GARCH are in [Haas, Mittnik, and Paoletta \(2004\)](#) and [Smith \(2002\)](#). A number of papers have also investigated the presence of regimes in the typical factors (level, slope, and convexity) that characterise the no-arbitrage dynamics of the term structure, showing the predictive benefits of incorporating MS (see, for example, [Guidolin and Pedio, 2019](#), [Hevia et al., 2015](#)). Alternatively, a few studies have tried to capture the time-varying, nonlinear dynamics of interest rates using threshold models. As discussed by [Pai and Pedersen \(1999\)](#), threshold models have an advantage compared to MS ones: the regimes are not determined by an unobserved latent variable, thus fostering interpretability. In most of the applications to interest rates, the regimes are determined by the lagged level of the short rate itself, in a self-exciting fashion. For instance, [Pfann, Schotman, and Tschernig \(1996\)](#) explored nonlinear dynamics of the US short-term interest rate using a (self-

exciting) threshold autoregressive model augmented by conditional heteroskedasticity (namely, a TAR-GARCH model) and found strong evidence of the presence of two regimes. More recently, also [Gospodinov \(2005\)](#) used a TAR-GARCH to predict the short-term rate and showed that this model can capture some well-documented features of the data, such as high persistence and conditional heteroskedasticity.

Another advantage of nonlinear models is that they can reproduce the empirical puzzles that plague the expectations hypothesis of interest rates (EH), according to which it is a weighted average of short-term rates to drive longer-term rates (see, for example, [Bansal et al., 2004](#), [Dai et al., 2007](#)). For instance, while [Bekaert, Hodrick, and Marshall \(2001\)](#) show single-state VARs cannot generate distributions consistent with the EH, [Guidolin and Timmermann \(2009\)](#) find that the optimal combinations of lagged short and forward rates depend on regimes so that the EH holds only in some states. As widely documented (see, for instance, [Guidolin & Thornton, 2018](#)), the predictable component in mean rates is hardly significant. As a result, the random walk remains a hard benchmark to outperform as far as the prediction of the mean is concerned. However, density forecasts reflect all moments and the models that capture the dynamics of higher-order moments tend to perform best. MS models appear at the forefront of a class of non-linear models that produce accurate density predictions (see, for example, [Hong et al., 2004](#), [Maheu and Yang, 2016](#)). Alternatively, [Pfann et al. \(1996\)](#) and more recently [Dellaportas, Denison, and Holmes \(2007\)](#) estimated TAR models to also forecast conditional higher order moments and all report reasonable accuracy.

Finally, a literature has strived to fit rates not only under the physical measure, i.e., in time series, but to predict rates when MS enters the pricing kernel, the fundamental pricing operator. A few papers have assumed that regimes represent a new risk factor (see, for instance, [Dai & Singleton, 2003](#)). This literature reports that MS models lead to a range of shapes for nominal and real term structures (see, for instance, [Veronesi & Yared, 1999](#)). Often the model specifications that are not rejected by formal tests include regimes ([Ang et al., 2008](#), [Bansal and Zhou, 2002](#)).

To conclude, it is worthwhile noting that, while threshold models are more interpretable, MS remain a more popular alternative for the prediction of interest rates. This is mainly due to the fact that statistical inference for threshold regime switching

models poses some challenges, because the likelihood function is discontinuous with respect to the threshold parameters.

3.3.7. House price forecasting¹¹¹

The boom and bust in housing markets in the early and mid 2000s and its decisive role in the Great Recession has generated a vast interest in the dynamics of house prices and emphasised the importance of accurately forecasting property price movements during turbulent times. International organisations, central banks and research institutes have become increasingly engaged in monitoring the property price developments across the world.¹¹² At the same time, a substantial empirical literature has developed that deals with predicting future house price movements (for a comprehensive survey see [Ghysels, Plazzi, Valkanov, & Torous, 2013](#)). Although this literature concentrates almost entirely on the US (see, for example, [Bork and Møller, 2015](#), [Rapach and Strauss, 2009](#)), there are many other countries, such as the UK, where house price forecastability is of prime importance. Similarly to the US, in the UK, housing activities account for a large fraction of GDP and of households' expenditures; real estate property comprises a significant component of private wealth and mortgage debt constitutes a main liability of households ([Office for National Statistics, 2019](#)).

The appropriate forecasting model has to reflect the dynamics of the specific real estate market and take into account its particular characteristics. In the UK, for instance, there is a substantial empirical literature that documents the existence of strong spatial linkages between regional markets, whereby the house price shocks emanating from southern regions of the country, and in particular Greater London, have a tendency to spread out and affect neighbouring regions with a time lag (see, for example, [Antonakakis et al., 2018](#), [Cook and Thomas, 2003](#), [Holly et al., 2010](#) *inter alia*); see also Section [2.3.10](#) on forecasting functional data.

Recent evidence also suggests that the relationship between real estate valuations and conditioning macro and financial variables displayed a complex of time-varying patterns over the previous decades ([Aizenman & Jinjarak, 2013](#)). Hence, predictive methods that do not allow for time-variation in both predictors and their marginal effects may not be able to capture the complex house price dynamics in the UK (see [Yusupova, Pavlidis, & Pavlidis, 2019](#) for a comparison of forecasting accuracy of a battery of static and dynamic econometric methods).

An important recent trend is to attempt to incorporate information from novel data sources (such as newspaper articles, social media, etc.) in forecasting models as a measure of expectations and perceptions of economic agents (see also Section 2.9.3). It has been shown that changes in uncertainty about house prices impact on housing investment and real estate construction decisions (Banks et al., 2015, Cunningham, 2006, Oh and Yoon, 2020), and thus incorporating a measure of uncertainty in the forecasting model can improve the forecastability of real estate prices. For instance in the UK, the House Price Uncertainty (HPU) index (Yusupova, Pavlidis, Paya, & Peel, 2020), constructed using the methodology outlined in Baker, Bloom, and Davis (2016),¹¹³ was found to be important in predicting property price inflation ahead of the house price collapse of the third quarter of 2008 and during the bust phase (Yusupova et al., 2019). Along with capturing the two recent recessions (in the early 1990s and middle 2000s) this index also reflects the uncertainty related to the EU Referendum, Brexit negotiations and COVID-19 pandemic.

3.3.8. Exchange rate forecasting¹¹⁴

Exchange rates have long fascinated and puzzled researchers in international finance. The reason is that following the seminal paper of Meese and Rogoff (1983), the common wisdom is that macroeconomic models cannot outperform the random walk in exchange rate forecasting (see Rossi, 2013 for a survey). This view is difficult to reconcile with the strong belief that exchange rates are driven by fundamentals, such as relative productivity, external imbalances, terms of trade, fiscal policy or interest rate disparity (Couharde et al., 2018, Lee et al., 2013, MacDonald, 1998). These two contradicting assertions by the academic literature is referred to as “exchange rate disconnect puzzle”.

The literature provides several explanations for this puzzle. First, it can be related to the forecast estimation error (see Section 2.5.2). The studies in which models are estimated with a large panels of data (Engel et al., 2008, Ince, 2014, Mark and Sul, 2001), long time series (Lothian & Taylor, 1996) or calibrated (Ca’ Zorzi & Rubaszek, 2020) deliver positive results on exchange rate forecastability. Second, there is ample evidence that the adjustment of exchange rates to equilibrium is non-linear (Curran and Velic, 2019, Taylor and Peel, 2000), which might diminish the out-of-sample performance of macroeconomic models (Kilian and Taylor, 2003, Lopez-Suarez and Rodriguez-Lopez, 2011). Third, few economists argue that the role of macroeconomic fundamentals may

be varying over time and this should be accounted for in a forecasting setting (Beckmann and Schussler, 2016, Byrne et al., 2016).

The dominant part of the exchange rate forecasting literature investigates which macroeconomic model performs best out-of-sample. The initial studies explored the role of monetary fundamentals to find that these models deliver inaccurate short-term and not so bad long-term predictions in comparison to the random walk (Mark, 1995, Meese and Rogoff, 1983). In a comprehensive study from mid-2000s, Cheung, Chinn, and Pascual (2005) showed that neither monetary, uncovered interest parity (UIP) nor behavioural equilibrium exchange rate (BEER) model are able to outperform the no-change forecast. A step forward was made by Molodtsova and Papell (2009), who proposed a model combining the UIP and Taylor rule equations and showed that it delivers competitive exchange rate forecasts. This result, however, has not been confirmed by more recent studies (Cheung et al., 2019, Engel et al., 2019). In turn, Ca' Zorzi and Rubaszek (2020) argue that a simple method assuming gradual adjustment of the exchange rate towards the level implied by the Purchasing Power Parity (PPP) performs well over shorter as well as longer horizon. This result is consistent with the results of Ca' Zorzi et al. (2017) and Eichenbaum, Johansson, and Rebelo (2017), who showed that exchange rates are predictable within a general equilibrium DSGE framework (see Section 2.3.15), which encompasses an adjustment of the exchange rate to a PPP equilibrium. Finally, Ca' Zorzi, Cap, Mijakovic, and Rubaszek (2020) discuss how extending the PPP framework for other fundamentals within the BEER framework is not helping in exchange rate forecasting. Overall, at the current juncture it might be claimed that “exchange rate disconnect puzzle” is still puzzling, with some evidence that methods based on PPP and controlling the estimation forecast error can deliver more accurate forecast than the random walk benchmark. A way forward to account for macroeconomic variables in exchange rate forecasting could be to use variable selection methods that allow to control for the estimation error (see Section 2.5.3).

3.3.9. Financial time series forecasting with range-based volatility models¹¹⁵

The range-based (RB) volatility models is a general term for the models constructed with high and low prices, and most often with their difference i.e., the price range. A

short review and classification of such models is contained in Section 2.3.14. From practical point of view, it is important that low and high prices are almost always available with daily closing prices for financial series. The price range (or its logarithm) is a significantly more efficient estimator of volatility than the estimator based on closing prices (Alizadeh et al., 2002). Similarly the co-range (the covariance based on price ranges) is a significantly more efficient estimator of the covariance of returns than the estimator based on closing prices (Brunetti & Lildholdt, 2002). For these reasons models based on the price range and the co-range better describe variances and covariances of financial returns than the ones based on closing prices.

Forecasts of volatility from simple models like moving average, EWMA, AR, ARMA based on the RB variance estimators are more accurate than the forecasts from the same models based on squared returns of closing prices (Rajvanshi, 2015, Vipul and Jacob, 2007). Forecasts of volatility from the AR model based on the Parkinson estimator are more precise even than the forecasts from the standard GARCH models (see Section 2.3.11) based on closing prices (Li & Hong, 2011).

In plenty of studies it was shown that forecasts of volatility of financial returns from the univariate RB models are more accurate than the forecasts from standard GARCH models based on closing prices (see, for example, Mapa, 2003 for the GARCH-PARK-R model; Chou, 2005 for the CARR model; Fiszeder, 2005 for the GARCH-TR model; Brandt & Jones, 2006 for the REGARCH model; Chen et al., 2008 for the TARR model; Lin et al., 2012 for the STARR model; Fiszeder & Perczak, 2016 for the GARCH model estimated with low, high and closing prices during crisis periods; Molnár, 2016 for the RGARCH model).

The use of daily low and high prices in the multivariate volatility models leads to more accurate forecasts of covariance or covariance matrix of financial returns than the forecasts from the models based on closing prices (see, for example, Chou et al., 2009 for the RB DCC model; Harris & Yilmaz, 2010 for the hybrid EWMA model; Fiszeder, 2018 for the BEKK-HL model; Fiszeder & Faldziński, 2019 for the co-range DCC model; Fiszeder et al., 2019 for the DCC-RGARCH model).

The RB models were used in many financial applications. They lead for example to more precise forecasts of value-at-risk measures in comparison to the application of only closing prices (see, for example, Chen, Gerlach, Hwang, & McAleer, 2012 for the threshold CAViaR model; Asai & Brugal, 2013 for the HVAR model; Fiszeder et

al., 2019 for the DCC-RGARCH model; Meng & Taylor, 2020 for scoring functions). The application of the multivariate RB models provides also the increase in the efficiency of hedging strategies (see, for example, Chou et al., 2009 for the RB DCC model; Harris & Yilmaz, 2010 for the hybrid EWMA model; Su & Wu, 2014 for the RB-MS-DCC model). Moreover, the RB volatility models have more significant economic value than the return-based ones in the portfolio construction (Chou & Liu, 2010 for the RB DCC model; Wu & Liang, 2011 for the RB-copula model). Some studies show that based on the forecasts from the volatility models with low and high prices it is possible to construct profitable investment strategies (He, Kwok, & Wan, 2010 for the VECM model; Kumar, 2015 for the CARRS model).

3.3.10. Copula forecasting with multivariate dependent financial times series¹¹⁶

In this section, we focus on the practical advances on jointly forecasting multivariate financial time series with copulas. In the copula framework (see Section 2.4.3), because marginal models and copula models are separable, point forecasts are straightforward with marginal models, but dependence information is ignored. A joint probabilistic forecast with copulas involves both estimations of the copula distribution and marginal models.

In financial time series, an emerging interest is to model and forecast the asymmetric dependence. A typical asymmetric dependence phenomenon is that two stock returns exhibit greater correlation during market downturns than market upturns. Patton (2006) employs the asymmetric dependence between exchange rates with a time-varying copula construction with AR and GARCH margins. A similar study for measuring financial contagion with copulas allows the parameters of the copula to change with the states of the variance to identify shifts in the dependence structure in times of crisis (Rodriguez, 2007).

In stock forecasting, Almeida and Czado (2012) employ a stochastic copula autoregressive model to model DJI and Nasdaq, and the dependence at the time is modelled by a real-valued latent variable, which corresponds to the Fisher transformation of Kendall's . Li and Kang (2018) use a covariate-dependent copula framework to forecast the time varying dependence that improves both the probabilistic forecasting performance and the forecasting interpretability. Liquidity risk is another

focus in finance. [Weiß and Supper \(2013\)](#) forecast three types of liquidity-adjusted intraday Value-at-Risk (L-IVaR) with a vine copula structure. The liquidity-adjusted intraday VaR is based on simulated portfolio values, and the results are compared with the realised portfolio profits and losses.

In macroeconomic forecasting, most existing reduced-form models for multivariate time series produce symmetric forecast densities. Gaussian copulas with skew Student's- t margins depict asymmetries in the predictive distributions of GDP growth and inflation ([Smith & Vahey, 2016](#)). Real-time macroeconomic variables are forecasted with heteroscedastic inversion copulas ([Smith & Maneesoonthorn, 2018](#)) that allow for asymmetry in the density forecasts, and both serial and cross-sectional dependence could be captured by the copula function ([Loaiza-Maya & Smith, 2020](#)).

Copulas are also widely used to detect and forecast default correlation, which is a random variable called *time-until-default* to denote the survival time of each defaultable entity or financial instrument ([Li, 2000](#)). Then copulas are used in modelling the dependent defaults ([Li, 2000](#)), forecasting credit risk ([Bielecki & Rutkowski, 2013](#)), and credit derivatives market forecasting ([Schönbucher, 2003](#)). A much large volume of literature is available for this specific area. See the aforementioned references therein. For particular applications in default risk and credit default swap (CDS) forecasting, see [Li and He \(2019\)](#) and [Oh and Patton \(2018\)](#) respectively.

In energy economics, [Aloui, Hammoudeh, and Nguyen \(2013\)](#) employ the time-varying copula approach, where the marginal models are from ARMA(,)–GARCH(1,1) to investigate the conditional dependence between the Brent crude oil price and stock markets in the Central and Eastern European transition economies. [Bessa, Miranda, Botterud, Zhou, and Wang \(2012\)](#) propose a time-adaptive quantile-copula where the copula density is estimated with a kernel density forecast method. The method is applied to wind power probabilistic forecasting (see also Section 3.4.6) and shows its advantages for both system operators and wind power producers. Vine copula models are also used to forecast wind power farms' uncertainty in power system operation scheduling. [Wang, Wang, Liu, Wang, and Hou \(2017\)](#) shows vine copulas have advantages of providing reliable and sharp forecast intervals, especially in the case with limited observations available.

3.3.11. Financial forecasting with neural networks¹⁷

Neural Networks (NNs; see Section 2.7.8) are capable of successfully modelling non-stationary and non-linear series. This property has made them one of the most popular (if not the most) non-linear specification used by practitioners and academics in Finance. For example, 89% of European banks use NNs to their operations ([European Banking Federation, 2019](#)) while 25.4% of the NNs applications in total is in Finance ([Wong, Bodnovich, & Selvi, 1995](#)).

The first applications of NNs in Finance and currently the most widespread, is in financial trading. In the mid-80s when computational power became cheaper and more accessible, hedge fund managers started to experiment with NNs in trading. Their initial success led to even more practitioners to apply NNs and nowadays 67% of hedge fund managers use NNs to generate trading ideas ([BarclayHedge, 2018](#)). A broad measure of the success of NNs in financial trading is provided by the EurekaHedge AI Hedge Fund Index¹¹⁸ where it is noteworthy the 13.02% annualised return of the selected AI hedge funds over the last 10 years.

In academia, financial trading with NNs is the focus of numerous papers. Notable applications of NNs in trading financial series were provided by [Dunis et al., 2010](#), [Kaastra and Boyd, 1996](#), [Panda and Narasimhan, 2007](#), [Tenti, 1996](#), and [Zhang and Ming \(2008\)](#). The aim of these studies is to forecast the sign or the return of financial trading series and based on these forecasts to generate profitable trading strategies. These studies are closely related to the ones presented in Section 3.3.13 but the focus is now in profitability. The second major field of applications of NNs in Finance is in derivatives pricing and financial risk management. The growth of the financial industry and the provided financial services have made NNs and other machine learning algorithms a necessity for tasks such as fraud detection, information extraction and credit risk assessment ([Buchanan, 2019](#)). In derivatives pricing, NNs try to fill the limitations of the Black–Scholes model and are being used in options pricing and hedging. In academia notable applications of NNs in risk management are provided by [Liu \(2005\)](#) and [Locarek-Junge and Prinzler \(1998\)](#) and in derivatives by [Bennell and Sutcliffe \(2004\)](#) and [Psaradellis and Sermpinis \(2016\)](#).

As discussed before, financial series due to their non-linear nature and their wide applications in practice seems the perfect forecasting data set for researchers that want to test their NN topologies. As a result, there are thousands of forecasting papers in the field of NNs in financial forecasting. However, caution is needed in interpretation of

their results. NNs are sensitive to the choice of their hyperparameters. For a simple MLP, a practitioner needs to set (among others) the number and type of inputs, the number of hidden nodes, the momentum, the learning rate, the number of epochs and the batch size. This complexity in NN modelling leads inadvertently to the data snooping bias (see also Section 2.12.6). In other words, a researcher that experiments long enough with the parameters of a NN topology can have excellent in-sample and out-of-sample results for a series. However, this does not mean that the results of his NN can be generalised. This issue has led the related literature to be stained by studies cannot be extended in different samples.

3.3.12. Forecasting returns to investment style¹¹⁹

Investment style or factor portfolios are constructed from constituent securities on the basis of a variety of a-priori observable characteristics, thought to affect future returns. For example a ‘Momentum’ portfolio might be constructed with positive (‘long’) exposures to stocks with positive trailing 12-month returns, and negative (‘short’) exposure to stocks with negative trailing 12-month returns (for full background and context, see, for example [Bernstein, 1995](#), [Haugen, 2010](#)).¹²⁰ Explanations as to why such characteristics seem to predict returns fall in to two main camps: firstly that the returns represent a risk premium, earned by the investor in return for taking on some kind of (undiversifiable) risk, and secondly that such returns are the result of behavioural biases on the part of investors. In practice, both explanations are likely to drive style returns to a greater or lesser extent. Several such strategies have generated reasonably consistent positive risk-adjusted returns over many decades, but as with many financial return series, return volatility is large relative to the mean, and there can be periods of months or even years when returns deviate significantly from their long-run averages. The idea of timing exposure to styles is therefore at least superficially attractive, although the feasibility of doing so is a matter of some debate ([Arnott et al., 2016](#), [Asness, 2016](#), [Bender et al., 2018](#)). Overconfidence in timing ability has a direct cost in terms of trading frictions and opportunity cost in terms of potential expected returns and diversification forgone.

A number of authors write on the general topic of style timing (recent examples include [Dichtl et al., 2019](#), [Hodges et al., 2017](#)), and several forecasting methodologies have been suggested, falling in to three main camps:

- 1.

Serial Correlation: Perhaps the most promising approach is exploiting serial correlation in style returns. [Babu, Levine, Ooi, Pedersen, and Stamelos \(2020\)](#) and [Tarun and Bryan \(2019\)](#) outline two such approaches and [Ehsani and Linnainmaa \(2020\)](#) explore the relationship between momentum in factor portfolios and momentum in underlying stock returns. As with valuation spreads mentioned below, there is a risk that using momentum signals to time exposure to momentum factor portfolios risks unwittingly compounding exposure. A related strand of research relates (own) factor volatility to future returns, in particular for momentum factors ([Barroso, 2015](#), [Daniel and Moskowitz, 2016](#)).

- 2. Valuation Spreads: Using value signals (aggregated from individual stock value exposures) to time exposure to various fundamental-based strategies is a popular and intuitively appealing approach ([Asness, 2016](#)); however evidence of value added from doing so is mixed, and the technique seems to compound risk exposure to value factors.
- 3. Economic & Financial Conditions: [Polk, Haghbin, and de Longis \(2020\)](#) explore how economic and financial conditions affect style returns (an idea that dates back at least to [Bernstein \(1995\)](#) and references therein).

Style returns exhibit distinctly non-normal distributions. On a univariate basis, most styles display returns which are highly negatively skewed and demonstrate significant kurtosis. The long-run low correlation between investment styles is often put forward as a benefit of style-based strategies, but more careful analysis reveals that non-normality extends to the co-movements of investment style returns; factors exhibit significant tail dependence. [Christoffersen and Langlois \(2013\)](#) explores this issue, also giving details of the skew and kurtosis of weekly style returns. These features of the data mean that focusing solely on forecasting the mean may not be sufficient, and building distributional forecasts becomes important for proper risk management. [Jondeau \(2007\)](#) writes extensively on modelling non-gaussian distributions.

3.3.13. Forecasting stock returns¹²¹

Theory and intuition suggest a plethora of potentially relevant predictors of stock returns. Financial statement data (e.g., [Chan and Genovese, 2001](#), [Yan and Zheng, 2017](#)) provide a wealth of information, and variables relating to liquidity, price trends, and

sentiment, among numerous other concepts, have been used extensively by academics and practitioners alike to predict stock returns. The era of big data further increases the data available for forecasting returns. When forecasting with large numbers of predictors, conventional ordinary least squares (OLS) estimation is highly susceptible to overfitting, which is exacerbated by the substantial noise in stock return data (reflecting the intrinsically large unpredictable component in returns); see Section 2.7.11.

Over the last decade or so, researchers have explored methods for forecasting returns with large numbers of predictors. Principal component regression extracts the first few principal components (or factors) from the set of predictors; the factors then serve as predictors in a low-dimensional predictive regression, which is estimated via OLS (see Section 2.7.1). Intuitively, the factors combine the information in the individual predictors to reduce the dimension of the regression, which helps to guard against overfitting. Ludvigson and Ng (2007) find that a few factors extracted from hundreds of macroeconomic and financial variables improve out-of-sample forecasts of the US market return. Kelly and Pruitt (2013) and Huang, Jiang, Tu and Zhou (2015) use partial least squares (Wold, 1966) to construct target-relevant factors from a cross section of valuation ratios and a variety of sentiment measures, respectively, to improve market return forecasts.

Since Bates and Granger (1969), it has been known that combinations of individual forecasts often perform better than the individual forecasts themselves (Timmermann, 2006 and Section 2.6.1). Rapach, Strauss, and Zhou (2010) show that forecast combination can significantly improve out-of-sample market return forecasts. They first construct return forecasts via individual univariate predictive regressions based on numerous popular predictors from the literature (Goyal & Welch, 2008). They then compute a simple combination forecast by taking the average of the individual forecasts. Rapach et al. (2010) demonstrate that forecast combination exerts a strong shrinkage effect, thereby helping to guard against overfitting.

An emerging literature uses machine-learning techniques to construct forecasts of stock returns based on large sets of predictors. In an investigation of lead–lag relationships among developed equity markets, Rapach, Strauss, and Zhou (2013) appear to be the first to employ machine-learning tools to predict market returns. They use the elastic net (ENet: Zou & Hastie, 2005), a generalisation of the popular least absolute shrinkage and selection operator (LASSO: Tibshirani, 1996). The LASSO and ENet employ

penalised regression to guard against overfitting in high-dimensional settings by shrinking the parameter estimates toward zero. [Chinco, Clark-Joseph, and Ye \(2019\)](#) use the LASSO to forecast high-frequency (one-minute-ahead) individual stock returns and report improvements in out-of-sample fit, while [Rapach et al. \(2019\)](#) use the LASSO to improve monthly forecasts of industry returns.

Incorporating insights from [Diebold and Shin \(2019\)](#), [Han, He, Rapach, and Zhou \(2021\)](#) use the LASSO to form combination forecasts of cross-sectional stock returns based on a large number of firm characteristics from the cross-sectional literature (e.g., [Harvey et al., 2016](#), [Hou et al., 2020](#), [McLean and Pontiff, 2016](#)), extending the conventional OLS approach of [Green, Hand, and Zhang \(2017\)](#), [Haugen and Baker \(1996\)](#), and [Lewellen \(2015\)](#). [Dong, Li, Rapach, and Zhou \(2021\)](#) and [Rapach and Zhou \(2020\)](#) use the ENet to compute combination forecasts of the market return based on popular predictors from the time-series literature and numerous anomalies from the cross-sectional literature, respectively. Forecasting individual stock returns on the basis of firm characteristics in a panel framework, [Freyberger, Neuhierl, and Weber \(2020\)](#) and [Gu, Kelly, and Xiu \(2020\)](#) employ machine-learning techniques – such as the nonparametric additive LASSO ([Huang, Horowitz, & Wei, 2010](#)), random forests ([Breiman, 2001](#)), and artificial neural networks – that allow for nonlinear predictive relationships.

3.3.14. Forecasting crashes in stock markets¹²²

Time series data on financial asset returns have special features. Returns themselves are hard to forecast, while it seems that volatility of returns can be predicted. Empirical distributions of asset returns show occasional clusters of large positive and large negative returns. Large negative returns, that is, crashes seem to occur more frequently than large positive returns. Forecasting upcoming increases or decreases in volatility can be achieved by using variants of the Autoregressive Conditional Heteroskedasticity (ARCH) model ([Bollerslev, 1986](#), [Engle, 1982](#) and Section 2.3.11) or realized volatility models ([Taylor, 1986a](#)). These models take (functions of) past volatility and past returns as volatility predictors, although also other explanatory variables can be incorporated in the regression.

An important challenge that remains is to predict crashes. [Sornette \(2003\)](#) summarises potential causes for crashes and these are computer trading, increased trading in derivatives, illiquidity, trade and budget deficits, and especially, herding behaviour of

investors. Yet, forecasting the exact timing of crashes may seem impossible, but on the other hand, it may be possible to forecast the probability that a crash may occur within a foreseeable future. Given the herding behaviour, any model to use for prediction should include some self-exciting behaviour. For that purpose, [Aït-Sahalia, Cacho-Diaz, and Laeven \(2015\)](#) propose mutually exciting jump processes, where jumps can excite new jumps, also across assets or markets (see also [Chavez-Demoulin, Davison, & McNeil, 2005](#)). Another successful approach is the Autoregressive Conditional Duration (ACD) model ([Engle and Russell, 1997](#), [Engle and Russell, 1998](#)), which refers to a time series model for durations between (negative) events.

An alternative view on returns' volatility and the potential occurrence of crashes draws upon the earthquake literature ([Ogata, 1978](#), [Ogata, 1988](#)). The idea is that tensions in and across tectonic plates build up, until an eruption, and after that, tension starts to build up again until the next eruption. By modelling the tension-building-up process using so-called Hawkes processes ([Hawkes, 1971](#), [Hawkes, 2018](#), [Hawkes and Oakes, 1974](#), [Ozaki, 1979](#)), one can exploit the similarities between earthquakes and financial crashes (see also Section 2.8.4). [Gresnigt, Kole, and Franses \(2015\)](#) take Hawkes processes to daily S&P 500 data and show that it is possible to create reliable probability predictions of a crash occurrence within the next five days. [Gresnigt et al., 2017a](#), [Gresnigt et al., 2017b](#) further develop a specification strategy for any type of asset returns, and document that there are spillovers across assets and markets.

Given investor behaviour, past crashes can ignite future crashes. Hawkes processes are particularly useful to describe this feature and can usefully be implemented to predict the probability of nearby crashes. By the way, these processes can also be useful to predict social conflicts, as also there one may discern earthquake-like patterns. [van den Hengel and Franses \(2020\)](#) document their forecasting power for social conflicts in Africa.

3.4. Energy

3.4.1. Building energy consumption forecasting and optimisation¹²³

In Europe, buildings account for 40% of total energy consumed and 36% of total emissions ([Patti, Acquaviva, Jahn, Pramudianto, Tomasi, Rabourdin, Virgone, & Macii, 2016](#)). Given that energy consumption of buildings is expected to increase in the

coming years, forecasting electricity consumption becomes critical for improving energy management and planning by supporting a large variety of optimisation procedures. The main challenge in electricity consumption forecasting is that building energy systems are complex in nature, with their behaviour depending on various factors related to the type (e.g., residential, office, entertainment, business, and industrial) and the end-uses (e.g., heating, cooling, hot water, and lighting) of the building, its construction, its occupancy, the occupants' behaviour and schedule, the efficiency of the installed equipment, and the weather conditions (Zhao & Magoulès, 2012). Special events, holidays, and calendar effects can also affect the behaviour of the systems and further complicate the consumption patterns, especially when forecasting at hourly or daily level (see Section 2.3.5). As a result, producing accurate forecasts typically requires developing tailored, building-specific methods.

To deal with this task, the literature focuses on three main classes of forecasting methods, namely engineering, statistical, and ML (Mat Daut, Hassan, Abdullah, Rahman, Abdullah, & Hussin, 2017). Engineering methods, typically utilised through software tools such as DOE-2, EnergyPlus, BLAST, and ESP-r, build on physical models that forecast consumption through detailed equations which account for the particularities of the building (Al-Homoud, 2001, Fouquier et al., 2013, Zhao and Magoulès, 2012). Statistical methods usually involve linear regression (see Section 2.3.2), ARIMA/ARIMAX (see Section 2.3.4), and exponential smoothing (see Section 2.3.1) models that forecast consumption using past consumption data or additional explanatory variables, such as weather or occupancy and calendar related information (Deb, Zhang, Yang, Lee, & Shah, 2017). Finally, ML methods (see Section 2.7.10) typically involve neural networks (see Section 2.7.8), support vector machines, and grey models that account for multiple non-linear dependencies between the electricity consumed and the factors influencing its value (Ahmad, Hassan, Abdullah, Rahman, Hussin, Abdullah, & Saidur, 2014). Till present, the literature has been inconclusive about which class of methods is the most appropriate, with the conclusions drawn being subject to the examined building type, data set used, forecasting horizon considered, and data frequency at which the forecasts are produced (Wei, Li, Peng, Zeng, & Lu, 2019). To mitigate this problem, combinations of methods (see Section 2.6) and hybrids (see Section 2.7.13) have been proposed, reporting encouraging results (Mohandes et al., 2019, Zhao and Magoulès, 2012).

Other practical issues refer to data pre-processing. Electricity consumption data is typically collected at high frequencies through smart meters and therefore display noise and missing or extreme values due to monitoring issues (see Section 2.7.11). As a result, verifying the quality of the input data through diagnostics and data cleansing techniques (see Section 2.2.3 and Section 2.2.4), as well as optimising the selected time frames, are important for improving forecasting performance (Bourdeau, qiang Zhai, Nefzaoui, Guo, & Chatellier, 2019). Similarly, it is critical to engineer (see Section 2.2.5) and select (see Section 2.5.3) appropriate regressor variables which are of high quality and possible to accurately predict to assist electricity consumption forecasting. Finally, it must be carefully decided whether the bottom-up, the top-down or a combination method (see Section 2.10.1) will be used for producing reconciled forecasts at both building and end-use level (Kuster, Rezgui, & Mourshed, 2017), being also possibly mixed with temporal aggregation approaches (Spiliotis, Petropoulos et al., 2020 but also Section 2.10.3). Provided that accurate forecasts are available, effective energy optimisation can take place at a building level or across blocks of buildings (see Section 3.4.10) to reduce energy cost, improve network stability, and support efforts towards a carbon-free future, by exploiting smart grid, internet of things (IoT), and big data technologies along with recommendation systems (Marinakis et al., 2020).

An example for a typical application in this area is the optimisation of heating, ventilation, and air conditioning (HVAC) systems. The goal is to minimise the energy use of the HVAC system under the constraints of maintaining certain comfort levels in the building (Marinakis, Doukas, Spiliotis, & Papastamatiou, 2017). Though this is predominantly an optimisation exercise, forecasting comes in at different points of the system as input into the optimisation, and many problems in this space involve forecasting as a sub-problem, including energy consumption forecasting, room occupancy forecasting, inside temperature forecasting, (hyper-local) forecasts of outside temperature, and air pressure forecasting for ventilation, among others. For instance, Krüger and Givoni (2004) use a linear regression approach to predict inside temperatures in 3 houses in Brazil, and Ruano, Crispim, Conceição, and Lúcio (2006) propose the use of a neural network to predict temperatures in a school building. Madaus, McDermott, Hacker, and Pullen (2020) predict hyper-local extreme heat events, combining global climate models and machine learning models. Jing, Cai,

Chen, Zhai, Cui, and Yin (2018) predict air pressure to tackle the air balancing problem in ventilation systems, using a support vector machine.

Predicting energy demand on a building/household level from smart meter data is an important research topic not only for energy savings. In the building space, Ahmad, Mourshed, and Rezugui (2017), Touzani, Granderson, and Fernandes (2018), and Wang, Wang, Zeng, Srinivasan, and Ahrentzen (2018) predict building energy consumption of residential and commercial buildings using decision tree-based algorithms (random forests and gradient boosted trees) and neural networks to improve energy efficiency. A recent trend in forecasting are global forecasting models, built across sets of time series (Januschowski et al., 2020). (Recurrent) neural networks (Bandara et al., 2020a, Hewamalage et al., 2021) are particularly suitable for this type of processing due to their capabilities to deal with external inputs and cold-start problems. Such capabilities are necessary if there are different regimes in the simulations under which to predict, an example of such a system for HVAC optimisation is presented by Godahewa, Deng, Prouzeau, and Bergmeir (2020).

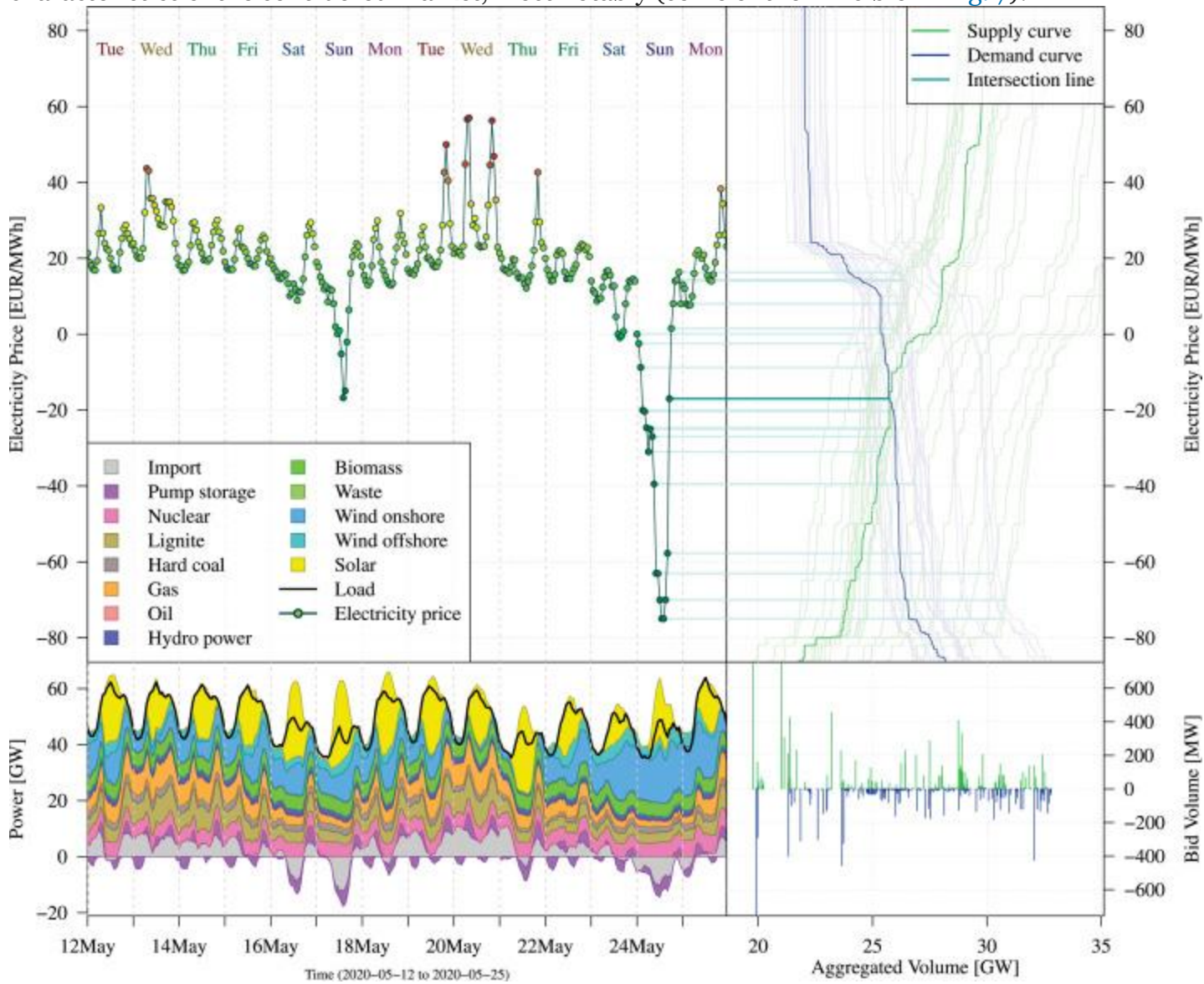
More generally, many challenges in the space of building energy optimisation are classical examples of so-called “predict then optimise” problems (Demirovic et al., 2019, Elmachtoub and Grigas, 2017). Here, different possible scenario predictions are obtained from different assumptions in the form of input parameters. These input parameters are then optimised to achieve a desired predicted outcome. As both prediction and optimisation are difficult problems, they are usually treated separately (Elmachtoub & Grigas, 2017), though there are now recent works where they are considered together (Demirovic et al., 2019, El Balghiti et al., 2019), and this will certainly be an interesting avenue for future research.

3.4.2. Electricity price forecasting¹²⁴

Forecasting electricity prices has various challenges that are highlighted in the detailed review papers by Weron (2014). Even though there are economically well motivated fundamental electricity price models, forecasting models based on evaluating historic price data are dominating the academic literature. In recent years the focus on probabilistic forecasting grew rapidly, as they are highly relevant for many applications in energy trading and risk management, storage optimisation and predictive maintenance, (Nowotarski and Weron, 2018, Ziel and Steinert, 2018). Electricity price data is highly complex and is influenced by regulation. However, there is electricity

trading based on auctions and on continuous trading. Many markets like the US and European markets organise day-ahead auctions for electricity prices, see Fig. 7. Thus, we have to predict multivariate time series type data (Ziel & Weron, 2018). In contrast, intraday markets usually apply continuous trading to manage short term variations due to changes in forecasts of renewable energy and demand, and outages (Kiesel & Paraschiv, 2017).

The key challenge in electricity price forecasting is to address all potential characteristics of the considered market, most notably (some of them visible in Fig. 7):



1. [Download: Download high-res image \(1MB\)](#)
2. [Download: Download full-size image](#)

Fig. 7. Hourly German day-ahead electricity price data resulting from a two-sided auction (top left) with corresponding 24 sale/supply and purchase/demand curves for 24 May 2020 and highlighted curves for 17:00 (top right), power generation and consumption time series (bottom left), and bid structure of 24 May 2020 17:00 (bottom right).

- 1.
(time-varying) autoregressive effects and (in)stationarity,
- 2.
calendar effects (daily, weekly and annual seasonality, holiday effects, clock-change),
- 3.
(time-varying) volatility and higher moment effects,
- 4.
price spikes (positive and negative), and
- 5.
price clustering.

Some of those impacts can be explained by external inputs, that partially have to be predicted in advance:

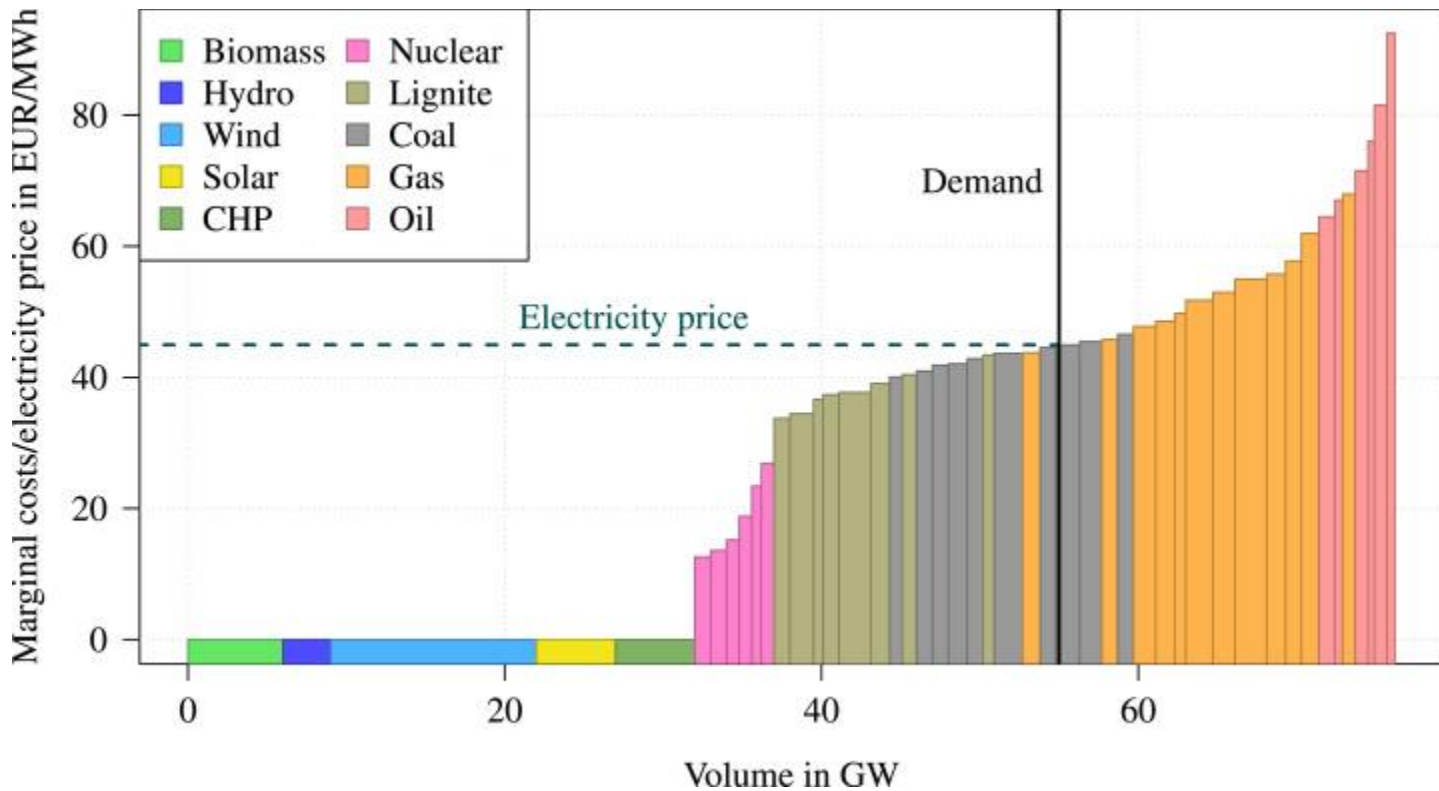
- 1.
load/demand/consumption (see Section [3.4.3](#)),
- 2.
power generation, especially from the renewable energy sources (RES) of wind and solar (see Sections [3.4.6 Wind power forecasting](#), [3.4.8 Solar power forecasting](#)),
- 3.
relevant fuel prices (especially oil, coal, natural gas; see also Section [3.4.4](#)),
- 4.
prices of emission allowances (costs),
- 5.
related power market prices (future, balancing and neighboring markets),
- 6.
availabilities of power plants and interconnectors,
- 7.

- import/export flow related data, and
- 8. weather effects (e.g. temperature due to cooling and heating and combined heat and power (CHP) effects; see also Section 3.5.2).

Note that other weather effects might be relevant as well, but should be covered from the fundamental point of view by the listed external inputs. Obvious examples are wind speed for the wind power prediction, cloud cover for the solar power production and illumination effects in the electricity consumption.

Many of those external effects may be explained by standard economic theory from fundamental electricity price models (Cludius et al., 2014, Kulakov and Ziel, 2021). Even the simple supply stack model (merit order model), see Fig. 8, explains many features and should be kept in mind when designing an appropriate electricity price forecasting model.

In recent years, statistical and machine learning methods gained a lot of attraction in day-ahead electricity price forecasting. Even though the majority of effects is linear there are specific non-linear dependencies that can be explored by using non-linear models, especially neural networks (Dudek, 2016, Lago et al., 2018, Marcjasz et al., 2019, Ugurlu et al., 2018). Of course this comes along with higher computational costs compared to linear models. Fezzi and Mosetti (2020) illustrate that even simple linear models can give highly accurate forecasts, if correctly calibrated. However, there seems to be consensus that forecast combination is appropriate, particularly for models that have different structures or different calibration window length (Gaillard et al., 2016, Hubicka et al., 2018, Mirakyan et al., 2017).



1. [Download: Download high-res image \(395KB\)](#)
2. [Download: Download full-size image](#)

Fig. 8. Illustrative example of a supply stack model with inelastic demand for different power plant types, roughly covering the situation in Germany 2020.

Another increasing stream of electricity price forecasting models do not focus on the electricity price itself, but the bid/sale/sell/supply and ask/sell/purchase/demand curves of the underlying auctions (see Fig. 7, but also [Kulakov, 2020](#), [Mestre et al., 2020](#), [Shah and Lisi, 2020](#), [Ziel and Steinert, 2016](#)). This sophisticated forecasting problem allows more insights for trading applications and the capturing of price clusters.

In forecasting intraday markets the literature just started to grow quickly. As the aforementioned market characteristics get less distinct if information from day-ahead markets is taken into account appropriately. However, intraday prices are usually more volatile and exhibit stronger price spikes. Thus, probabilistic forecasting is even more relevant ([Janke and Steinke, 2019](#), [Narajewski and Ziel, 2020b](#)). Recent studies showed that European markets are close to weak-form efficiency. Thus naive point forecasting benchmarks perform remarkably well ([Marcjasz et al., 2020](#), [Narajewski and Ziel, 2020a](#), [Oksuz and Ugurlu, 2019](#)).

As pointed out above, predicting price spikes is particularly important in practice, due to the high impact in decision making problems which occur usually in extreme situations, see Fig. 8. Very high electricity prices are usually observed in connection to high demand and low renewable energy generation, sometimes together with sudden power plant failures. In contrast, negative price spikes occur in oversupply situation, when there is low demand but high penetration from wind and solar power. The presence of spikes is explored in two main streams in literature: spike forecasting and prediction of prices under normal regime through robust estimators.

Within the first set of papers, spikes are often modelled as one regime of non-linear models for time series. This approach is followed by [Mount, Ning, and Cai \(2006\)](#) focusing on regime-switching models with parameters driven by time-varying variables and by [Becker, Hurn, and Pavlov \(2008\)](#) who adopt Markov switching models for spikes prediction. [Christensen et al., 2009](#), [Christensen et al., 2012](#) suggest treating and forecasting price spikes through Poisson autoregressive and discrete-time processes, respectively. [Herrera and González \(2014\)](#) use a Hawkes model combined with extreme events theory. Interregional links among different electricity markets are used by [Clements, Herrera, and Hurn \(2015\)](#) and [Manner, Türk, and Eichler \(2016\)](#) to forecast electricity price spikes. A new procedure for the simulation of electricity spikes has been recently proposed by [Muniain and Ziel \(2020\)](#) utilising bivariate jump components in a mean reverting jump diffusion model in the residuals.

The second stream of literature includes papers developing outlier detection methods or robust estimators to improve the forecasting performance of the models. [Martínez-Álvarez, Troncoso, Riquelme, and Aguilar-Ruiz \(2011\)](#) tackle the issue of outlier detection and prediction defining “motifs”, that is patches of units preceding observations marked as anomalous in a training set. [Janczura, Trück, Weron, and Wolff \(2013\)](#) focus on the detection and treatment of outliers in electricity prices. A very similar approach, based on seasonal autoregressive models and outlier filtering, is followed by [Afanasyev and Fedorova \(2019\)](#). [Grossi and Nan \(2019\)](#) introduced a procedure for the robust statistical prediction of electricity prices.

The econometric framework is represented by the robust estimation of non-linear SETAR processes. A similar approach has been followed by [Wang, Yang, Du, and Niu \(2020\)](#) using an outlier-robust machine learning algorithm.

3.4.3. Load forecasting¹²⁵

Load forecasting forms the basis where power system operation and planning builds upon. Based on the time horizon of the forecasts, load forecasting can be classified into very short-term (VSTLF), that refers to horizon from several minutes ahead up to 1 h, short-term (STLF), that spans from 1 h to 168 h ahead, medium-term (MTLF), that spans from 168 h to 1 year ahead and finally, and long-term (LTTLF) that concerns predictions from 1 year to several years ahead. In VSTLF and STLF applications, the focus is on the sub-hourly or hourly load. In MTLF and LTTLF, the variables of interest can be either monthly electricity peak load and total demand for energy.

Inputs differ in the various horizons. In VSTLF and STLF, apart from meteorological data, day type identification codes are used. In LTTLF, macroeconomic data are used since total demand of energy is influenced by the long-term modifications of the social and economic environments. Among the horizons, special attention is placed at STLF. This is reflected by the research momentum that have been placed in the load forecasting related literature by other researchers ([Hong & Fan, 2016](#)). Processes like unit commitment and optimal power flow rely on STLF ([Bo and Li, 2012](#), [Saksornchai et al., 2005](#)). Additionally, since competitive energy markets continually evolve, STLF becomes vital for new market entities such as retailers, aggregators, and prosumers for applications such as strategic bidding, portfolio optimisation, and tariff design ([Ahmad et al., 2019](#), [Danti and Magnani, 2017](#)).

The models that can be found in the load forecasting related literature can in general categorised into three types: time-series, machine learning, and hybrid. Time-series models historically precede the others. Typical examples of this family are ARMA, ARIMA, and others (see also Section 2.3.4). In the machine learning models, the structure is usually determined via the training process. NNs are commonly used. Once a NN is sufficiently trained, it can provide forecasts for all types of forecasting horizons ([Hippert, Pedreira, & Souza, 2001](#)). The third category of models refers to the integration of two or more individual forecasting approaches (see also see Section 2.7.13). For instance, a NN can be combined with time series methods, with unsupervised machine learning algorithms, data transformation, and with meta-heuristics algorithms ([Bozkurt et al., 2017](#), [El-Hendawi and Wang, 2020](#), [López et al., 2017](#), [Lu et al., 2019](#)).

Hybrid systems has been tested on validation data (through forecasting competitions), power system aggregated load, and application oriented tasks. [Ma \(2021\)](#) proposed an

ensemble method based on a combination of various single forecasters on GEFCom2012 forecasting competition data that outperformed benchmark forecasters such as Theta method, NN, ARIMA, and others (see Section 2.12.7 for further discussions on forecasting competitions). For aggregated load cases, researchers focus on different countries and energy markets. [Zhang, Wei, Li, Tan, and Zhou \(2018\)](#) combined empirical mode decomposition (EMD), ARIMA, and wavelet neural networks (WNN) optimised by the fruit fly algorithm on Australian Market data and New York City data. Their approach was to separate the linear and nonlinear components from original electricity load; ARIMA is used for linear part while the WNN for the non-linear one. [Sideratos, Ikonomopoulos, and Hatziargyriou \(2020\)](#) proposed that a radial basis network that performs the initial forecasting could serve as input to a convolutional neural network that performs the final forecasting. The proposed model led to lower error compared to the persistence model, NN, and SVM. [Semero, Zhang, and Zheng \(2020\)](#) focused on the energy management of a microgrid located in China using EMD to decompose the load, adaptive neuro-fuzzy inference system (ANFIS) for forecasting and particle swarm intelligence (PSO) to optimize ANFIS parameters. The results show that the proposed approach yielded superior performance over four other methods. [Faraji, Ketabi, Hashemi-Dezaki, Shafie-Khah, and Catalão \(2020\)](#) proposed a hybrid system for the scheduling of a prosumer microgrid in Iran. Various machine learning algorithms provided load and weather forecasts. Through an optimisation routine, the best individual forecast is selected. The hybrid system displayed better accuracy from the sole application of the individual forecasters.

3.4.4. Crude oil price forecasting¹²⁶

Crude oil, one of the leading energy resources, has contributed to over one-third of the world's energy consumption ([Alvarez-Ramirez, Soriano, Cisneros, & Suarez, 2003](#)). The fluctuations of the crude oil price have a significant impact on industries, governments as well as individuals, with substantial up-and-downs of the crude oil price bringing dramatic uncertainty for the economic and political development ([Cunado and De Gracia, 2005](#), [Kaboudan, 2001](#)). Thus, it is critical to develop reliable methods to accurately forecast crude oil price movement, so as to guard against the crude oil market extreme risks and improve macroeconomic policy responses. However, the crude oil price movement suffers from complex features such as nonlinearity, irregularities, dynamics and high volatility ([Alquist et al., 2013](#), [Herrera et al., 2018](#), [Kang et al.,](#)

2009 and also Section 2.3.11), making the crude oil price forecasting still one of the most challenging forecasting problems.

Some prior studies have suggested that the crude oil price movement is inherently unpredictable, and it would be pointless and futile to attempt to forecast future prices, see Miao, Ramchander, Wang, and Yang (2017) for a detailed summary. These agnostics consider the naive no-change forecast as the best available forecast value of future prices. In recent years, however, numerous studies result in forecasts that are more accurate than naive no-change forecasts, making the forecasting activities of crude oil prices promising (Alquist et al., 2013, Baumeister et al., 2015).

Extensive research on crude oil price forecasting has focused predominantly on the econometric models, such as VAR, ARCH-type, ARIMA, and Markov models (see, for example, Agnolucci, 2009, e Silva et al., 2010, Mirmirani and Li, 2004, Mohammadi and Su, 2010 and Section 2.3). In the forecasting literature, unit root tests (see Section 2.3.4) are commonly applied to examine the stationarity of crude oil prices prior to econometric modelling (Rahman and Serletis, 2012, Serletis and Rangel-Ruiz, 2004, Silvapulle and Moosa, 1999). It is well-documented that crude oil prices are driven by a large set of external components, which are themselves hard to predict, including supply and demand forces, stock market activities, oil-related events (e.g., war, weather conditions), political factors, etc. In this context, researchers have frequently considered structural models (see Section 2.3.9), which relate the oil price movements to a set of economic factors. With so many econometric models, is there an optimal one? Recently, de Albuquerque, de Medeiros, da Nóbrega Besarria, and Maia (2018) proposed a SETAR model, allowing for predictive regimes changing after a detected threshold, and achieved performance improvements over six widely used econometric models. Despite their high computational efficiency, the econometric models are generally limited in the ability to nonlinear time series modelling.

On the other hand, artificial intelligence and machine learning techniques, such as belief networks, support vector machines (SVMs), recurrent neural networks (RNNs), and extreme gradient boosting (XGBoost), provided powerful solutions to recognise the nonlinear and irregular patterns of the crude oil price movement with high automation (see, for example, Abramson and Finizza, 1991, Gumus and Kiran, 2017, Mingming and Jinliang, 2012, Xie et al., 2006). However, challenges also exist in these techniques, such as computational cost and overfitting. In addition, a large

number of studies have increasingly focused on the hybrid forecasting models (see also Section 2.7.13) based on econometrics models and machine learning techniques (Baumeister and Kilian, 2015, Chiroma et al., 2015, He et al., 2012, Jammazi and Aloui, 2012), achieving improved performance. Notably, the vast majority of the literature has focused primarily on the deterministic prediction, with much less attention paid to the probabilistic prediction and uncertainty analysis. However, the high volatility of crude oil prices makes probabilistic prediction more crucial to reduce the risk in decision-making (Abramson and Finizza, 1995, Sun et al., 2018).

3.4.5. Forecasting renewable energy technologies¹²⁷

The widespread adoption of renewable energy technologies, RETs, plays a driving role in the transition to low-carbon energy systems, a key challenge to face climate change and energy security problems. Forecasting the diffusion of RETs is critical for planning a suitable energy agenda and setting achievable targets in terms of electricity generation, although the available time series are often very short and pose difficulties in modelling. According to Rao and Kishore (2010), renewables' typical characteristics such as low load factor, need for energy storage, small size, high upfront costs create a competitive disadvantage, while Meade and Islam (2015b) suggested that renewable technologies are different from other industrial technological innovations because, in the absence of focused support, they are not convenient from a financial point of view. In this sense, policy measures and incentive mechanisms, such as feed-in tariffs, have been used to stimulate the market. As highlighted in Lee and Huh (2017b), forecasting RETs requires to capture different socio-economic aspects, such as policy choices by governments, carbon emissions, macroeconomic factors, economic and financial development of a country, competitive strength of traditional energy technologies. The complex and uncertain environment concerning RETs deployment has been faced in literature in several ways, in order to account for various determinants of the transition process. A first stream of research employed a bottom-up approach, where forecasts at a lower level are aggregated to higher levels within the forecasting hierarchy. For instance Park, Yun, Yun, Lee, and Choi (2016) realised a bottom-up analysis to study the optimum renewable energy portfolio, while Lee and Huh (2017a) performed a three-step forecasting analysis, to reflect the specificities of renewable sources, by using different forecasting methods for each of the sources considered. A similar bottom-up perspective was adopted in Zhang, Bauer, Yin, and Xie (2020), by conducting a multi-

region study, to understand how multi-level learning may affect RETs dynamics, with the regionalised model of investment and technological development, a general equilibrium model linking a macro-economic growth with a bottom-up engineering-based energy system model.

The relative newness of RETs has posed the challenge of forecasting with a limited amount of data: in this perspective, several contributions applied the ‘Grey System’ theory, a popular methodology for dealing with systems with partially unknown parameters (Kayacan, Ulutas, & Kaynak, 2010). Grey prediction models for RETs forecasting were proposed in Liu and Wu, 2021, Lu, 2019, Moonchai and Chutsagulprom, 2020, Tsai et al., 2017 and Wu, Ma, Zeng, Wang, and Cai (2019).

Other studies developed forecasting procedures based on growth curves and innovation diffusion models (see Sections 2.3.18 Innovation diffusion models, 2.3.19 The natural law of growth in competition, 2.3.20 Synchronic and diachronic competition): from the seminal work by Marchetti and Nakicenovic (1979), contributions on the diffusion of RETs were proposed by Bunea et al., 2020, Dalla Valle and Furlan, 2011, Guidolin and Mortarino, 2010, Lee and Huh, 2017b and Meade and Islam (2015b). Forecasting the diffusion of renewable energy technologies was also considered within a competitive environment in Furlan and Mortarino, 2018, Guidolin and Alpcan, 2019, Guidolin and Guseo, 2016 and Huh and Lee (2014).

3.4.6. Wind power forecasting¹²⁸

Wind energy is a leading source of renewable energy, meeting 4.8% of global electricity demand in 2018, more than twice that of solar energy (IEA, 2020). Kinetic energy in the wind is converted into electrical energy by wind turbines according to a characteristic ‘power curve’. Power production is proportion to the cube of the wind speed at low-to-moderate speeds, and above this is constant at the turbine’s rated power. At very high or low wind speeds no power is generated. Furthermore, the power curve is influenced by additional factors including air density, icing, and degradation of the turbine’s blades. Forecasts of wind energy production are required from minutes to days-ahead to inform the operation of wind farms, participation in energy markets and power systems operations. However, the limited predictability of the weather (see also Section 3.5.2) and the complexity of the power curve make this challenging. For this reason, probabilistic forecasts are increasingly used in practice (Bessa et al., 2017). Their value for energy trading is clear (Pinson, Chevallier, & Kariniotakis, 2007), but quantifying

value for power system operation is extremely complex. Wind power forecasting may be considered a mature technology as many competing commercial offerings exist, but research and development efforts to produce novel and enhanced products is ongoing (see also Section 3.4.5).

Short-term forecasts (hours to days ahead) of wind power production are generally produced by combining numerical weather predictions (NWP) with a model of the wind turbine, farm or even regional power curve, depending on the objective. The power curve may be modelled using physical information, e.g. provided by the turbine manufacturer, in which case it is also necessary to post-process NWP wind speeds to match the same height-above-ground as the turbine's rotor. More accurate forecasts can be produced by learning the NWP-to-energy relationship from historic data when it is available. State-of-the-art methods for producing wind power forecasts leverage large quantities of NWP data to produce a single forecast (Andrade, Filipe, Reis, & Bessa, 2017) and detailed information about the target wind farm (Gilbert, Browell, & McMillan, 2020a). A number of practical aspects may also need to be considered by users, such as maintenance outages and requirements to reduce output for other reasons, such as noise control or electricity network issues.

Very short-term forecast (minutes to a few hours ahead) are also of value, and on these time scales recent observations are the most significant input to forecasting models and more relevant than NWP. Classical time series methods perform well (see Section 2.3), and those which are able to capture spatial dependency between multiple wind farms are state-of-the-art, notably vector autoregressive models and variants (Cavalcante et al., 2016, Messner and Pinson, 2018). Care must be taken when implementing these models as wind power time series are bounded by zero and the wind farm's rated power meaning that errors may not be assumed to be normally distributed. The use of transformations is recommended (see also Section 2.2.1), though the choice of transformation depends on the nature of individual time series (Pinson, 2012).

Wind power forecasting is reviewed in detail in Giebel and Kariniotakis (2017), Hong, Pinson, Wang, Weron, Yang, and Zareipour (2020) and Zhang, Wang, and Wang (2014), and research is ongoing in a range of directions including: improving accuracy and reducing uncertainty in short-term forecasting, extending forecast horizons to weeks and months ahead, and improving very short-term forecast with remote sensing and data sharing (Sweeney, Bessa, Browell, & Pinson, 2019 and Section 3.4.10).

3.4.7. Wave forecasting¹²⁹

Ocean waves are primarily generated by persistent winds in one direction. The energy thus propagated by the wind is referred to as wave energy flux and follows a linear function of wave height squared and wave period. Wave height is typically measured as significant wave height, the average height of the highest third of the waves. The mean wave period, typically measured in seconds, is the average time between the arrival of consecutive crests, whereas the peak wave period is the wave period at which the highest energy occurs at a specific point.

The benefit of wave energy is that it requires significantly less reserve compared to those from wind (see Section 3.4.6) and solar (see Section 3.4.8) renewable energy sources (Hong et al., 2016). For example, the forecast error at one hour ahead for the simulated wave farms is typically in the range of 5%–7%, while the forecast errors for solar and wind are 17 and 22% respectively (Reikard, Pinson, & Bidlot, 2011). Solar power is dominated by diurnal and annual cycles but also exhibits nonlinear variability due to factors such as cloud cover, temperature and precipitation. Wind power is dominated by large ramp events such as irregular transitions between states of high and low power. Wave energy exhibits annual cycles and is generally smoother although there are still some large transitions, particularly during the winter months. In the first few hours of forecasting wave energy, time series models are known to be more accurate than numerical wave prediction. Beyond these forecast horizons, numerical wave prediction models such as SWAN (Simulating WAVes Nearshore, Booij, Ris, & Holthuijsen, 1999) and WAVEWATCH III® (Tolman, 2008) are widely used. As there is as yet no consensus on the most efficient model for harnessing wave energy, potential wave energy is primarily measured with energy flux, but the wave energy harnessed typically follows non-linear functions of wave height and wave period in the observations of the six different types of wave energy converters (Reikard, Robertson, Buckham, Bidlot, & Hiles, 2015).

To model the dependencies of wind speed, wave height, wave period and their lags, Reikard et al. (2011) uses linear regressions, which were then converted to forecasts of energy flux. Pinson, Reikard, and Bidlot (2012) use Reikard et al.'s (2011) regression model and log-normal distribution assumptions to produce probabilistic forecasts. López-Ruiz, Bergillos, and Ortega-Sánchez (2016) model the temporal dependencies of significant wave heights, peak wave periods and mean wave direction

using a vector autoregressive model, and used them to produce medium to long term wave energy forecasts. [Jeon and Taylor \(2016\)](#) model the temporal dependencies of significant wave heights and peak wave periods using a bivariate VARMA-GARCH (see also Section 2.3.11) to convert the two probabilistic forecasts into a probabilistic forecast of wave energy flux, finding this approach worked better than either univariate modelling of wave energy flux or bivariate modelling of wave energy flux and wind speed. [Taylor and Jeon \(2018\)](#) produce probabilistic forecasts for wave heights using a bivariate VARMA-GARCH model of wave heights and wind speeds, and using forecasts so as to optimise decision making for scheduling offshore wind farm maintenance vessels dispatched under stochastic uncertainty. On the same subject, [Gilbert, Browell, and McMillan \(2020b\)](#) use statistical post-processing of numerical wave predictions to produce probabilistic forecasts of wave heights, wave periods and wave direction and a logistic regression to determine the regime of the variables. They further applied the Gaussian copula to model temporal dependency but this did not improve their probabilistic forecasts of wave heights and periods.

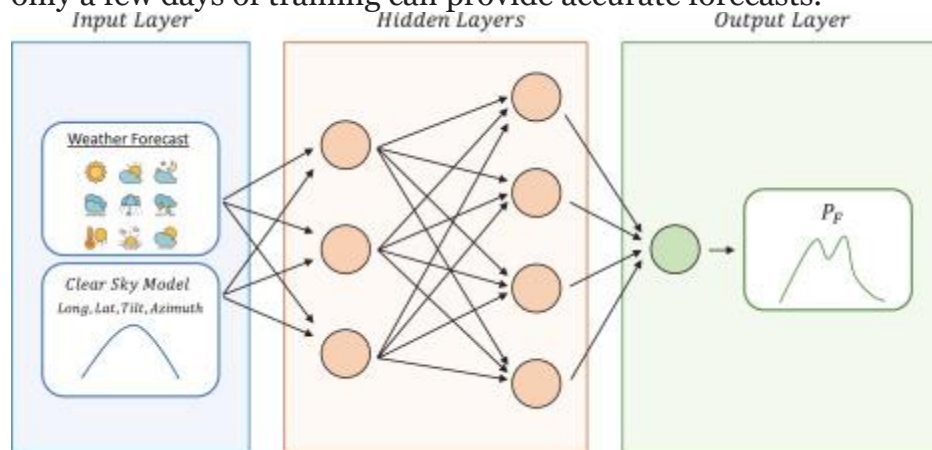
3.4.8. Solar power forecasting¹³⁰

Over the past few years, a number of forecasting techniques for photovoltaic (PV) power systems has been developed and presented in the literature. In general, the quantitative comparison among different forecast techniques is challenging, as the factors influencing the performance are numerous: the historical data, the weather forecast, the temporal horizon and resolution, and the installation conditions. A recent review by [Sobri, Koohi-Kamali, and Rahim \(2018\)](#) presents a comparative analysis of previous works, also including statistical errors. However, since the conditions and metrics used in each work were different, the comparison is not very meaningful. [Dolara, Grimaccia, Leva, Mussetta, and Ogliari \(2018\)](#) present relevant evaluation metrics for PV forecasting accuracy, while [Leva, Mussetta, & Ogliari, 2019](#) compare their effectiveness and immediate comprehension. In term of forecast horizon for PV power systems, intraday ([Nespoli et al., 2019](#)) and the 24 h of the next day ([Mellit et al., 2020](#)) are considered the most important.

[Nespoli et al. \(2019\)](#) compared two of the most widely used and effective methods for the forecasting of the PV production: a method based on Multi-Layer Perceptron (MLP) and a hybrid method using artificial neural network combined with clear sky solar radiation (see also Sections 2.7.8 Neural networks, 2.7.13 Hybrid methods). In the

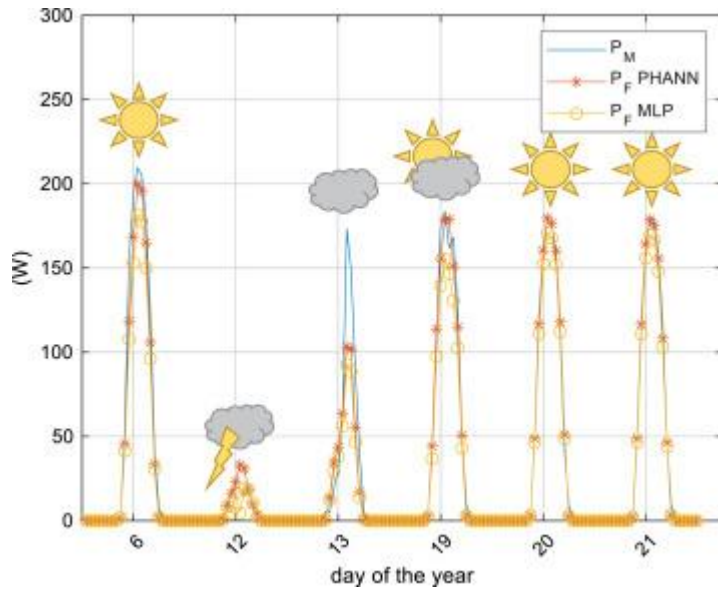
second case, the simulations are based on a feed-forward neural network (FFNN) but, among the inputs, the irradiation in clear sky conditions is provided. This method is called Physical Hybrid Artificial Neural Network (PHANN) and is graphically depicted in Fig. 9 (Dolara, Grimaccia, Leva, Mussetta, & Ogliari, 2015). PHANN method demonstrates better performance than classical NN methods. Fig. 10 shows a comparison between the measured and forecasted hourly output power of the PV plant for both sunny and cloudy days. The PHANN method shows good forecasting performance, especially for sunny days.

Ogliari, Dolara, Manzolini, and Leva (2017) compared the PV output power day-ahead forecasts performed by deterministic (based on three and five parameters electric equivalent circuit) and stochastic hybrid (based on artificial neural network models) methods aiming to find the best performance conditions. In general, there is no significant difference between the two deterministic models, with the three-parameter approach being slightly more accurate. Fig. 11 shows the daily value of normalised mean absolute error (NMAE%) for 216 days evaluated by using PHANN and three parameters electric circuit. The PHANN hybrid method achieves the best forecasting results, and only a few days of training can provide accurate forecasts.



1. [Download: Download high-res image \(236KB\)](#)
2. [Download: Download full-size image](#)

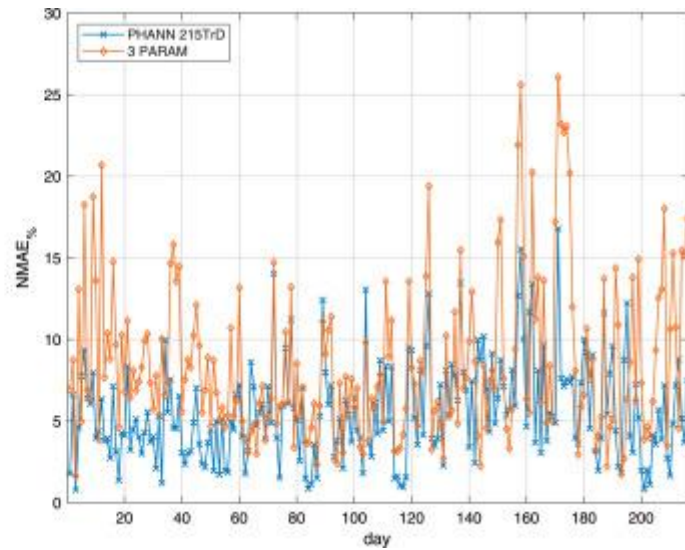
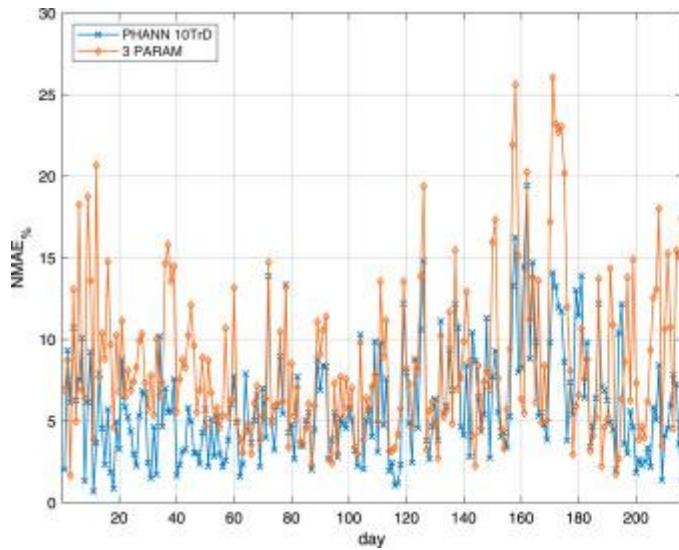
Fig. 9. Physical Hybrid Artificial Neural Network (PHANN) for PV power forecasting.



1. [Download: Download high-res image \(314KB\)](#)
2. [Download: Download full-size image](#)

Fig. 10. Measured versus forecasted output power by MLP and PHANN methods.

[Dolara et al. \(2018\)](#) analysed the effect of different approaches in the composition of a training data-set for the day-ahead forecasting of PV power production based on NN. In particular, the influence of different data-set compositions on the forecast outcome has been investigated by increasing the size of the training set size and by varying the lengths of the training and validation sets, in order to assess the most effective training method of this machine learning approach. As a general comment on the reported results, it can be stated that a method that employs the same chronologically consecutive samples for training is best suited when the availability of historical data is limited (for example, in newly deployed PV plant), while training based on randomly mixed samples method, appears to be most effective in the case of a greater data availability. Generally speaking, ensembles composed of independent trials are most effective.



1. [Download: Download high-res image \(977KB\)](#)
2. [Download: Download full-size image](#)

Fig. 11. Daily NMAE% of the PHANN method trained with 10 days (left) and with 215 days (right) compared with the three-parameters model.

3.4.9. Long-term simulation for large electrical power systems¹³¹

In large electrical power systems with renewable energy dependence, the power generators need to be scheduled to supply the system demand (de Queiroz, 2016). In general, for modelling long-term renewables future behaviour, such as hydro, wind and solar photovoltaics (PV), stochastic scenarios should be included in the scheduling, usually in a dispatch optimisation problem under uncertainty – like described, for small systems, in Section 3.4.1 and, for wave forecasting, in Section 3.4.7. Due to the complexity and uncertainty associated, this problem is, in general, modelled with time series scenarios and multi-stage stochastic approaches. de Queiroz (2016) presented a review for hydrothermal systems, with a focus on the optimisation algorithms.

Sections 3.4.6 Wind power forecasting, 3.4.8 Solar power forecasting explore the up-to-date methods for wind and PV solar power forecasting.

Here, we emphasise the importance of forecasting with simulation in the long-term renewable energy planning, especially in hydroelectric systems. In this context, due to the data spatial and temporal dependence structure, time series models are useful for future scenarios generation. Although the proposal could be forecasting for short-term planning and scheduling (as described in Section 3.4.6, 3.4.7, and

Section 3.4.8), simulation strategies are explored for considering and estimating uncertainty in medium and/or long-term horizons.

According to [Hipel and McLeod \(1994\)](#), stochastic processes of natural phenomena, such as the renewables ones, are, in general, stationary. One of the main features of hydroelectric generation systems is the strong dependence on hydrological regimes. To deal with this task, the literature focuses on two main classes for forecasting/simulation streamflow data: physical and data-driven models ([Zhang, Peng, Zhang, & Wang, 2015](#)). Water resources management for hydropower generation and energy planning is one of the main challenges for decision-makers. At large, the hydrological data are transformed into the so-called affluent natural energy, that is used for scenarios simulation and serve as input for the optimisation algorithms ([Oliveira, Souza, & Marcato, 2015](#)). The current state-of-the-art models for this proposal are the periodic ones. [Hipel and McLeod \(1994\)](#) presented a wide range of possibilities, but the univariate periodic autoregressive (PAR, a periodic extension version of the ones presented in Section 2.3.4) is still the benchmark, with several enhanced versions. The approach fits a model to each period of the historical data and the residuals are simulated to generate new future versions of the time series, considered stationary. Among many others, important variations and alternative proposals to PAR with bootstrap procedures (see bootstrap details in Section 2.7.5), Bayesian dynamic linear models, spatial information and copulas versions (for copulas references, see Section 2.4.3) are detailed in [Souza, Marcato, Dias, and Oliveira \(2012\)](#), [Marangon Lima, Popova, and Damien \(2014\)](#), [Lohmann, Hering, and Rebennack \(2016\)](#), and [de Almeida Pereira and Veiga \(2019\)](#), respectively.

It is worth considering the need for renewables portfolio simulation. This led [Pinheiro Neto, Domingues, Coimbra, de Almeida, Alves, and Calixto \(2017\)](#) to propose a model to integrate hydro, wind and solar power scenarios for Brazilian data. For the Eastern United States, [Shahriari and Blumsack \(2018\)](#) add to the literature on the wind, solar and blended portfolios over several spatial and temporal scales. For China, [Liu et al. \(2020\)](#) proposed a multi-variable model, with a unified framework, to simulate wind and PV scenarios to compensate hydropower generation. However, in light of the aforementioned, one of the key challenges and trends for renewable electrical power systems portfolio simulation are still related to the inclusion of

exogenous variables, such as climate, meteorological, calendar and economic ones, as mentioned in Section 3.4.2.

3.4.10. Collaborative forecasting in the energy sector¹³²

As mentioned in Section 3.4.6, the combination of geographically distributed time series data, in a collaborative forecasting (or data sharing) framework, can deliver significant improvements in the forecasting accuracy of each individual renewable energy power plant. The same is valid for hierarchical load forecasting (Hong et al., 2019) and energy price forecasting (see Section 3.4.2). A review of multivariate time series forecasting methods can be found in Section 2.3.9 2.3.11 and Section 2.4.3. However, this data might have different owners, which are unwilling to share their data due to the following reasons: (i) personal or business sensitive information, (ii) lack of understanding about which data can and cannot be shared, and (iii) lack of information about economic (and technical) benefits from data sharing.

In order to tackle these limitations, recent research in energy time series forecasting is exploring two alternative (and potentially complementary) pathways: (i) privacy-preserving analytics, and (ii) data markets.

The role of privacy-preserving techniques applied collaborative forecasting is to combine time series data from multiple data owners in order to improve forecasting accuracy and keep data private at the same time. For solar energy forecasting, Berdugo, Chaussin, Dubus, Hebrail, and Leboucher (2011) described a method based on local and global analog-search that uses solar power time series from neighbouring sites, where only the timestamps and normalised weights (based on similarity) are exchanged and not the time series data. Zhang and Wang (2018) proposed, for wind energy forecasting with spatio-temporal data, a combination of ridge linear quantile regression and Alternating Direction Method of Multipliers (ADMM) that enables each data owner to autonomously solve its forecasting problem, while collaborating with the others to improve forecasting accuracy. However, as demonstrated by Gonçalves, Bessa, and Pinson (2021a), the mathematical properties of these algorithms should be carefully analysed in order to avoid privacy breaches (i.e., when a third party recovers the original data without consent).

An alternative approach is to design a market (or auction) mechanism for time series or forecasting data where the data owners are willing to sell their private (or confidential) data in exchange for an economic compensation (Agarwal, Dahleh, & Sarkar, 2019). The

basic concept consists in pricing data as a function of privacy loss, but it can be also pricing data as a function of tangible benefits such as electricity market profit maximization. [Gonçalves, Pinson, and Bessa \(2021b\)](#) adapted for renewable energy forecasting the model described in [Agarwal et al. \(2019\)](#), by considering the temporal nature of the data and relating data price with the extra revenue obtained in the electricity market due to forecasting accuracy improvement. The results showed a benefit in terms of higher revenue resulting from the combination of electricity and data markets. With the advent of peer-to-peer energy markets at the domestic consumer level ([Parag & Sovacool, 2016](#)), smart meter data exchange between peers is also expected to increase and enable collaborative forecasting schemes. For this scenario, [Yassine, Shirehjini, and Shirmohammadi \(2015\)](#) proposed a game theory mechanism where a energy consumer maximizes its reward by sharing consumption data and a data aggregator can this data with a data analyst (which seeks data with the lowest possible price).

Finally, promoting data sharing via privacy-preserving or data monetisation can also solve data scarcity problems in some use cases of the energy sector, such as forecasting the condition of electrical grid assets ([Fan, Nowaczyk, & Rögnvaldsson, 2020](#)). Moreover, combination of heterogeneous data sources (e.g., numerical, textual, categorical) is a challenging and promising avenue of future research in collaborative forecasting ([Obst, Ghattas, Claudel, Cugliari, Goude, & Oppenheim, 2019](#)).

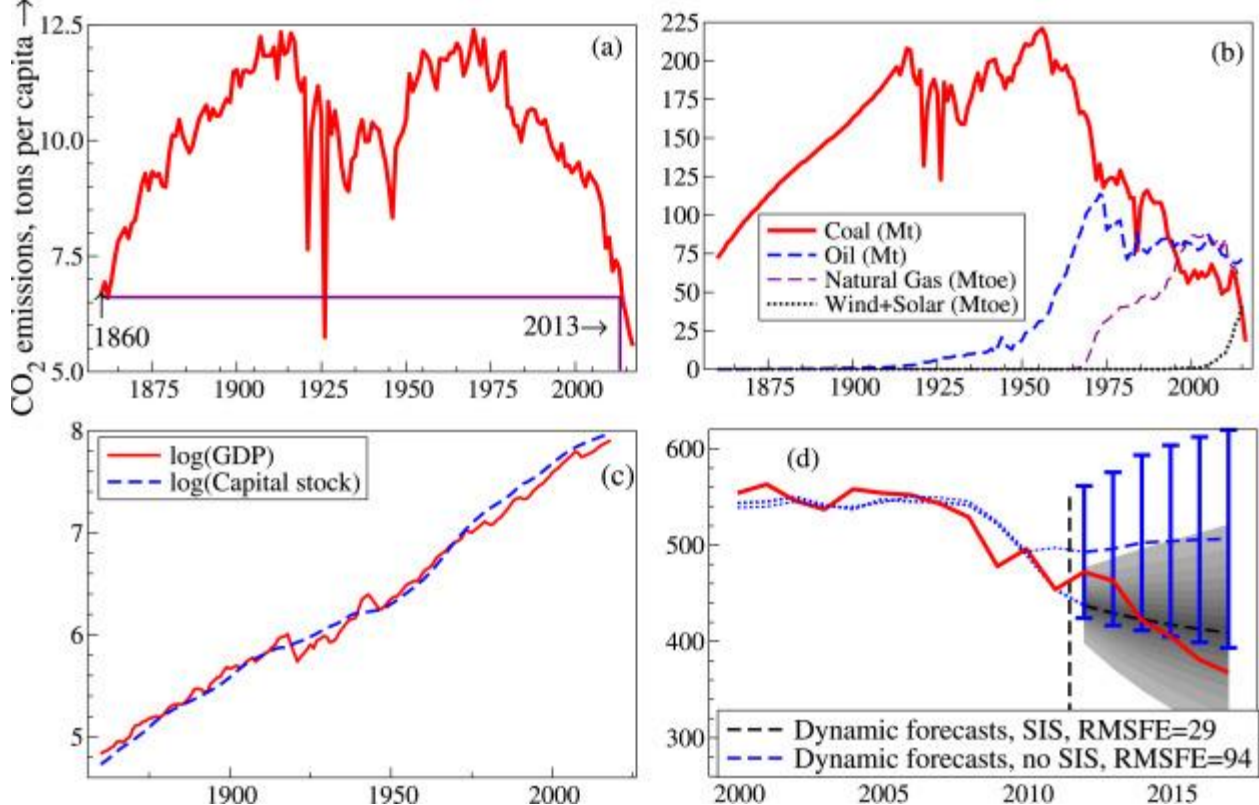
3.5. Environmental applications

3.5.1. Forecasting two aspects of climate change¹³³

First into the Industrial Revolution, the UK is one of the first out: in 2013 its per capita CO emissions dropped below their 1860 level, despite per capita real incomes being around 7-fold higher ([Hendry, 2020](#)). The model for forecasting UK CO emissions was selected from annual data 1860–2011 on CO emissions, coal and oil usage, capital and GDP, their lags and non-linearities (see Section 3.5.2 for higher frequency weather forecasts). [Figs. 12\(a\) to 12\(c\)](#) show the non-stationary time series with strong upward then downward trends, punctuated by large outliers from world wars, miners strikes plus shifts from legislation and technological change: ([Castle & Hendry, 2020a](#)). Saturation estimation at 0.1% using *Autometrics* ([Doornik, 2018](#)) retaining all other

regressors, detected 4 step shifts coinciding with major policy interventions like the 2008 Climate Change Act, plus numerous outliers, revealing a cointegrated relation. The multi-step forecasts over 2012–2017 from a VAR in panel (d) of Fig. 12 show the advantage of using step-indicator saturation (SIS: Castle et al., 2015b).

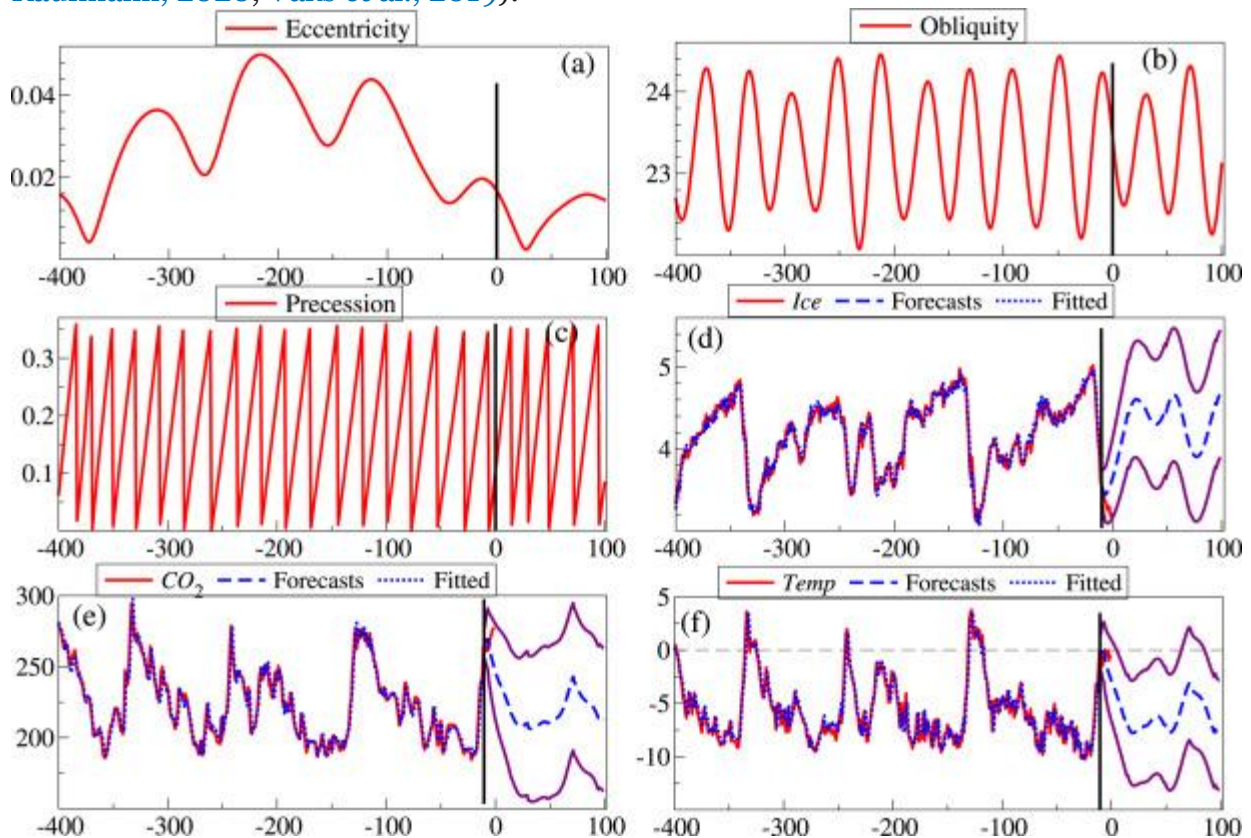
We formulated a 3-equation simultaneous model of atmospheric CO and Antarctic Temperature and Ice volume over 800,000 years of Ice Ages in 1000-year frequency (Kaufmann and Juselius, 2013, Paillard, 2001). Driven by non-linear functions of eccentricity, obliquity, and precession (see panels (a), (b), and (c) of Fig. 13 respectively), the model was selected with saturation estimation. Earth’s orbital path is calculable into the future (Croll, 1875 and Milankovitch, 1969), allowing 100,000 years of multi-step forecasts at endogenous emissions. Humanity has affected climate since 10 thousand years ago (kya: Ruddiman, 2005), so we commence forecasts there. Forecasts over to 100 with time series from 400kya in panels (d) to (f) of Fig. 13 show paths within the ranges of past data 2.2SE (Pretis & Kaufmann, 2018).



1. [Download: Download high-res image \(653KB\)](#)
2. [Download: Download full-size image](#)

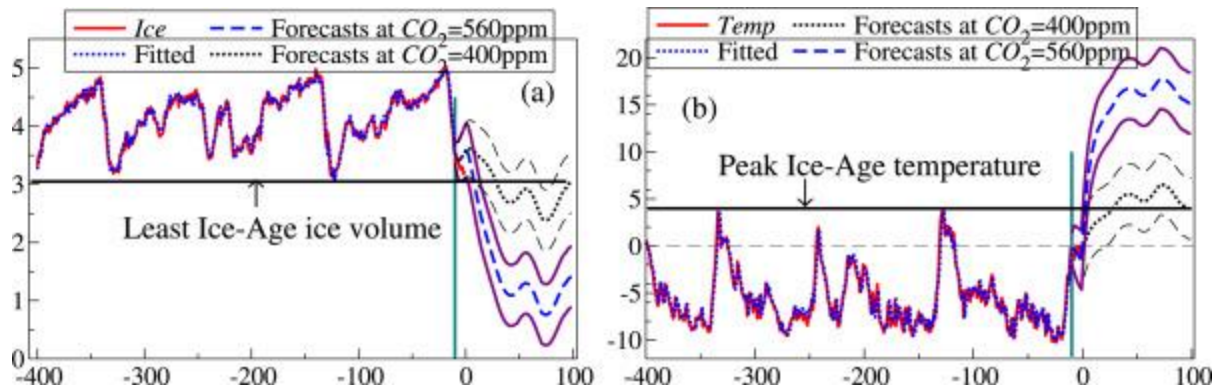
Fig. 12. (a) UK emissions, (b) energy sources in megatonnes (Mt) and megatonnes of oil equivalent (Mtoe), (c) economic variables, and (d) multi-step forecasts of CO₂ emissions in Mt.

Atmospheric CO already exceeds 400 ppm (parts per million), dramatically outside the Ice-Age range (Sundquist & Keeling, 2009). Consequently, we conditionally forecast the next 100,000 years, simulating the potential climate for anthropogenic CO (Castle & Hendry, 2020b) noting the ‘greenhouse’ temperature is proportional to the logarithm of CO (Arrhenius, 1896). The orbital drivers will continue to influence all three variables but that relation is switched off in the scenario for ‘exogenised’ CO. The 110 dynamic forecasts conditional on 400 ppm and 560 ppm with 2SE bands are shown in Fig. 14, panels (a) and (b) for Ice and Temperature respectively. The resulting global temperature rises inferred from these Antarctic temperatures would be dangerous, at more than 5 °C, with Antarctic temperatures positive for thousands of years (Pretis and Kaufmann, 2020, Vaks et al., 2019).



1. [Download: Download high-res image \(1MB\)](#)
2. [Download: Download full-size image](#)

Fig. 13. Ice-Age data, model fits, and forecasts with endogenous CO₂.



1. [Download: Download high-res image \(472KB\)](#)
2. [Download: Download full-size image](#)

Fig. 14. Ice-Age simulations with exogenous CO₂.

3.5.2. Weather forecasting¹³⁴

The weather has a huge impact on our lives, affecting health, transport, agriculture (see also Section 3.8.10), energy use (see also Section 3.4), and leisure. Since [Bjerknes \(1904\)](#) introduced hydrodynamics and thermodynamics into meteorology, weather prediction has been based on merging physical principles and observational information. Modern weather forecasting is based on numerical weather prediction (NWP) models that rely on accurate estimates of the current state of the climate system, including ocean, atmosphere and land surface. Uncertainty in these estimates is propagated through the NWP model by running the model for an ensemble of perturbed initial states, creating a weather forecast ensemble ([Buizza, 2018](#), [Toth and Buizza, 2019](#)).

One principal concern in NWP modelling is that small-scale phenomena such as clouds and convective precipitation are on too small a scale to be represented directly in the models and must, instead, be represented by approximations known as parameterisations. Current NWP model development aims at improving both the grid resolution and the observational information that enters the models ([Bannister et al., 2020](#), [Leuenberger et al., 2020](#)). However, for fixed computational resources, there is a trade-off between grid resolution and ensemble size, with a larger ensemble generally providing a better estimate of the prediction uncertainty. Recent advances furthermore include machine learning approaches (see Section 2.7.10) to directly model the small-scale processes, in particular cloud processes (see, for example, [Gentine et al., 2018](#), [Rasp et al., 2018](#)).

Despite rapid progress in NWP modelling, the raw ensemble forecasts exhibit systematic errors in both magnitude and spread (Buizza, 2018). Statistical post-processing is thus routinely used to correct systematic errors in calibration and accuracy before a weather forecast is issued; see Vannitsem, Wilks, and Messner (2018) for a recent review but also Sections 2.12.4 Evaluating probabilistic forecasts, 2.12.5 Assessing the reliability of probabilistic forecasts. A fundamental challenge here is to preserve physical consistency across space, time and variables (see, for example, Heinrich et al., 2020, Möller et al., 2013, Schefzik et al., 2013). This is particularly important when the weather forecast is used as input for further prediction modelling, e.g., in hydrometeorology (Hemri, 2018, Hemri et al., 2015). At time scales beyond two weeks, the weather noise that arises from the growth of the initial uncertainty, becomes large (Royer, 1993). Sources of long-range predictability are usually associated with the existence of slowly evolving components of the earth system, including the El Niño Southern Oscillation (ENSO), monsoon rains, the Madden Julian Oscillation (MJO), the Indian Ocean dipole, and the North Atlantic Oscillation (NAO), spanning a wide range of time scales from months to decades (Hoskins, 2013, Vitart et al., 2012). It is expected that, if a forecasting system is capable of reproducing these slowly evolving components, they may also be able to forecast them (Van Schaeybroeck & Vannitsem, 2018). The next step is then to find relationships between modes of low-frequency variability and the information needed by forecast users such as predictions of surface temperature and precipitation (Roulin and Vannitsem, 2019, Smith et al., 2020).

3.5.3. Air quality forecasting¹³⁵

To preserve human health, European Commission stated in the Directive (2008/50/EC) that member states have to promptly inform the population when the particulate matter (PM) daily mean value exceeds (or is expected to exceed) the threshold of . Therefore, systems have been designed in order to produce forecasts for up to three days in advance using as input the measured value of concentration and meteorological conditions. These systems can be classified in (i) data-driven models (Carnevale et al., 2016, Corani, 2005, Stadlober et al., 2018 and Section 2.7), and (ii) deterministic chemical and transport models (Honoré et al., 2007, Manders et al., 2009). In this section, a brief overview of the application of these systems to the high polluted area of Lombardy region, in Italy, will be presented.

Carnevale, Finzi, Pederzoli, Turrini, and Volta (2018) compared the results of three different forecasting systems based on neural networks, lazy learning models, and regression trees respectively. A single model has been identified for each monitoring station. In the initial configuration, only the last three PM measurements available were used to produce the forecast. In this configuration, the systems offered reasonable performance, with correlation coefficients ranging from 0.6 (lazy learning method) to 0.75 (neural network). The work also demonstrated that the performance of the ensemble of the three systems was better than the best model for each monitoring station (see also Section 2.6 for further discussions on forecast combinations).

Starting from the results of this work, a second configuration was implemented, using as input also the wind speed measured in the meteorological monitoring station closest to the measurement point of PM. The researchers observed an improvement in all performance indices, with the median of the correlation for the best model (neural networks) increasing from 0.75 to 0.82 and the RMSE dropping from to .

One of the main drawbacks of data-driven models for air quality is that they provide information only in the point where the measurements are available. To overcome this limitation, recent literature has presented mixed deterministic and data-driven approaches (see, for example, Carnevale, Angelis, Finzi, Turrini, & Volta, 2020) which use the data assimilation procedure and offer promising forecasting performance.

From a practical point of view, critical issues regarding forecasting air quality include:

- •
Information collection and data access: even if regional authorities have to publicly provide data and information related to air quality and meteorology, the measured data are not usually available in real-time and the interfaces are sometimes not automated;
- •
Data quantity: the amount of information required by air quality forecasting systems is usually large, in particular towards the definition of the training and validation sets;
- •
Non-linear relationships: the phenomenon of accumulation of pollutants in atmosphere is usually affected by strong nonlinearities, which significantly impact the selection of the models and their performance;

-
-

Unknown factors: it is a matter of fact that the dynamic of pollutants in atmosphere is affected by a large number of non-measurable variables (such as meteorological variables or the interaction with other non-measurable pollutants), largely affecting the capability of the models to reproduce the state of the atmosphere.

3.5.4. Forecasting and decision making for floods and water resources management¹³⁶

In Water Resources and Flood Risk Management, decision makers are frequently confronted with the need of taking the most appropriate decisions not knowing what will occur in the future. To support their decision-making under uncertainty, decision theory (Berger, 1985, Bernardo, 1994, DeGroot, 2004) invokes Bayesian informed decision approaches, which find the most appropriate decision by maximising (or minimising) the expected value of a “utility function”, thus requiring its definition, together with the estimation of a “predictive probability” density (Berger, 1985) due to the fact that utility functions are rarely linear or continuous. Consequently, their expected value does not coincide with the value assumed on the predicted “deterministic” expected value. Accordingly, overcoming the classical 18th century “mechanistic” view by resorting into probabilistic forecasting approaches becomes essential (see also Section 2.6.2).

The failure of decision-making based on deterministic forecasts in the case of Flood Risk Management is easily shown through a simple example. At a river section, the future water level provided by a forecast is uncertain and can be described by a Normal distribution with mean 10 meters and standard deviation of 5 m. Given a dike elevation of 10.5 m, damages may be expected as zero if water level falls below the dike elevation and linearly growing when level exceeds it with a factor of dollars. If one assumes the expect value of forecast as the deterministic prediction to compute the damage the latter will result equal to zero, while if one correctly integrates the damage function times the predictive density the estimated expected damage will results into 6.59 millions of dollars and educated decisions on alerting or not the population or evacuating or not a flood-prone area can be appropriately taken (see also Section 3.6).

Water resources management, and in particular reservoirs management, aim at deriving appropriate operating rules via long term expected benefits maximisation. Nonetheless, during flood events decision makers must decide how much to preventively release from multi-purpose reservoirs in order to reduce dam failure and downstream flooding risks the optimal choice descending from trading-off between loosing future water resource vs the reduction of short term expected losses.

This is obtained by setting up an objective function based on the linear combination of long and short term “expected losses”, once again based on the available probabilistic forecast. This Bayesian adaptive reservoir management approach incorporating into the decision mechanism the forecasting information described by the short-term predictive probability density, was implemented on the lake Como since 1997 (Todini, 1999, Todini, 2017) as an extension of an earlier original idea (Todini, 1991). This resulted into:

- •
a reduction of over 30% of the city of Como frequency;
- •
an average reduction of 12% of the water deficit;
- •
an increase of 3% in the electricity production.

Lake Como example clearly shows that instead of basing decisions on the deterministic prediction, the use of a Bayesian decision scheme, in which model forecasts describe the predictive probability density, increases the reliability of the management scheme by essentially reducing the probability of wrong decisions (Todini, 2017, Todini, 2018).

3.6. Social good and demographic forecasting

3.6.1. Healthcare¹³⁷

There are many decisions that depend on the quality of forecasts in the health care system, from capacity planning to layout decisions to the daily schedules. In general, the role of forecasting in health care is to inform both clinical and non-clinical decisions. While the former concerns decisions related to patients and their treatments (Makridakis, Kirkham, Wakefield, Papadaki, Kirkham, & Long, 2019), the

latter involves policy/management, and supply chain decisions that support the delivery of high-quality care for patients.

A number of studies refer to the use of forecasting methods to inform clinical decision making. These methods are used to screen high risk patients for preventative health care (Chen et al., 2015, Santos et al., 2015, Uematsu et al., 2014, van der Mark et al., 2014), to predict mental health issues (Shen et al., 2017, Tran et al., 2013), to assist diagnosis and disease progression (Ghassemi et al., 2015, Ma et al., 2017, Pierce et al., 2010, Qiao et al., 2019), to determine prognosis (Dietzel et al., 2010, Ng et al., 2007), and to recommend treatments for patients (Kedia and Williams, 2003, Scerri et al., 2006, Shang et al., 2019). Common forecasting methods to inform clinical decisions include time series (see Sections 2.3.1 Exponential smoothing models, 2.3.4 Autoregressive integrated moving average (ARIMA) models, 2.3.5 Forecasting for multiple seasonal cycles), regression models (see Section 2.3.2), classification trees (see Section 2.7.12), neural networks (see Section 2.7.8), Markov models (see Section 2.3.12) and Bayesian networks. These models utilise structured and unstructured data including clinician notes (Austin and Kusumoto, 2016, Labarere et al., 2014) which makes the data pre-processing a crucial part of the forecasting process in clinical health care.

One of the aspects of the non-clinical forecasting that has received the most attention in both research and application is the policy and management. Demand forecasting is regularly used in Emergency Departments (Arora et al., 2020, Choudhury and Urena, 2020, Khaldi et al., 2019, Rostami-Tabar and Ziel, 2020), ambulance services (Al-Azzani et al., 2020, Setzler et al., 2009, Vile et al., 2012, Zhou and Matteson, 2016) and hospitals with several different specialities (McCoy et al., 2018, Ordu et al., 2019, Zhou et al., 2018) to inform operational, tactical and strategic planning. The common methods used for this purpose include classical ARIMA and exponential smoothing methods, regression, singular spectrum analysis, Prophet, Double-Seasonal Holt-Winter, TBATS and Neural Networks. In public health, forecasting can guide policy and planning. Although it has a wider definition, the most attention is given to Epidemic forecasting (see also Section 3.6.2).

Forecasting is also used in both national and global health care supply chains, not only to ensure the availability of medical products for the population but also to avoid excessive inventory. Additionally, the lack of accurate demand forecast in a health

supply chain may cost lives (Baicker, Chandra, & Skinner, 2012) and has exacerbated risks for suppliers (Levine, Pickett, Sekhri, & Yadav, 2008). Classical exponential smoothing, ARIMA, regression and Neural Network models have been applied to estimate the drug utilisation and expenditures (Dolgin, 2010, Linnér et al., 2020), blood demand (Fortsch & Khapalova, 2016), hospital supplies (Gebicki et al., 2014, Riahi et al., 2013) and demand for global medical items (Amarasinghe et al., 2010, Hecht and Gandhi, 2008, van der Laan et al., 2016). It is important to note that, while the demand in a health care supply chain has often grouped and hierarchical structures (Mircetica, Rostami-Tabar, Nikolicica, & Maslarica, 2020 see also Section 2.10.1), this has not been well investigated and needs more attention.

3.6.2. Epidemics and pandemics¹³⁸

Pandemics and epidemics both refer to disease outbreaks. An epidemic is a disease outbreak that spreads across a particular region. A pandemic is defined as spread of a disease worldwide. Forecasting the evolution of a pandemic or an epidemic, the growth of cases and fatalities for various horizons and levels of granularity, is a complex task with raw and limited data – as each disease outbreak type has unique features with several factors affecting the severity and the contagiousness. Be that as it may, forecasting becomes an paramount task for the countries to prepare and plan their response (Nikolopoulos, 2020), both in healthcare and the supply chains (Beliën & Forcé, 2012 see also Section 3.6.1 and Section 3.2.2).

Successful forecasting methods for the task include time-series methods (see Section 2.3), epidemiological and agent-based models (see Section 2.7.3), metapopulation models, approaches in metrology (Nsoesie, Mararthe, & Brownstein, 2013), machine and deep learning methods (Yang, Zeng, Wang, Wong, Liang, Zanin, Liu, Cao, Gao, Mai, Liang, Liu, Li, Li, Ye, Guan, Yang, Li, Luo, Xie, Liu, Wang, Zhang, Wang, Zhong, & He, 2020). Andersson, Köhlmann-Berenzon, Linde, Schiöler, Rubinova, and Frisé (2008) used regression models for the prediction of the peak time and volume of cases for a pandemic with evidence from seven outbreaks in Sweden. Yaffee, Nikolopoulos, Reilly, Crone, Wagoner, Douglas, Amman, Ksiazek, and Mills (2011) forecasted the evolution of the Hantavirus epidemic in USA and compared causal and machine-learning methods with time-series methods and found that univariate methods quite successful. Soebiyanto, Adimi, and Kiang (2010) used ARIMA models for successfully short-term forecasting of influenza weekly cases. Shaman and

[Karspeck \(2012\)](#) used Kalman filter based SIR epidemiological models to forecast the peak time of influenza 6–7 weeks ahead.

For COVID-19, [Petropoulos and Makridakis \(2020\)](#) applied a multiplicative exponential smoothing model (see also Section 2.3.1) for predicting global number of confirmed cases, with very successful results both for point forecasts and prediction intervals. This article got serious traction with 100,000 views and 300 citations in the first twelve months since its publication, thus evidencing the importance of such empirical investigations. There has been a series of studies focusing on predicting deaths in the USA and European countries for the first wave of the COVID-19 pandemic ([IHME COVID-19 health service utilization forecasting team and Murray, 2020a](#), [IHME COVID-19 health service utilization forecasting team and Murray, 2020b](#)).

Furthermore, [Petropoulos, Makridakis, and Stylianou \(2020\)](#) expanded their investigation to capture the continuation of both cases and deaths as well as their uncertainty, achieving high levels of forecasting accuracy for ten-days-ahead forecasts over a period of four months. Along the same lines, [Doornik et al. \(2020b\)](#) have been publishing real-time accurate forecasts of confirmed cases and deaths from mid-March 2020 onwards. Their approach is based on extraction of trends from the data using machine learning.

[Pinson and Makridakis \(2020\)](#) organised a debate between Taleb and Ioannidis on forecasting pandemics. [Ioannidis, Cripps, and Tanner \(2020\)](#) claim that forecasting for COVID-19 has by and large failed. However they give recommendations of how this can be averted. They suggest that the focus should be on predictive distributions and models should be continuously evaluated. Moreover, they emphasise the importance of multiple dimensions of the problem (and its impact). [Taleb et al. \(2020\)](#) discuss the dangers of using naive, empirical approaches for fat-tailed variables and tail risk management.

They also reiterate the inefficiency of point forecasts for such phenomena.

Finally, [Nikolopoulos, Punia, Schäfers, Tsinopoulos, and Vasilakis \(2020\)](#) focused on forecast-driven planning, predicting the growth of COVID-19 cases and the respective disruptions across the supply chain at country level with data from the USA, [India](#), UK, Germany, and Singapore. Their findings confirmed the excess demand for groceries and electronics, and reduced demand for automotive – but the model also proved that the earlier a lock-down is imposed, the higher the excess demand will be for groceries.

Therefore, governments would need to secure high volumes of key products before

imposing lock-downs; and, when this is not possible, seriously consider more radical interventions such as rationing.

Dengue is one of the most common epidemic diseases in tropical and sub-tropical regions of the world. Estimates of World Health Organisation reveals that about half of the world's population is now at risk for Dengue infection (Romero, Olivero, Real, & Guerrero, 2019). *Aedes aegypti* and *Aedes albopictus* are the principal vectors of dengue transmission and they are highly domesticated mosquitoes. Rainfall, temperature and relative humidity are thought of as important factors attributing towards the growth and dispersion of mosquito vectors and potential of dengue outbreaks (Banu, Hu, Hurst, & Tong, 2011).

In reviewing the existing literature, two data types have been used to forecast dengue incidence: (i) spatio-temporal data: incidence of laboratory-confirmed dengue cases among the clinically suspected patients (Naish, Dale, Mackenzie, McBride, Mengersen, & Tong, 2014), (ii) web-based data: Google trends, tweets associated with Dengue cases (de Almeida Marques-Toledo et al., 2017).

SARIMA models (see also Section 2.3.4) have been quite popular in forecasting laboratory-confirmed dengue cases (Gharbi et al., 2011, Martinez and Silva, 2011, Promprou et al., 2006). Chakraborty, Chattopadhyay, and Ghosh (2019) used a hybrid model combining ARIMA and neural network autoregressive (NNAR) to forecast dengue cases. In light of biological relationships between climate and transmission of *Aedes* mosquitoes, several studies have used additional covariates such as, rainfall, temperature, wind speed, and humidity to forecasts dengue incidence (Banu et al., 2011, Naish et al., 2014, Talagala, 2015). Poisson regression model has been widely used to forecast dengue incidence using climatic factors and lagged time between dengue incidence and weather variables (Hii et al., 2012, Koh et al., 2018). Several researchers looked at the use of Quasi-Poisson and negative binomial regression models to accommodate over dispersion in the counts (Lowe et al., 2011, Wang et al., 2014). Cazelles, Chavez, McMichael, and Hales (2005) used wavelet analysis to explore the dynamic of dengue incidence and wavelet coherence analyses was used to identify time and frequency specific association with climatic variables. de Almeida Marques-Toledo et al. (2017) took a different perspective and look at weekly tweets to forecast Dengue cases. Rangarajan, Mody, and Marathe (2019) used Google trend data to forecast Dengue cases. Authors hypothesised that web query search related to dengue

disease correlated with the current level of dengue cases and thus may be helpful in forecasting dengue cases.

A direction for future research in this field is to explore the use of spatio-temporal hierarchical forecasting (see Section 2.10).

3.6.3. Forecasting mortality¹³⁹

Actuarial, Demographic, and Health studies are some examples where mortality data are commonly used. A valuable source of mortality information is the Human Mortality Database (HMD), a database that provides mortality and population data for 41 mainly developed countries. Additionally, at least five country-specific databases are devoted to subnational data series: Australian, Canadian, and French Human Mortality Databases, United States and Japan Mortality Databases. In some situations, the lack of reliable mortality data can be a problem, especially in developing countries, due to delays in registering or miscounting deaths (Checchi & Roberts, 2005). Analysis of National Causes of Death for Action (ANACONDA) is a valuable tool that assesses the accuracy and completeness of data for mortality and cause of death by checking for potential errors and inconsistencies (Mikkelsen, Moesgaard, Hegnauer, & Lopez, 2020).

The analysis of mortality data is fundamental to public health authorities and policymakers to make decisions or evaluate the effectiveness of prevention and response strategies. When facing a new pandemic, mortality surveillance is essential for monitoring the overall impact on public health in terms of disease severity and mortality (Setel et al., 2020, Vestergaard et al., 2020). A useful metric is excess mortality and is the difference between the observed number of deaths and the expected number of deaths under “normal” conditions (Aron and Muellbauer, 2020, Checchi and Roberts, 2005). Thus, it can only be estimated with accurate and high-quality data from previous years. Excess mortality has been used to measure the impact of heat events (Limaye et al., 2018, Matte et al., 2016), pandemic influenza (Nielsen et al., 2013, Nunes et al., 2011), and nowadays COVID-19 (Nogueira et al., 2020, Ritchie et al., 2020, Shang and Xu, 2021, Sinnathamby et al., 2020 and Section 3.6.2), among others. Excess mortality data have been making available by the media publications *The Economist*, *The New York Times* and *The Financial Times*. Moreover, a monitoring system of the weekly excess mortality in Europe has been performed by the EuroMOMO project (Vestergaard et al., 2020).

An essential use of mortality data for those individuals at age over 60 is in the pension and insurance industries, whose profitability and solvency crucially rely on accurate mortality forecasts to adequately hedge longevity risks (see, e.g., [Shang and Haberman, 2020a](#), [Shang and Haberman, 2020b](#)). Longevity risk is a potential systematic risk attached to the increasing life expectancy of annuitants, and it is an important factor to be considered when determining a sustainable government pension age (see, e.g., [Hyndman, Zeng, & Shang, 2021](#) for Australia). The price of a fixed-term or lifelong annuity is a random variable, as it depends on the value of zero-coupon bond price and mortality forecasts. The zero-coupon bond price is a function of interest rate (see Section 3.3.6) and is comparably more stable than the retirees' mortality forecasts. Several methodologies were developed for mortality modelling and forecasting ([Booth and Tickle, 2008](#), [Janssen, 2018](#)). These methods can be grouped into three categories: expectation, explanation, and extrapolation ([Booth & Tickle, 2008](#)).

The expectation approach is based on the subjective opinion of experts (see also Section 2.11.4), who set a long-run mortality target. Methods based on expectation make use of experts' opinions concerning future mortality or life expectancy with a specified path and speed of progression towards the assumed value ([Continuous Mortality Investigation, 2020](#)). The advantage of this approach is that demographic, epidemiological, medical, and other relevant information may be incorporated into the forecasts. The disadvantages are that such information is subjective and biased towards experts' opinions, and it only produces scenario-based (see Section 2.11.5) deterministic forecasts ([Ahlburg and Vaupel, 1990](#), [Wong-Fupuy and Haberman, 2004](#)).

The explanation approach captures the correlation between mortality and the underlying cause of death. Methods based on the explanation approach incorporate medical, social, environmental, and behavioural factors into mortality modelling. Example include smoking and disease-related mortality models. The benefit of this approach is that mortality change can be understood from changes in related explanatory variables; thus, it is attractive in terms of interpretability ([Guterman & Vanderhoof, 1998](#)).

The extrapolative approach is considered more objective, easy to use and more likely to obtain better forecast accuracy than the other two approaches ([Janssen, 2018](#)). The extrapolation approach identifies age patterns and trends in time which can be then forecasted via univariate and multivariate time series models (see Section 2.3). In the

extrapolation approach, many parametric and nonparametric methods have been proposed (see, e.g., [Alho and Spencer, 2005](#), [Hyndman and Ullah, 2007](#), [Shang et al., 2011](#)). Among the parametric methods, the method of [Heligman and Pollard \(1980\)](#) is well-known. Among the nonparametric methods, the Lee–Carter model ([Lee & Carter, 1992](#)), Cairns-Blake-Dowd model ([Cairns et al., 2009](#), [Dowd et al., 2010](#)), and functional data model ([Hyndman & Ullah, 2007](#) and Section 2.3.10), as well as their extensions and generalisations are dominant. The time-series extrapolation approach has the advantage of obtaining a forecast probability distribution rather than a deterministic point forecast and, also, enable the determination of forecast intervals ([Booth & Tickle, 2008](#)).

[Janssen \(2018\)](#) presents a review of the advances in mortality forecasting and possible future research challenges.

3.6.4. Forecasting fertility¹⁴⁰

Aside from being a driver of population forecasts (see Section 2.3.7), fertility forecasts are vital for planning maternity services and anticipating demand for school places. The key challenge relates to the existence of, and interaction between, the quantum (how many?) and tempo (when?) components ([Booth, 2006](#)). This intrinsic dependence on human decisions means that childbearing behaviour is influenced by numerous factors acting at different levels, from individual characteristics to societal change ([Balbo, Billari, & Mills, 2013](#)). An important methodological challenge for many low- and middle-income countries is fertility estimation, due to deficiencies in vital statistics caused by inadequate birth registration systems ([AbouZahr et al., 2015](#), [Moultrie et al., 2013](#), [Phillips et al., 2018](#)). Such countries are often also in the process of transitioning from high to low fertility, which induces greater forecast uncertainty compared to low-fertility countries ([Programme, 2019](#)).

A range of statistical models have been proposed to forecast fertility – see [Booth \(2006\)](#), [Bohk-Ewald, Li, and Myrskylä \(2018\)](#), and [Shang and Booth \(2020\)](#) for reviews. The Lee–Carter model ([Lee & Carter, 1992](#)) originally developed for mortality forecasting (see Section 3.6.3) has been applied to fertility ([Lee, 1993](#)), with extensions in functional data ([Hyndman & Ullah, 2007](#)) and Bayesian ([Wiśniowski, Smith, Bijak, Raymer, & Forster, 2015](#)) contexts. Other notable extrapolative methods include the cohort-ARIMA model of [De Beer, 1985](#), [De Beer, 1990](#) – see Section 2.3.4 – and the linear extrapolation method of [Myrskylä, Goldstein, and Cheng \(2013\)](#). Many parametric models have been specified to describe the shapes of fertility curves ([Brass,](#)

1974, Evans, 1986, Hoem et al., 1981, Schmertmann, 2003), with forecasts obtained through time series extrapolations of the parameters (Congdon, 1990, De Iaco and Maggio, 2016, Knudsen et al., 1993). Bayesian methods have been used to borrow strength across countries (for example, Alkema et al., 2011, Schmertmann et al., 2014), with Ellison, Dodd, and Forster (2020) developing a hierarchical model in the spirit of the latter. The top-down approach (see Section 2.10.1) of the former, which is used by the United Nations, projects the aggregate Total Fertility Rate (TFR) measure probabilistically (also see Tuljapurkar & Boe, 1999) before decomposing it by age. Hyppöla, Tunkelo, and Törnqvist (1949) provide one of the earliest known examples of probabilistic fertility forecasting (Alho & Spencer, 2005). Little work has been done to compare forecast performance across this broad spectrum of approaches. The study of Bohk-Ewald et al. (2018) is the most comprehensive to date. Most striking is their finding that few methods can better the naive freezing of age-specific rates, and those that can differ greatly in method complexity (see also Section 2.5.2). A recent survey of fertility forecasting practice in European statistical offices (Gleditsch & Syse, 2020) found that forecasts tend to be deterministic and make use of expert panels (see Section 2.11.4). Expert elicitation techniques are gaining in sophistication, highlighted by the protocol of Statistics Canada (Dion, Galbraith, & Sirag, 2020) which requests a full probability distribution of the TFR. A promising avenue is the development of forecasting methods that incorporate birth order (parity) information, supported by evidence from individual-level analyses (for example, Fiori, Graham, & Feng, 2014). Another underexplored area is the integration of survey data into fertility forecasting models, which tend to use vital statistics alone when they are of sufficient quality (see Rendall et al., 2009, Zhang and Bryant, 2019 for Bayesian fertility estimation with imperfect census data). Alternative data sources also have great potential. For example, Wilde, Chen, and Lohmann (2020) use Google data to predict the effect of COVID-19 on US fertility in the absence of vital statistics. Lastly, investigation of the possible long-term impacts of delayed motherhood in high-income countries, alongside developments in assisted reproduction technology such as egg freezing, is required (see, for example, Sobotka & Beaujouan, 2018).

3.6.5. Forecasting migration¹⁴¹

Migration forecasts are needed both as crucial input into population projections (see Section 2.3.7), as well as standalone predictions, made for a range of users, chiefly in the

areas of policy and planning. At the same time, migration processes are highly uncertain and complex, with many underlying and interacting drivers, which evade precise conceptualisation, definitions, measurement, and theoretical description (Bijak & Czaika, 2020). Given the high level of the predictive uncertainty, and the non-stationary character of many migration processes (Bijak & Wiśniowski, 2010), the current state of the art of forward-looking migration studies reflects therefore a shift from prediction to the use of forecasts as contingency planning tools (idem).

Reviews of migration forecasting methods are available in Bijak (2010) and Sohst, Tjaden, de Valk, and Melde (2020). The applications in official statistics, with a few exceptions, are typically based on various forms scenario-based forecasting with judgment (see Section 2.11.5), based on pre-set assumptions (for an example, see Abel, 2018). Such methods are particularly used for longer time horizons, of a decade or more, so typically in the context of applications in population projections, although even for such long time horizons calibrated probabilistic methods have been used as well (Azose et al., 2016).

The mainstream developments in migration forecasting methodology, however, include statistical and econometric methods discussed in Section 2.3, such as time series models, both uni- and multivariate (for example, Bijak, 2010, Bijak et al., 2019, Gorbey et al., 1999), econometric models (for example, Brücker and Siliverstovs, 2006, Cappelen et al., 2015), Bayesian hierarchical models (Azose & Raftery, 2015), and dedicated methods, for example for forecasting data structured by age (Raymer & Wiśniowski, 2018). In some cases, the methods additionally involve selection and combining forecasts through Bayesian model selection and averaging (Bijak, 2010 see also Section 2.5 and Section 2.6). Such models can be expected to produce reasonable forecasts (and errors) for up to a decade ahead (Bijak & Wiśniowski, 2010), although this depends on the migration flows being forecast, with some processes (e.g., family migration) more predictable than other (e.g., asylum). Another recognised problem with models using covariates is that those can be endogenous to migration (e.g., population) and also need predicting, which necessitates applying structured models to prevent uncertainty from exploding.

The methodological gaps and current work in migration forecasting concentrate in a few key areas, notably including causal (mechanistic) forecasting based on the process of migrant decision making (Willekens, 2018); as well as early warnings and ‘nowcasting’

of rapidly changing trends, for example in asylum migration ([Napierała, Hilton, Forster, Carammia, & Bijak, 2021](#)). In the context of early warnings, forays into data-driven methods for changepoint detection, possibly coupled with the digital trace and other high-frequency 'Big data', bear particular promise. At the same time, coherent uncertainty description across a range of time horizons, especially in the long range ([Azose & Raftery, 2015](#)), remains a challenge, which needs addressing for the sake of proper calibration of errors in the population forecasts, to which these migration components contribute.

3.6.6. Forecasting risk for violence and wars¹⁴²

Can we predict the occurrence of WW3 in the next 20 years? Is there any trend in the severity of wars?

The study of armed conflicts and atrocities, both in terms of frequency over time and the number of casualties, has received quite some attention in the scientific literature and the media (e.g., [Cederman, 2003](#), [Friedman, 2015](#), [Hayes, 2002](#), [Norton-Taylor, 2015](#), [Richardson, 1948](#), [Richardson, 1960](#)), falling within the broader discussion about violence ([Berlinski, 2009](#), [Goldstein, 2011](#), [Spagat et al., 2009](#)), with the final goal of understanding whether humanity is becoming less belligerent ([Pinker, 2011](#)), or not ([Braumoeller, 2019](#)).

Regarding wars and atrocities, the public debate has focused its attention on the so-called *Long Peace Theory* ([Gaddis, 1989](#)), according to which, after WW2, humanity has experienced the most peaceful period in history, with a decline in the number and in the severity of bloody events. Scholars like [Mueller, 2009a](#), [Mueller, 2009b](#) and [Pinker, 2011](#), [Pinker, 2018](#) claim that sociological arguments and all statistics suggest we live in better times, while others like [Gray, 2015a](#), [Gray, 2015b](#) and [Mann \(2018\)](#) maintain that those statistics are often partial and misused, the derived theories weak, and that war and violence are not declining but only being transformed. For Mann, the Long Peace proves to be ad-hoc, as it only deals with Western Europe and North America, neglecting the rest of the world, and the fact that countries like the US have been involved in many conflicts out of their territories after WW2.

Recent statistical analyses confirm Gray's and Mann's views: empirical data do not support the idea of a decline in human belligerence (no clear trend appears), and in its severity. Armed conflicts show long inter-arrival times, therefore a relative peace of a few decades means nothing statistically ([Cirillo & Taleb, 2016b](#)). Moreover, the

distribution of war casualties is extremely fat-tailed (Clauset, 2018, Clauset and Gleditsch, 2018), often with a tail exponent (Cirillo & Taleb, 2016b), indicating a possibly infinite mean, i.e., a tremendously erratic and unforeseeable phenomenon (see Section 2.3.22). An only apparently infinite-mean phenomenon though (Cirillo & Taleb, 2019), because no single war can kill more than the entire world population, therefore a finite upper bound exists, and all moments are necessarily finite, even if difficult to estimate. Extreme value theory (Embrechts et al., 2013) can thus be used to correctly model tail risk and make prudential forecasts (with many caveats like in Scharpf, Schneider, Nöh, & Clauset, 2014), while avoiding naive extrapolations (Taleb et al., 2020).

As history teaches (Nye, 1990), humanity has already experienced periods of relative regional peace, like the famous Paces Romana and Sinica. The present Pax Americana is not enough to claim that we are structurally living in a more peaceful era. The Long Peace risks to be another apophenia, another example of Texan sharpshooter fallacy (Carroll, 2003).

Similar mistakes have been made in the past. Buckle (1858) wrote: “that [war] is, in the progress of society, steadily declining, must be evident, even to the most hasty reader of European history. If we compare one country with another, we shall find that for a very long period wars have been becoming less frequent; and now so clearly is the movement marked, that, until the late commencement of hostilities, we had remained at peace for nearly forty years: a circumstance unparalleled [...] in the affairs of the world”. Sadly, Buckle was victim of the illusion coming from the Pax Britannica (Johnston, 2008): the century following his prose turned out to be the most murderous in human history.

3.7. Systems and humans

3.7.1. Support systems¹⁴³

Forecasting in businesses is a complicated procedure, especially when predicting numerous, diverse series (see Section 2.7.4), dealing with unstructured data of multiple sources (see Section 2.7.1), and incorporating human judgment (Lim & O’Connor, 1996a but also Section 2.11). In this respect, since the early 80’s, various Forecasting Support Systems (FSSs) have been developed to facilitate forecasting and support decision making (Kusters, McCullough, & Bell, 2006). Rycroft (1993) provides an early

comparative review of such systems, while many studies strongly support their utilisation over other forecasting alternatives (Sanders and Manrodt, 2003, Tashman and Leach, 1991).

In a typical use-case scenario, the FSSs will retrieve the data required for producing the forecasts, will provide some visualisations and summary statistics to the user, allow for data pre-processing, and then produce forecasts that may be adjusted according to the preferences of the user. However, according to Ord and Fildes (2013), effective FSS should be able to produce forecasts by combining relevant information, analytical models, judgment, visualisations, and feedback. To that end, FSSs must (i) elaborate accurate, efficient, and automatic statistical forecasting methods, (ii) enable users to effectively incorporate their judgment, (iii) allow the users to track and interact with the whole forecasting procedure, and (iv) be easily customised based on the context of the company.

Indeed, nowadays, most off-the-self solutions, such as SAP, SAS, JDEdwards, and ForecastPro, offer a variety of both standard and advanced statistical forecasting methods (see Section 2.3), as well as data pre-processing (see Section 2.2) and performance evaluation algorithms (see Section 2.12). On the other hand, many of them still struggle to incorporate state-of-the-art methods that can further improve forecasting accuracy, such as automatic model selection algorithms and temporal aggregation (see also Section 2.10.2), thus limiting the options of the users (Petropoulos, 2015). Similarly, although many FSSs support judgmental forecasts (see Section 2.11.1) and judgmental adjustments of statistical forecasts (see Section 2.11.2), this is not done as suggested by the literature, i.e., in a guided way under a well-organised framework. As a result, the capabilities of the users are restrained and methods that could be used to mitigate biases, overshooting, anchoring, and unreasonable or insignificant changes that do not rationalise the time wasted, are largely ignored (Fildes and Goodwin, 2013, Fildes et al., 2006).

Other practical issues of FSSs are related with their engine and interfaces which are typically designed so that they are generic and capable to serve different companies and organisations of diverse needs (Kusters et al., 2006). From a developing and economic perspective, this is a reasonable choice. However, the lack of flexibility and customisability can lead to interfaces with needless options, models, tools, and features that may confuse inexperienced users and undermine their performance (Fildes et al.,

2006). Thus, simple, yet exhaustive interfaces should be designed in the future to better serve the needs of each company and fit its particular requirements (Spiliotis, Raptis, & Assimakopoulos, 2015). Ideally, the interfaces should be adapted to the strengths and weaknesses of the user, providing useful feedback when possible (Goodwin, Fildes, Lawrence, & Nikolopoulos, 2007). Finally, web-based FSSs could replace windows-based ones that are locally installed and therefore of limited accessibility, availability, and compatibility (Asimakopoulos & Dix, 2013). Cloud computing and web-services could be exploited in that direction.

3.7.2. Cloud resource capacity forecasting¹⁴⁴

One of the central promises in cloud computing is that of elasticity. Customers of cloud computing services can add compute resources in real-time to meet and satisfy increasing demand and, when demand for a cloud-hosted application goes down, it is possible for cloud computing customers to down-scale. The benefit of the latter is particularly economically interesting during the current pandemic. Popular recent cloud computing offerings take this elasticity concept one step further. They abstract away the computational resources completely from developers, so that developers can build serverless applications. In order for this to work, the cloud provider handles the addition and removal of compute resources “behind the scenes”.

To keep the promise of elasticity, a cloud provider must address a number of forecasting problems at varying scales along the operational, tactical and strategic problem dimensions (Januschowski & Kolassa, 2019). As an example for a strategic forecasting problems: where should data centres be placed? In what region of a country and in what geographic region? As an example for tactical forecasting problems, these must take into account energy prices (see Section 3.4.2) and also, classic supply chain problems (Larson, Simchi-Levi, Kaminsky, & Simchi-Levi, 2001). After all, physical servers and data centres are what enables the cloud and these must be ordered and have a lead-time. The careful incorporation of life cycles of compute types is important (e.g., both the popularity of certain machine types and the duration of a hard disk). Analogous to the retail sector, cloud resource providers have tactical cold-start forecasting problems. For example, while GPU or TPU instances are still relatively recent but already well established, the demand for quantum computing is still to be decided. In the class of operational forecasting problems, cloud provider can choose to address short-term resource forecasting problems for applications such as adding

resources to applications predictively and make this available to customers (Barr, 2018). The forecasting of the customer's spend for cloud computing is another example. For serverless infrastructure, a number of servers is often maintained in a ready state (Gias & Casale, 2020) and the forecasting of the size of this 'warmpool' is another example. We note that cloud computing customers have forecasting problems that mirror the forecasting challenges of the cloud providers. Interestingly, forecasting itself has become a software service that cloud computing companies offer (Januschowski, Arpin et al., 2018, Liberty et al., 2020, Poccia, 2019)

Many challenges in this application area are not unique to cloud computing. Cold start problems exist elsewhere for example. What potentially stands out in cloud computing forecasting problems may be the scale (e.g., there are a lot of physical servers available), the demands on the response time and granularity of a forecast and the degree of automation. Consider the operational forecasting problem of predictive scaling. Unlike in retail demand forecasting, no human operator will be able to control this and response times to forecasts are in seconds. It will be interesting to see whether approaches based on reinforcement learning (Dempster et al., 2001, Gamble and Gao, 2018) can partially replace the need to have forecasting models (Januschowski, Gasthaus, Wang, Rangapuram & Callot, 2018).

3.7.3. Judgmental forecasting in practice¹⁴⁵

Surveys of forecasting practice (De Baets, 2019) have shown that the use of pure judgmental forecasting by practitioners has become less common. About 40 years ago, Sparkes and McHugh (1984) found that company action was more likely to be influenced by judgmental forecasts than by any other type of forecast. In contrast, Fildes and Petropoulos (2015) found that only 15.6% of forecasts in the surveyed companies were made by judgment alone. The majority of forecasts (55.6%) were made using a combination of statistical and judgmental methods. In this section, we discuss forecasting using unaided judgment (pure judgmental forecasting; see also Section 2.11.1), judgmental adjustments (judgment in combination with statistical models; see also Section 2.11.2), and the role of judgment in forecasting support systems.

On the first theme, the survey results discussed above beg the question of whether pure judgmental forecasting is still relevant and reliable. Answers here depend on the type of information on which the judgmental forecasts are based (Harvey, 2007 see also

Section 2.11.1). For instance, people have difficulty making cross-series forecasts, as they have difficulty learning the correlation between variables and using it to make their forecasts (Harvey et al., 1994, Lim and O'Connor, 1996b, Lim and O'Connor, 1996c). Additionally, they appear to take account of the noise as well as the pattern when learning the relation between variables; hence, when later using one of the variables to forecast the other, they add noise to their forecasts (Gray, Barnes, & Wilkinson, 1965). Judgmental extrapolation from a single time series is subject to various effects. First, people are influenced by optimism. For example, they over-forecast time series labelled as 'profits' but under-forecast the same series labelled as 'losses' (Harvey & Reimers, 2013). Second, they add noise to their forecasts so that a sequence of forecasts looks similar to ('represents') the data series (Harvey, 1995). Third, they damp trends in the data (Eggleton, 1982, Harvey and Reimers, 2013, Lawrence and Makridakis, 1989). Fourth, forecasts from un-trended independent series do not lie on the series mean but between the last data point and the mean; this is what we would expect if people perceived a positive autocorrelation in the series (Reimers & Harvey, 2011). These last two effects can be explained in terms of the under-adjustment that characterises use of the anchor-and-adjust heuristic: forecasters anchor on the last data point and adjust towards the trend line or mean – but do so insufficiently. However, in practice, this under-adjustment may be appropriate because real linear trends do become damped and real series are more likely to contain a modest autocorrelation than be independent (Harvey, 2011). We should therefore be reluctant to characterise these last two effects as biases.

Given these inherent flaws in people's decision making, practitioners might be hesitant to base their predictions on judgment. However, the reality is that companies persist in incorporating judgment into their forecasting. Assertions that they are wrong to do so represent an over-simplified view of the reality in which businesses operate. Statistical models are generally not able to account for external events, events with low frequency, or a patchy and insufficient data history (Armstrong and Collopy, 1998, Goodwin, 2002, Hughes, 2001). Hence, a balance may be found in the combination of statistical models and judgment (see Section 2.11.2).

In this respect, judgmental adjustments to statistical model outputs are the most frequent form of judgmental forecasting in practice (Arvan et al., 2019, Eksoz et al., 2019, Lawrence et al., 2006, Petropoulos et al., 2016). Judgmental adjustments give

practitioners a quick and convenient way to incorporate their insights, their experience and the additional information that they possess into a set of statistical baseline forecasts. Interestingly, [Fildes et al. \(2009\)](#) examined the judgmental adjustment applications in four large supply-chain companies and found evidence that the adjustments in a 'negative' direction improved the accuracy more than the adjustments in a 'positive' direction. This effect may be attributable to wishful thinking or optimism that may underlie positive adjustments. Adjustments that were 'larger' in magnitude were also more beneficial in terms of the final forecast accuracy than 'smaller' adjustments ([Fildes et al., 2009](#)). This may simply be because smaller adjustments are merely a sign of tweaking the numbers, but large adjustments are carried out when there is a highly valid reason to make them. These findings have been confirmed in other studies (see, for example, [Franses and Legerstee, 2009b](#), [Syntetos et al., 2009](#)). What are the main reasons behind judgmental adjustments? [Önkal and Gönül \(2005\)](#) conducted a series of interviews and a survey on forecasting practitioners ([Gönül, Önkal, & Goodwin, 2009](#)) to explore these. The main reasons given were (i) to incorporate the practitioners' intuition and experience about the predictions generated externally, (ii) to accommodate sporadic events and exceptional occasions, (iii) to integrate confidential/insider information that may have not been captured in the forecasts, (iv) to hold responsibility and to gain control of the forecasting process, (v) to incorporate the expectations and viewpoints of the practitioners, and (vi) to compensate for various judgmental biases that are believed to exist in the predictions. These studies also revealed that forecasting practitioners are very fond of judgmental adjustments and perceive them as a prominent way of 'completing' and 'owning' the predictions that are generated by others. While the first three reasons represent the integration of an un-modelled component into the forecast, potentially improving accuracy, the other reasons tend to harm accuracy rather than improve it. In such cases, the forecast would be better off if left unadjusted. [Önkal and Gönül \(2005\)](#) and [Gönül et al. \(2009\)](#) report that the occasions when forecasters refrain from adjustments are (i) when the practitioners are adequately informed and knowledgeable about the forecasting method(s) that are used to generate the baseline forecasts, (ii) when there are accompanying explanations and convincing communications that provide the rationale behind forecast method selection, (iii) when baseline predictions are supplemented by additional supportive materials such as

scenarios and alternative forecasts, (*iv*) when the forecasting source is believed to be trustworthy and reliable, and (*v*) when organisational policy or culture prohibits judgmental adjustments. In these circumstances, the baseline forecasts are more easily accepted by practitioners and their adjustments tend to be less frequent.

Ideally, a Forecast Support System (FSS; see Section 3.7.1) should be designed to ensure that it encourages adjustment or non-adjustment whichever is appropriate (Fildes et al., 2006). But how can this be achieved? The perceived quality and accessibility of a FSS can be influenced by its design. More on this can be found in the literature on the Technology Acceptance Model (Davis, Bagozzi, & Warshaw, 1989) and decision making (for instance, by means of framing, visual presentation or nudging; e.g., Gigerenzer, 1996, Kahneman and Tversky, 1996, Payne, 1982, Thaler and Sunstein, 2009). A number of studies have investigated the design aspects of FSS, with varying success. One of the more straightforward approaches is to change the look and feel of the FSS as well as its presentation style. Harvey and Bolger (1996) found that trends were more easily discernible when the data was displayed in graphical rather than tabular format. Additionally, simple variations in presentation such as line graphs versus point graphs can alter accuracy (Theocharis, Smith, & Harvey, 2018). The functionalities of the FSS can also be modified (see Section 2.11.2). Goodwin (2000b) investigated three ways of improving judgmental adjustment via changes in the FSS: a 'no adjustment' default, requesting forecasters specify the size of an adjustment rather than give a revised forecast, and requiring a mandatory explanation for the adjustment. Only the default option and the explanation feature were successful in increasing the acceptance of the statistical forecast and so improving forecast accuracy. Goodwin et al. (2011) reported an experiment that investigated the effects of (*i*) 'guidance' in the form of providing information about when to make adjustments and (*ii*) 'restriction' of what the forecaster could do (e.g., prohibiting small adjustments). They found that neither restrictiveness nor guidance was successful in improving accuracy, and both were met with resistance by the forecasters. While these studies focused on voluntary integration, Goodwin, 2000a, Goodwin, 2002 examined the effectiveness of various methods of mechanical integration and concluded that automatic correction for judgmental biases by the FSS was more effective than combining judgmental and statistical inputs automatically with equal or varying weights. Another approach to mechanical integration was investigated by Baecke et al.

(2017). They compared ordinary judgmental adjustment with what they termed “integrative judgment”. This takes the judgmental information into account as a predictive variable in the forecasting model and generates a new forecast. This approach improved accuracy. It also had the advantage that forecasters still had their input into the forecasting process and so the resistance found by [Goodwin et al. \(2011\)](#) should not occur. Finally, it is worth emphasising that an effective FSS should not only improve forecast accuracy but should also be easy to use, understandable, and acceptable ([Fildes et al., 2006](#) see also Section [2.11.6](#) and Section [3.7.1](#)).

3.7.4. Trust in forecasts¹⁴⁶

Regardless of how much effort is poured into training forecasters and developing elaborate forecast support systems, decision-makers will either modify or discard the predictions if they do not trust them (see also Section [2.11.2](#), Section [2.11.6](#), Section [3.7.1](#), and Section [3.7.3](#)). Hence, trust is essential for forecasts to be actually used in making decisions ([Alvarado-Valencia and Barrero, 2014](#), [Önkal et al., 2019](#)). Given that trust appears to be the most important attribute that promotes a forecast, what does it mean to practitioners? Past work suggests that trusting a forecast is often equated with trusting the forecaster, their expertise and skills so that predictions could be used without adjustment to make decisions ([Önkal et al., 2019](#)). It is argued that trust entails relying on credible forecasters that make the best use of available information while using correctly applied methods and realistic assumptions ([Gönül et al., 2009](#)) with no hidden agendas ([Gönül, Önkal, & Goodwin, 2012](#)). Research suggests that trust is not only about trusting forecaster’s competence; users also need to be convinced that no manipulations are made for personal gains and/or to mislead decisions ([Twyman, Harvey, & Harries, 2008](#)).

Surveys with practitioners show that key determinants of trust revolve around (i) forecast support features and tools (e.g., graphical illustrations, rationale for forecasts), (ii) forecaster competence/credibility, (iii) forecast combinations (from multiple forecasters/methods), and (iv) forecast user’s knowledge of forecasting methods ([Önkal et al., 2019](#)).

What can be done to enhance trust? If trust translates into accepting guidance for the future while acknowledging and tolerating potential forecast errors, then both the providers and users of forecasts need to work as partners towards shared goals and expectations. Important pathways to accomplish this include (i) honest communication

of forecaster's track record and relevant accuracy targets (Önkal et al., 2019), (ii) knowledge sharing (Özer et al., 2011, Renzl, 2008) and transparency of forecasting methods, assumptions and data (Önkal et al., 2019), (iii) communicating forecasts in the correct tone and jargon-free language to appeal to the user audience (Taylor & Thomas, 1982), (iv) users to be supported with forecasting training (Merrick, Hardin, & Walker, 2006), (v) providing explanations/rationale behind forecasts (Gönül et al., 2006, Önkal et al., 2008), (vi) presenting alternative forecasts under different scenarios (see Section 2.11.5), and (vii) giving combined forecasts as benchmarks (Önkal et al., 2019). Trust must be earned and deserved (Maister et al., 2012) and is based on building a relationship that benefits both the providers and users of forecasts. Take-aways for those who make forecasts and those who use them converge around clarity of communication as well as perceptions of competence and integrity. Key challenges for forecasters are to successfully engage with users throughout the forecasting process (rather than relying on a forecast statement at the end) and to convince them of their objectivity and expertise. In parallel, forecast users face challenges in openly communicating their expectations from forecasts (Gönül et al., 2009), as well as their needs for explanations and other informational addendum to gauge the uncertainties surrounding the forecasts. Organisational challenges include investing in forecast management and designing resilient systems for collaborative forecasting.

3.7.5. Communicating forecast uncertainty¹⁴⁷

Communicating forecast uncertainty is a critical issue in forecasting practice. Effective communication allows forecasters to influence end-users to respond appropriately to forecasted uncertainties. Some frameworks for effective communication have been proposed by decomposing the communication process into its elements: the communicator, object of uncertainty, expression format, audience, and its effect (National Research Council, 2006, van der Bles et al., 2019).

Forecasters have long studied part of this problem focusing mostly in the manner by which we express forecast uncertainties. Gneiting and Katzfuss (2014) provides a review of recent probabilistic forecasting methods (see also Section 2.12.4 and Section 2.12.5). Forecasting practice however revealed that numeracy skills and cognitive load can often inhibit end-users from correctly interpreting these uncertainties (Joslyn and Nichols, 2009, Raftery, 2016). Attempts to improve understanding through the use of less technical vocabulary also creates new challenges. Research in psychology show that

wording and verbal representation play important roles in disseminating uncertainty (Joslyn, Nadav-Greenberg, Taing, & Nichols, 2009). Generally forecasters are found to be consistent in their use of terminology, but forecast end-users often have inconsistent interpretation of these terms even those commonly used (Budescu and Wallsten, 1985, Clark, 1990, Ülkümen et al., 2016). Pretesting verbal expressions and avoiding commonly misinterpreted terms are some easy ways to significantly reduce biases and improve comprehension.

Visualisations can also be powerful in communicating uncertainty. Johnson and Slovic (1995) and Spiegelhalter, Pearson, and Short (2011) propose several suggestions for effective communication (e.g., multiple-format use, avoiding framing bias, and acknowledging limitations), but also recognise the limited amount of existing empirical evidence. Some domain-specific studies do exist. For example, Riveiro, Helldin, Falkman, and Lebram (2014) showed uncertainty visualisation helped forecast comprehension in a homeland security context.

With respect to the forecaster and her audience, issues such as competence, trust, respect, and optimism have been recently examined as a means to improve uncertainty communication. Fiske and Dupree (2014) discusses how forecast recipients often infer apparent intent and competence from the uncertainty provided and use these to judge trust and respect (see also Sections 2.11.6 [Trusting model and expert forecasts](#), 3.7.4 [Trust in forecasts](#) for discussion on trust and forecasting). This suggests that the amount of uncertainty information provided should be audience dependent (Han et al., 2009, Politi et al., 2007). Raftery (2016) acknowledges this by using strategies depending on the audience type (e.g., low-stakes user, risk avoider, etc.). Fischhoff and Davis (2014) suggests a similar approach by examining how people are likely to use the information (e.g., finding a signal, generating new options, etc.)

When dealing with the public, experts assert that communicating uncertainty helps users understand forecasts better and avoid a false sense of certainty (Morss, Demuth, & Lazo, 2008). Research however shows that hesitation to include forecast uncertainty exists among experts because it provides an opening for criticism and the possibility of misinterpretation by the public (Fischhoff, 2012). This is more challenging when the public has prior beliefs on a topic or trust has not been established. Uncertainty can be used by individuals to reinforce a motivated-reasoning bias that allows them to “see what they want to see” (Dieckmann, Gregory, Peters, & Hartman, 2017). Recent work

however suggests that increasing transparency for uncertainty does not necessarily affect trust in some settings. [van der Bles, van der Linden, Freeman, and Spiegelhalter \(2020\)](#) recently showed in a series of experiments that people recognise greater uncertainty with more information but expressed only a small decrease in trust in the report and trustworthiness of the source.

3.8. Other applications

3.8.1. Tourism demand forecasting¹⁴⁸

As seen throughout 2020, (leisure) tourism demand is very sensitive to external shocks such as natural and human-made disasters, making tourism products and services extremely perishable ([Frechtling, 2001](#)). As the majority of business decisions in the tourism industry require reliable demand forecasts ([Song, Witt, & Li, 2008](#)), improving their accuracy has continuously been on the agenda of tourism researchers and practitioners alike. This continuous interest has resulted in two tourism demand forecasting competitions to date ([Athanasopoulos et al., 2011](#), [Song and Li, 2021](#)), the current one with a particular focus on tourism demand forecasting during the COVID-19 pandemic (for forecasting competitions, see Section 2.12.7). Depending on data availability, as well as on geographical aggregation level, tourism demand is typically measured in terms of arrivals, bed-nights, visitors, export receipts, import expenditures, etc.

Since there are no specific tourism demand forecast models, standard univariate and multivariate statistical models, including common aggregation and combination techniques, etc., have been used in quantitative tourism demand forecasting (see, for example, [Jiao and Chen, 2019](#), [Song et al., 2019](#) for recent reviews). Machine learning and other artificial intelligence methods, as well as hybrids of statistical and machine learning models, have recently been employed more frequently.

Traditionally, typical micro-economic demand drivers (own price, competitors' prices, and income) and some more tourism-specific demand drivers (source-market population, marketing expenditures, consumer tastes, habit persistence, and dummy variables capturing one-off events or qualitative characteristics) have been employed as predictors in tourism demand forecasting ([Song et al., 2008](#)). One caveat of some of

these economic demand drivers is their publication lag and their low frequency, for instance, when real GDP (per capita) is employed as a proxy for travellers' income. The use of leading indicators, such as industrial production as a leading indicator for real GDP (see also Section 3.3.2), has been proposed for short-term tourism demand forecasting and nowcasting (Chatziantoniou, Degiannakis, Eeckels, & Filis, 2016). During the past couple of years, web-based leading indicators have also been employed in tourism demand forecasting and have, in general, shown improvement in terms of forecast accuracy. However, this has not happened in each and every case, thereby confirming the traded wisdom that there is no single best tourism demand forecasting approach (Li, Song, & Witt, 2005). Examples of those web-based leading indicators include Google Trends indices (Bangwayo-Skeete & Skeete, 2015), Google Analytics indicators (Gunter & Önder, 2016), as well as Facebook 'likes' (Gunter, Önder, & Gindl, 2019).

The reason why these expressions of interaction of users with the Internet have proven worthwhile as predictors in a large number of cases is that it is sensible to assume potential travellers gather information about their destination of interest prior to the actual trip, with the Internet being characterised by comparably low search costs, ergo allowing potential travellers to forage information (Pirolli & Card, 1999) with only little effort (Zipf, 2016). A forecaster should include this information in their own set of relevant information at the forecast origin Lütkepohl (2005), if taking it into account results in an improved forecast accuracy, with web-based leading indicators thus effectively Granger-causing (Granger, 1969) actual tourism demand (see Section 2.5.1). Naturally, tourism demand forecasting is closely related to aviation forecasting (see Section 3.8.2), as well as traffic flow forecasting (see Section 3.8.3). A sub-discipline of tourism demand forecasting can be found with hotel room demand forecasting. The aforementioned perishability of tourism products and services is particularly evident for hotels as a hotel room not sold is lost revenue that cannot be regenerated. Accurate hotel room demand forecasts are crucial for successful hotel revenue management (Pereira, 2016) and are relevant for planning purposes such as adequate staffing during MICE (i.e., Meetings, Incentives, Conventions, and Exhibitions/Events) times, scheduling of renovation periods during low seasons, or balancing out overbookings and "no shows" given constrained hotel room supply (Ivanov & Zhechev, 2012).

Particularly since the onset of the COVID-19 pandemic in 2020, which has been characterised by global travel restrictions and tourism businesses being locked down to varying extents, scenario forecasting and other forms of hybrid and judgmental forecasting played an important role (Zhang, Song, Wen, & Liu, 2021 see Section 2.11.5), thereby highlighting an important limitation of quantitative tourism demand forecasting as currently practised. Based on the rapid development of information technology and artificial intelligence, Li and Jiao (2020), however, envisage a “super-smart tourism forecasting system” (Li & Jiao, 2020 p. 264) for the upcoming 75 years of tourism demand forecasting. According to these authors, this system will be able to automatically produce forecasts at the micro level (i.e., for the individual traveller and tourism business) in real time while drawing on a multitude of data sources and integrating multiple (self-developing) forecast models.

3.8.2. Forecasting for aviation¹⁴⁹

Airports and airlines have long invested in forecasting arrivals and departures of aircrafts. These forecasts are important in measuring airspace and airport congestions, designing flight schedules, and planning for the assignment of stands and gates (Barnhart & Cohn, 2004). Various techniques have been applied to forecast aircrafts’ arrivals and departures. For instance, Rebollo and Balakrishnan (2014) apply random forests to predict air traffic delays of the National Airspace System using both temporal and network delay states as covariates. Manna, Biswas, Kundu, Rakshit, Gupta, and Barman (2017) develop a statistical model based on a gradient boosting decision tree to predict arrival and departure delays, using the data taken from the United States Department of Transportation (Bureau of Transportation Statistics, 2020). Rodríguez-Sanz, Comendador, Valdés, Pérez-Castán, Montes, and Serrano (2019) develop a Bayesian Network model to predict flight arrivals and delays using the radar data, aircraft historical performance and local environmental data. There are also a few studies that have focused on generating probabilistic forecasts of arrivals and departures, moving beyond point estimates. For example, Tu, Ball, and Jank (2008) develop a predictive system for estimating flight departure delay distributions using flight data from Denver International Airport. The system employs the smoothing spline method to model seasonal trends and daily propagation patterns. It also uses mixture distributions to estimate the residual errors for predicting the entire distribution.

In the airline industry, accurate forecasts on demand and booking cancellations are crucial to revenue management, a concept that was mainly inspired by the airline and hotel industries (Lee, 1990, McGill and Van Ryzin, 1999 see also Section 3.8.1 for a discussion on hotel occupancy forecasting). The proposals of forecasting models for flight demand can be traced back to Beckmann and Bobkoski (1958), where these authors demonstrate that Poisson and Gamma models can be applied to fit airline data. Then, the use of similar flights' short-term booking information in forecasting potential future bookings has been discussed by airline practitioners such as Adams and Michael (1987) at Quantas as well as Smith, Leimkuhler, and Darrow (1992) at American Airlines. Regressions models (see Section 2.3.2) and time series models such as exponential smoothing (see Section 2.3.1) and ARIMA (see Section 2.3.4) have been discussed in Botimer, 1997, Sa, 1987, and Wickham (1995). There are also studies focusing on disaggregate airline demand forecasting. For example, Martinez and Sanchez (1970) apply empirical probability distributions to predict bookings and cancellations of individual passengers travelling with Iberia Airlines. Carson, Cenesizoglu, and Parker (2011) show that aggregating the forecasts of individual airports using airport-specific data could provide better forecasts at a national level. More recently, machine learning methods have also been introduced to generate forecasts for airlines. This can be seen in Weatherford, Gentry, and Wilamowski (2003) where they apply neural networks to forecast the time series of the number of reservations. Moreover, Hopman, Koole, and van der Mei (2021) show that an extreme gradient boosting model which forecasts itinerary-based bookings using ticket price, social media posts and airline reviews outperforms traditional time series forecasts. Forecasting passenger arrivals and delays in the airports have received also some attention in the literature, particularly in the past decade. Wei and Hansen (2006) build an aggregate demand model for air passenger traffic in a hub-and-spoke network. The model is a log-linear regression that uses airline service variables such as aircraft size and flight distance as predictors. Barnhart, Fearing, and Vaze (2014) develop a multinomial logit regression model, designed to predict delays of US domestic passengers. Their study also uses data from the US Department of Transportation (Bureau of Transportation Statistics, 2020). Guo, Grushka-Cockayne, and De Reyck (2020) recently develop a predictive system that generates distributional forecasts of connection times for transfer passengers at an airport, as well as passenger

flows at the immigration and security areas. Their approach is based on the application of regression trees combined with copula-based simulations. This predictive system has been implemented at Heathrow airport since 2017.

With an increasing amount of available data that is associated with activities in the aviation industry, predictive analyses and forecasting methods face new challenges as well as opportunities, especially in regard to updating forecasts in real time. The predictive system developed by [Guo et al. \(2020\)](#) is able to generate accurate forecasts using real-time flight and passenger information on a rolling basis. The parameters of their model, however, do not update over time. Therefore, a key challenge in this area is for future studies to identify an efficient way to dynamically update model parameters in real time.

3.8.3. Traffic flow forecasting¹⁵⁰

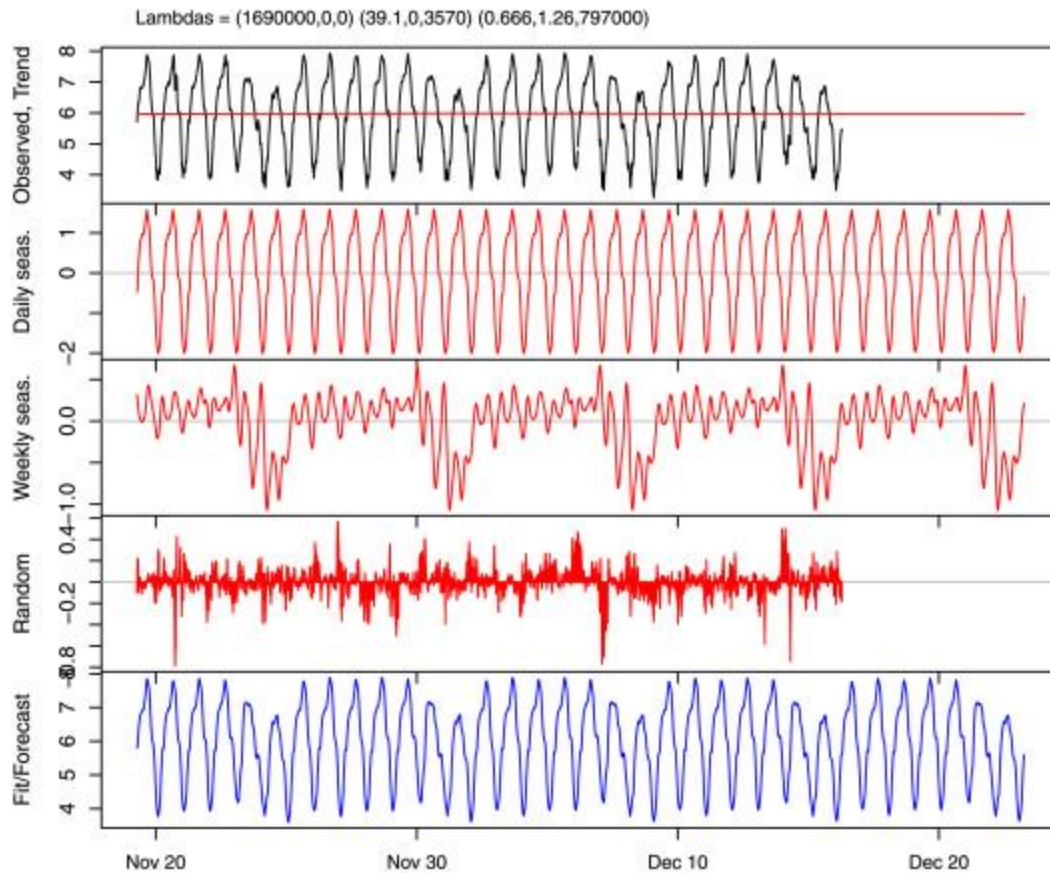
Traffic flow forecasting is an important task for traffic management bodies to reduce traffic congestion, perform planning and allocation tasks, as well as for travelling individuals to plan their trips. Traffic flow is complex spatial and time-series data exhibiting multiple seasonalities and affected by spatial exogenous influences such as social and economic activities and events, various government regulations, planned road works, weather, traffic accidents, etc. ([Polson & Sokolov, 2017](#)).

Methods to solve traffic flow forecasting problems vaguely fall into three categories. The first uses parametric statistical methods such as ARIMA, seasonal ARIMA, space–time ARIMA, Kalman filters, etc. (see, for example, [Kamarianakis and Prastacos, 2005](#), [Vlahogianni et al., 2004](#), [Vlahogianni et al., 2014](#), [Whittaker et al., 1997](#)). The second set of approaches uses purely of neural networks ([Mena-Oreja & Gozalvez, 2020](#)). The third group of methods uses various machine learning, statistical non-parametric techniques or mixture of them (see, for example, [Hong, 2011](#), [Zhang et al., 2017](#), [Zhang et al., 2016](#) but also Section 2.7.8 and Section 2.7.10 for an overview of NN and ML methods).

Although neural networks are probably the most promising technique for traffic flow forecasting (see, for example, [Do et al., 2019](#), [Polson and Sokolov, 2017](#)), statistical techniques, such as Seasonal-Trend decomposition based on Regression (STR, see Section 2.2.2), can outperform when little data is available or they can be used for imputation, de-noising, and other pre-processing before feeding data into neural

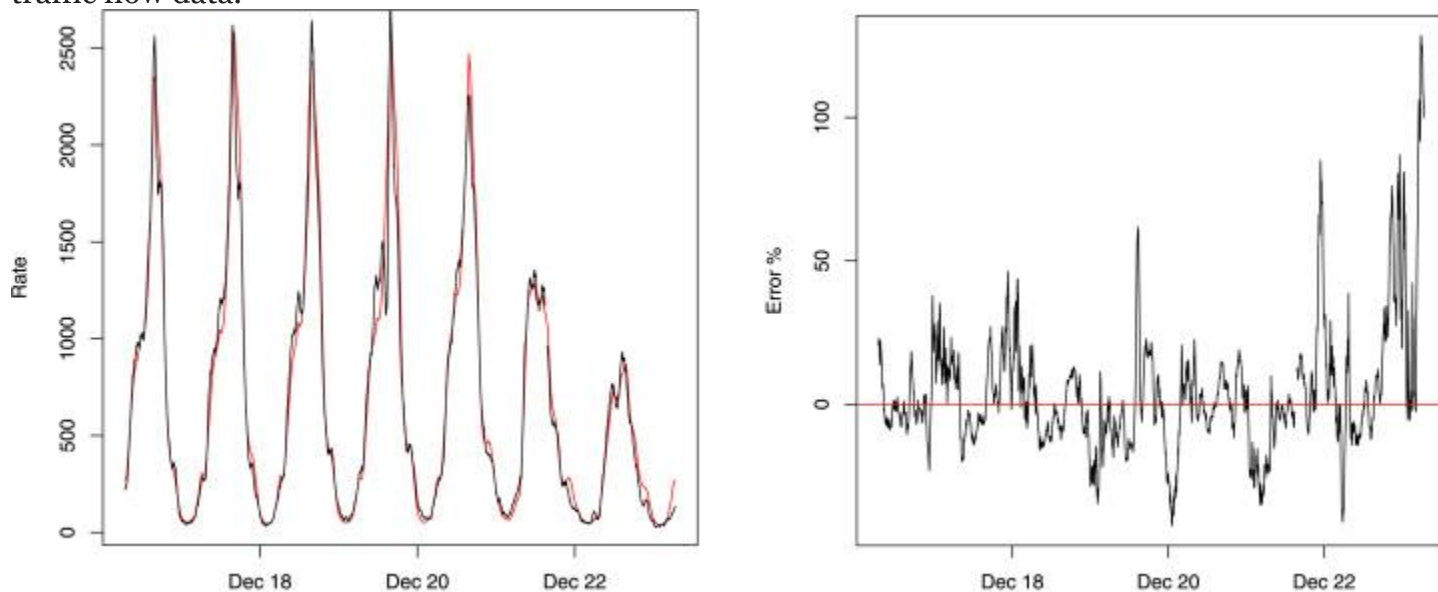
networks which often become less powerful when working with missing or very noisy data.

Traffic flow forecasting is illustrated below using vehicle flow rate data from road camera A1.GT.24538 on A1 highway in Luxembourg ([des Mobilités, 2020](#)) from 2019-11-19 06:44:00 UTC to 2019-12-23 06:44:00 UTC. Most of the data points are separated by 5 min intervals. Discarding points which do not follow this schedule leads to a data set where all data points are separated by 5 min intervals, although values at some points are missing. The data is split into training and test sets by setting aside last 7 days of data. As [Hou, Edara, and Sun \(2014\)](#) and [Polson and Sokolov \(2017\)](#) suggest, spatial factors are less important for long term traffic flow forecasting, and therefore they are not taken into account and only temporal data is used. Application of STR ([Dokumentov, 2017](#)) as a forecasting technique to the log transformed data leads to a forecast with Mean Squared Error 102.4, Mean Absolute Error 62.8, and Mean Absolute Percentage Error (MAPE) 14.3% over the test set, outperforming Double-Seasonal Holt-Winters by 44% in terms of MAPE. The decomposition and the forecast obtained by STR are shown on [Fig. 15](#) and the magnified forecast and the forecasting errors are on [Fig. 16](#).



1. [Download: Download high-res image \(983KB\)](#)
2. [Download: Download full-size image](#)

Fig. 15. STR decomposition of the log transformed training data and the forecasts for the traffic flow data.



1. [Download: Download high-res image \(341KB\)](#)

2. [Download: Download full-size image](#)

Fig. 16. Left: forecast (red) and the test data (black); Right: the prediction error over time for the traffic flow data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.8.4. Call arrival forecasting¹⁵¹

Forecasting of inbound call arrivals for call centres supports a number of key decisions primarily around staffing (Aksin, Armony, & Mehrotra, 2007). This typically involves matching staffing level requirements to service demand as summarised in Fig. 17. To achieve service level objectives, an understanding of the call load is required in terms of the call arrivals (Gans, Koole, & Mandelbaum, 2003). As such, forecasting of future call volume or call arrival rates is an important part of call centre management.

There are several properties to call arrival data. Depending on the level of aggregation and the frequency with which data is collected, e.g., hourly, call arrival data may exhibit intraday (within-day), intraweek, and intrayear multiple seasonal patterns (Avramidis et al., 2004, Brown et al., 2005a and Section 2.3.5). In addition, arrival data may also exhibit interday and intraday dependencies, with different time periods within the same day, or across days within the same week, showing strong levels of autocorrelation (Brown et al., 2005a, Shen and Huang, 2005, Tanir and Booth, 1999). Call arrivals may also be heteroscedastic with variance at least proportional to arrival counts (Taylor, 2008), and overdispersed under a Poisson assumption having variance per time period typically much larger than its expected value (Avramidis et al., 2004, Jongbloed and Koole, 2001, Steckley et al., 2005). These properties have implications for various approaches to modelling and forecasting call arrivals.



1. [Download: Download high-res image \(71KB\)](#)
2. [Download: Download full-size image](#)

Fig. 17. The staffing decision process in call centres.

The first family of methods are time series methods requiring no distributional assumptions. Early studies employed auto regressive moving average (ARMA; see Section 2.3.4) models (Andrews and Cunningham, 1995, Antipov and Meade, 2002, Tandberg et al., 1995, Xu, 1999), exponential smoothing (Bianchi et al.,

1993, Bianchi et al., 1998 see Section 2.3.1), fast Fourier transforms (Lewis, Herbert, & Bell, 2003), and regression (Tych, Pedregal, Young, & Davies, 2002 see Section 2.3.2). The first methods capable of capturing multiple seasonality were evaluated by Taylor (2008) and included double seasonal exponential smoothing (Taylor, 2003b) and multiplicative double seasonal ARMA (SARMA). Since then several advanced time series methods have been developed and evaluated (De Livera et al., 2011, Taylor, 2010, Taylor and Snyder, 2012), including artificial neural networks (Li et al., 2011, Millán-Ruiz and Hidalgo, 2013, Pacheco et al., 2009) and models for density forecasting (Taylor, 2012).

Another family of models relies on the assumption of a time-inhomogeneous Poisson process adopting fixed (Brown et al., 2005a, Jongbloed and Koole, 2001, Shen and Huang, 2008a, Taylor, 2012) and mixed modelling (Aldor-Noiman et al., 2009, Avramidis et al., 2004, Ibrahim and L'Ecuyer, 2013) approaches to account for the overdispersed nature of the data, and in some cases, interday and intraday dependence.

The works by Soyer and Tarimcilar (2008) and Weinberg, Brown, and Stroud (2007) model call volumes from a Bayesian point of view. Other Bayesian inspired approaches have been adopted mainly for estimating various model parameters, but also allowing for intraday updates of forecasts (Aktekin and Soyer, 2011, Landon et al., 2010).

A further class of approach addresses the dimensionality challenge related to high frequency call data using Singular Value Decomposition (SVD). Shen and Huang (2005) and Shen and Huang (2008a) use the same technique to achieve dimensionality reduction of arrival data, and to create a forecasting model that provides both interday forecasts of call volume, and an intraday updating mechanism. Several further studies have extended the basic SVD approach to realise further modelling innovations, for example, to forecast call arrival rate profiles and generate smooth arrival rate curves (Shen, 2009, Shen and Huang, 2008b, Shen et al., 2007). A more comprehensive coverage of different forecasting approaches for call arrival rate and volume can be found in a recent review paper by Ibrahim, Ye, L'Ecuyer, and Shen (2016).

3.8.5. Elections forecasting¹⁵²

With the exception of weather forecasts, there are few forecasts which have as much public exposure as election forecasts. They are frequently published by mass media, with

their number and disclosure reaching a frenzy as the Election Day approaches. This explains the significant amount of methods, approaches and procedures proposed and the paramount role these forecasts play in shaping people's confidence in (soft/social) methods of forecasting.

The problem escalates because, regardless whether the goal of the election forecast is an attempt to ascertain the winner in two-choice elections (e.g., a referendum or a Presidential election) or to reach estimates within the margins of error in Parliamentary systems, the knowledge of the forecasts influences electors' choices (Pavía, Gil-Carceller, Rubio-Mataix, Coll, Alvarez-Jareño, Aybar, & Carrasco-Arroyo, 2019). Election forecasts not only affect voters but also political parties, campaign organizations and (international) investors, who are also watchful of their evolution.

Scientific approaches to election forecasting include polls, information (stock) markets and statistical models. They can also be sorted by when they are performed; and new methods, such as social media surveillance (see also Section 2.9.3), are also emerging (Ceron et al., 2016, Huberty, 2015). Probabilistic (representative) polls are the most commonly used instrument to gauge public opinions. The progressive higher impact of non-sampling errors (coverage issues, non-response bias, measurement error: Biemer, 2010) is, however, severely testing this approach. Despite this, as Kennedy, Wojcik, and Lazer (2017) show in a recent study covering 86 countries and more than 500 elections, polls are still powerful and robust predictors of election outcomes after adjustments (see also Jennings, Lewis-Beck, & Wlezien, 2020). The increasing need of post-sampling adjustments of probabilistic samples has led to a resurgence of interest in non-probabilistic polls (Elliott and Valliant, 2017, Pavía and Larraz, 2012, Wang, Rothschild et al., 2015), abandoned in favour of probabilistic sampling in 1936, when Gallup forecasted Roosevelt's triumph over Landon using a small representative sample despite Literacy Digest failing to do so with a sample of near 2.5 million responses (Squire, 1988).

A person knows far more than just her/his voting intention (Rothschild, 2009) and when s/he makes a bet, the rationality of her/his prediction is reinforced because s/he wants to win. Expectation polls try to exploit the first issue (Graefe, 2014), while prediction markets, as efficient aggregators of information, exploit both these issues to yield election forecasts (see also Sections 2.6.4 *The wisdom of crowds*, 2.11.4 *Panels of experts*). Several studies have proven the performance of these approaches (Berg et al.,

2008, Erikson and Wlezien, 2012, Williams and Reade, 2016, Wolfers and Zitzewitz, 2004), even studying their links with opinion polls (Brown, Reade, & Vaughan Williams, 2019). Practice has also developed econometric models (Fair, 1978) that exploit structural information available months before the election (e.g., the evolution of the economy or the incumbent popularity). Lewis-Beck has had great success in publishing dozens of papers using this approach (see, e.g., Lewis-Beck, 2005).

Special mention also goes to Election-Day forecasting strategies, which have been systematically commissioned since the 1950s (Mitofsky, 1991). Exit (and entrance) polls (Klofstad and Bishin, 2012, Pavía, 2010), quick-counts (Pavía-Miralles & Larraz-Iribas, 2008), and statistical models (Bernardo, 1984, Moshman, 1964, Pavía-Miralles, 2005) have been used to anticipate outcomes on Election Day. Some of these strategies (mainly random quick-counts) can be also employed as auditing tools to disclose manipulation and fraud in weak democracies (Scheuren & Alvey, 2008).

3.8.6. Sports forecasting¹⁵³

Forecasting is inherent to sport. Strategies employed by participants in sporting contests rely on forecasts, and the decision by promoters to promote, and consumers to attend such events are conditioned on forecasts: predictions of how interesting the event will be. First in this section, we look at forecast competitions in sport, and following this we consider the role forecasts play in sporting outcomes.

Forecast competitions are common; see Section 2.12.7. Sport provides a range of forecast competitions, perhaps most notably the competition between bookmakers and their customers – betting. A bet is a contingent contract, a contract whose payout is conditional on specified future events occurring. Bets occur fundamentally because two agents disagree about the likelihood of that event occurring, and hence it is a forecast. Bookmakers have been extensively analysed as forecasters; Forrest, Goddard, and Simmons (2005) evaluated biases in the forecasts implied by bookmaker odds over a period where the betting industry became more competitive, and found that relative to expert forecasts, bookmaker forecasts improved.

With the internet age, prediction markets have emerged, financial exchanges where willing participants can buy and sell contingent contracts. In theory, such decentralised market structures ought to provide the most efficient prices and hence efficient forecasts (Nordhaus, 1987). A range of papers have tested this in the sporting

context ([Angelini and De Angelis, 2019](#), [Croxson and Reade, 2014](#), [Gil and Levitt, 2007](#)), with conclusions tending towards a lack of efficiency.

Judgmental forecasts by experts are commonplace too (see also Section [2.11](#)); traditionally in newspapers, but more recently on television and online. [Reade, Singleton, and Brown \(2020\)](#) evaluate forecasts of scorelines from two such experts against bookmaker prices, a statistical model, and the forecasts from users of an online forecasting competition. [Singleton, Reade, and Brown \(2019\)](#) find that when forecasters in the same competition revise their forecasts, their forecast performance worsens. This forecasting competition is also analysed by [Butler, Butler, and Eakins \(2020\)](#) and [Reade et al. \(2020\)](#).

Sport is a spectacle, and its commercial success is conditioned on this fact. Hundreds of millions of people globally watch events like the Olympics and the FIFA World Cup – but such interest is conditioned on anticipation, a forecast that something interesting will happen. A superstar is going to be performing, the match will be a close encounter, or it will matter a lot for a bigger outcome (the championship, say). These are the central tenets of sport economics back to [Neale \(1964\)](#) and [Rottenberg \(1956\)](#), most fundamentally the ‘uncertainty of outcome hypothesis’. A multitude of sport attendance prediction studies investigate this (see, for example, [Coates and Humphreys, 2010](#), [Forrest and Simmons, 2006](#), [Hart et al., 1975](#), [Sacheti et al., 2014](#), [van Ours, 2021](#)), and [Van Reeth \(2019\)](#) considers this for forecasting TV audiences for the Tour de France.

Cities and countries bid to host large events like the World Cup based on forecasts regarding the impact of hosting such events. Forecasts that are often inflated for political reasons ([Baade & Matheson, 2016](#)). Equally, franchise-based sports like many North American sports attract forecasts regarding the impact of a team locating in a city, usually resulting in public subsidies for the construction of venues for teams to play at [Coates and Humphreys \(1999\)](#). Governments invest in sporting development, primarily to achieve better performances at global events, most notably the Olympics ([Bernard & Busse, 2004](#)).

Many sporting events themselves rely on forecasts to function; high jumpers predict what height they will be able to jump over, and free diving contestants must state the depth they will dive to. Less formally, teams will set themselves goals: to win matches, to win competitions, to avoid the ‘wooden spoon’. Here, forecast outcomes are influenced

by the teams, and competitors, taking part in competitions and, as such, are perhaps less commonly thought of as genuine forecasts. Important works predicting outcomes range from [Dixon and Coles \(1997\)](#) in soccer, to [Kovalchik and Reid \(2019\)](#) for tennis, while the increasing abundance of data means that machine learning and deep learning methods are beginning to dominate the landscape. See, for example, [Hubáček, Šourek, and Železný \(2019\)](#) and [Maymin \(2019\)](#) for basketball, and [Mulholland and Jensen \(2019\)](#) for NFL.

3.8.7. Forecasting for megaprojects¹⁵⁴

Megaprojects are significant activities characterised by a multi-organisation structure, which produces highly visible infrastructure or asset with very crucial social impacts ([Aaltonen, 2011](#)). Megaprojects are complex, require huge capital investment, several stakeholders are identified and, usually a vast number of communities and the public are the receivers of the project's benefits. There is a need megaprojects especially those that deliver social and economic goods and create economic growth ([Flyvbjerg, Bruzelius, & Rothengatter, 2003](#)). Typical features of megaprojects include some or all the following: (i) delivering a substantial piece of physical infrastructure with a life expectancy that spans across decades, (ii) main contractor or group of contractors are privately owned and financed, and (iii) the contractor could retain an ownership stake in the project and the client is often a government or public sector organisation ([Sanderson, 2012](#)).

However, megaprojects are heavily laced with extreme human and technical complexities making their delivery and implementation difficult and often unsuccessful ([Merrow et al., 1988](#), [The R.F.E. Working Group Report, 2015](#)). This is largely due to the challenge of managing megaprojects including extreme complexity, increased risk, tight budget and deadlines, lofty ideals ([Fiori & Kovaka, 2005](#)). Due to the possibility and consequences of megaproject failure ([Mišić & Radujković, 2015](#)), forecasting the outcomes of megaprojects is becoming of growing importance. In particular, it is crucial to identify and assess the risks and uncertainties as well as other factors that contribute to disappointing outcomes of megaprojects in order to mitigate them ([Flyvbjerg et al., 2003](#), [Miller and Lessard, 2007](#)).

Literature review in forecasting in megaprojects is scarce. However, there are a few themes that have emerged in the extant literature as characteristics of megaprojects that should be skilfully managed to provide a guideline for the successful planning and

construction of megaprojects (Fiori and Kovaka, 2005, Flyvbjerg, 2007, Sanderson, 2012). Turner and Zolin (2012) even claim that we cannot even properly define what success is. They argue that we need to reliable scales in order to predict multiple perspectives by multiple stakeholders over multiple time frames — so definitely a very difficult long term problem. This could be done via a set of leading performance indicators that will enable managers of Megaprojects to forecast during project execution how various stakeholders will perceive success months or even years into the operation. At the very early stages of a project’s lifecycle, a number of decisions must be taken and are of a great importance for the performance and successful deliverables/outcomes. Flyvbjerg (2007) stress the importance of the front-end considerations particularly for Megaprojects Failure to account for unforeseen events frequently lead to cost overruns.

Litsiou et al. (2019) suggest that forecasting the success of megaprojects is particularly a challenging and critical task due to the characteristics of such projects. Megaproject stakeholders typically implement impact assessments and/or cost benefit Analysis tools (Litsiou et al., 2019). As Makridakis, Hogarth, and Gaba (2010) suggested, judgmental forecasting is suitable where quantitative data is limited, and the level of uncertainty is very high; elements that we find in megaprojects. By comparing the performance of three judgmental methods, unaided judgment, semi-structured analogies (sSA), and interaction groups (IG), used by a group of 69 semi-experts, Litsiou et al. (2019) found that, the use of sSA outperforms unaided judgment in forecasting performance (see also Section 2.11.4). The difference is amplified further when pooling of analogies through IG is introduced.

3.8.8. Competing products¹⁵⁵

Competition among products or technologies affects prediction due to local systematic deviations and saturating effects related to policies, and evolving interactions. The corresponding sales time series must be jointly modelled including the time varying reciprocal influence. Following the guidelines in subsection Section 2.3.20, some examples are reported below.

Based on IMS-Health quarterly number of cimetidine and ranitidine packages sold in Italy, the CRCDD model (Guseo & Mortarino, 2012) was tested to evaluate a diachronic competition that produced substitution. Cimetidine is a histamine antagonist that inhibits the production of stomach acid and was introduced by Smith, Kline & French in

1976. Ranitidine is an alternative active principle introduced by Glaxo in 1981 and was found to have far-improved tolerability and a longer-lasting action. The main effect in delayed competition is that the first compound spread fast but was suddenly outperformed by the new one principle that modified its stand-alone regime. [Guseo and Mortarino \(2012\)](#) give some statistical and forecasting comparisons with the restricted Krishnan-Bass-Kummar Diachronic model (KBKD) by [Krishnan et al. \(2000\)](#). Previous results were improved with the UCRC model in [Guseo and Mortarino \(2014\)](#) by considering a decomposition of word-of-mouth (WOM) effects in two parts: within-brand and cross-brand contributions. The new active compound exploited a large cross-brand WOM and a positive within-brand effect. After the start of competition, cimetidine experienced a negative WOM effect from its own adopters and benefited from the increase of the category's market potential driven by the antagonist. Forecasting is more realistic with the UCRC approach and it avoids mistakes in long-term prediction.

Restricted and unrestricted UCRC models were applied in Germany by [Guidolin and Guseo \(2016\)](#) to the competition between nuclear power technologies and renewable energy technologies (wind and solar; see also Sections [3.4.5 Forecasting renewable energy technologies](#), [3.4.6 Wind power forecasting](#), [3.4.8 Solar power forecasting](#)) in electricity production. Due to the 'Energiewende' policy started around 2000, the substitution effect, induced by competition, is confirmed by the electricity production data provided by BP.¹⁵⁶ An advance is proposed in [Furlan, Mortarino, and Zahangir \(2020\)](#) with three competitors (nuclear power, wind, and solar technologies) and exogenous control functions obtaining direct inferences that provide a deeper analysis and forecasting improvements in energy transition context.

Previous mentioned intersections between Lotka–Volterra approach and diffusion of innovations competition models suggested a more modulated access to the residual carrying capacity. The Lotka–Volterra with churn model (LVch) by [Guidolin and Guseo \(2015\)](#) represents 'churn effects' preserving within and cross-brand effects in a synchronic context.

An application of LVch model is discussed with reference to the competition/substitution between compact cassettes and compact discs for pre-recorded music in the US market. Obtained results of LVch outperform restricted and unrestricted UCRC analyses. In this context the residual market is not perfectly

accessible to both competitors and this fact, combined with WOM components, allows for better interpretation and forecasting especially in medium and long-term horizons. A further application of the LVch model, Lotka–Volterra with asymmetric churn (LVac), is proposed in [Guidolin and Guseo \(2020\)](#). It is based on a statistical reduction: The late entrant behaves as a standard [Bass \(1969\)](#) model that modifies the dynamics and the evolution of the first entrant in a partially overlapped market. The case study is offered by a special form of competition where the iPhone produced an inverse cannibalisation of the iPad. The former suffered a local negative interaction with some benefits: A long-lasting life cycle and a larger market size induced by the iPad.

A limitation in models for diachronic competition relates to high number of rivals, implying complex parametric representations with respect to the observed information. A second limitation, but also an opportunity, is the conditional nature of forecasting if the processes partially depend upon exogenous control functions (new policy regulations, new radical innovations, regular and promotional prices, etc.). These tools may be used to simulate the effect of strategic interventions, but a lack of knowledge of such future policies may affect prediction.

3.8.9. Forecasting under data integrity attacks¹⁵⁷

Data integrity attacks, where unauthorized parties access protected or confidential data and inject false information using various attack templates such as ramping, scaling, random attacks, pulse and smooth-curve, has become a major concern in data integrity control in forecasting ([Giani et al., 2013](#), [Singer and Friedman, 2014](#), [Sridhar and Govindarasu, 2014](#), [Yue, 2017](#)).

Several previous studies have given attention in anomaly detection pre-processing step in forecasting workflow with varying degree of emphasis. However, according to [Yue \(2017\)](#), the detection of data integrity attacks is very challenging as such attacks are done by highly skilled adversaries in a coordinated manner without notable variations in the historical data patterns ([Liang, He, & Chen, 2019](#)). These attacks can cause over-forecasts that demand unnecessary expenses for the upgrade and maintenance, and can eventually lead to poor planning and business decisions ([Luo et al., 2018a](#), [Luo et al., 2018b](#), [Wu et al., 2020](#)).

Short-term load forecasting (see Section [3.4.3](#)) is one major field that are vulnerable to malicious data integrity attacks as many power industry functions such as economic dispatch, unit commitment and automatic generation control heavily depend on

accurate load forecasts (Liang et al., 2019). The cyberattack on U.S. power grid in 2018 is one such major incident related to the topic. According to the study conducted by Luo et al. (2018a), the widely used load forecasting models fail to produce reliable load forecast in the presence of such large malicious data integrity attacks. A submission to the Global Energy Forecasting Competition 2014 (GEFCom2014) incorporated an anomaly detection pre-processing step with a fixed anomalous threshold to their load forecasting framework (Xie & Hong, 2016). The method was later improved by Luo, Hong, Yue (2018) by replacing the fixed threshold with a data driven anomalous threshold. Sridhar and Govindarasu (2014) also proposed a general framework to detect scaling and ramp attacks in power systems. Akouemo and Povinelli (2016) investigated the impact towards the gas load forecasting using hybrid approach based on Bayesian maximum likelihood classifier and a forecasting model. In contrast to the previous model based attempts, Yue, Hong, and Wang (2019) proposed a descriptive analytic-based approach to detect cyberattacks including long anomalous sub-sequences (see Section 2.2.3), that are difficult to detect by the conventional anomaly detection methods.

The problem of data integrity attacks is not limited to load forecasting. Forecasting fields such as elections forecasting (see Section 3.8.5), retail forecasting (see Section 3.2.4), airline flight demand forecasting (see Section 3.8.2) and stock price forecasting (see Section 3.3.13) are also vulnerable to data integrity attacks (Luo et al., 2018a, Seaman, 2018). For instant, Wu et al. (2020) explored the vulnerability of traffic modelling and forecasting in the presence of data integrity attacks with the aim of providing useful guidance for constrained network resource planning and scheduling. However, despite of the increasing attention toward the topic, advancements in cyberattacks on critical infrastructure raise further data challenges. Fooling existing anomaly detection algorithms via novel cyberattack templates is one such major concern. In response to the above concern, Liang et al. (2019) proposed a data poisoning algorithm that can fool existing load forecasting approaches with anomaly detection component while demanding further investigation into advanced anomaly detection methods. Further, adversaries can also manipulate other related input data without damaging the target data series. Therefore, further research similar to (Sobhani et al., 2020) are required to handle such data challenges.

3.8.10. The forecastability of agricultural time series¹⁵⁸

The forecasting of agricultural time series falls under the broader group of forecasting commodities, of which agricultural and related products are a critical subset. While there has been considerable work in the econometrics and forecasting literature on common factor models in general there is surprisingly little work so far on the application of such models for commodities and agricultural time series – and this is so given that there is considerable literature in the linkage between energy and commodities, including agricultural products, their prices and futures prices, their returns and volatilities. Furthermore, a significant number of papers is fairly recent which indicates that there are many open avenues of future research on these topics, and in particular for applied forecasting. The literature on the latter connection can consider many different aspects in modelling as we illustrate below. We can identify two literature strands, a much larger one on the various connections of energy with commodities and the agricultural sector (and in this strand we include forecasting agricultural series) and a smaller one that explores the issue of common factors. An early reference of the impact of energy on the agricultural sector is [Tewari \(1990\)](#) and then after a decade we find [Cohin and Chantret \(2010\)](#) on the long-run impact of energy prices on global agricultural markets. [Byrne, Fazio, and Fiess \(2013\)](#) is an early reference for co-movement of commodity prices followed by [Daskalaki, Kostakis, and Skiadopoulos \(2014\)](#) on common factors of commodity future returns and then a very recent paper from [Alquist, Bhattarai, and Coibion \(2020\)](#) who link global economic activity with commodity price co-movement. The impact of energy shocks on US agricultural productivity was investigated by [Wang and McPhail \(2014\)](#) while [Koirala, Mishra, D’Antoni, and Mehlhorn \(2015\)](#) explore the non-linear correlations of energy and agricultural prices with [Albulescu, Tiwari, and Ji \(2020\)](#) exploring the latter issue further, the last two papers using copulas. [Xiong, Li, Bao, Hu, and Zhang \(2015\)](#) is an early reference of forecasting agricultural commodity prices while [Kyriazi et al. \(2019\)](#), [Li, Li, Liu, Zhu and Wei \(2020\)](#), and [Wang, Wang, Li, and Zhou \(2019\)](#) consider three novel and completely different approaches on forecasting agricultural prices and agricultural futures returns. [López Cabrera and Schulz \(2016\)](#) explore volatility linkages between energy and agricultural commodity prices and then [Tian, Yang, and Chen \(2017\)](#) start a mini-stream on volatility forecasting on agricultural series followed among others by the work of [Luo, Klein, Ji, and Hou \(2019\)](#) and of [Degiannakis, Filis, Klein, and Walther \(2020\)](#). [de Nicola, De Pace, and](#)

Hernandez (2016) examine the co-movement of energy and agricultural returns while Kagraoka (2016) and Lübbers and Posch (2016) examine common factors in commodity prices. Pal and Mitra (2019) and Wei Su, Wang, Tao, and Oana-Ramona (2019) both investigate the linkages of crude oil and agricultural prices. Finally, Tiwari, Nasreen, Shahbaz, and Hammoudeh (2020) examine the time-frequency causality between various commodities, including agricultural and metals.

There is clearly room for a number of applications in the context of this recent research, such along the lines of further identifying and then using common factors in constructing forecasting models, exploring the impact of the COVID-19 crisis in agricultural production or that of climate changes on agricultural prices.

3.8.11. Forecasting in the food and beverage industry¹⁵⁹

Reducing the ecological impact and waste, and increasing the efficiency of the food and beverage industry are currently major worldwide issues. To this direction, efficient and sustainable management of perishable food and the control of the beverage quality is of paramount importance. A particular focus on this topic is placed on supply chain forecasting (see Section 3.2.2), with advanced monitoring technologies able to track the events impacting and affecting the food and beverage processes (La Scalia, Micale, Miglietta, & Toma, 2019). Such technologies are typically deployed inside manufacturing plants, yielding to Industry 4.0 solutions (Ojo, Shah, Coutroubis, Jiménez, & Ocana, 2018) that are enabled by state-of-the-art forecasting applications in smart factories. The transition from plain agriculture techniques to smart solutions for food processing is a trend that fosters emerging forecasting data-driven solutions in many parts of the world, with special attention to the sustainability aspects (Zailani, Jeyaraman, Vengadasan, & Premkumar, 2012).

Various forecasting approaches have been successfully applied in the context of the food and beverage industry, from Monte Carlo simulations based on a shelf-life model (La Scalia et al., 2019), to association rule mining (see Section 2.9.2) applied to sensor-based equipment monitoring measurements (Apiletti & Pastor, 2020), multi-objective mathematical models for perishable supply chain configurations, forecasting costs, delivery time, and emissions (Wang, Nhieu, Chung, & Pham, 2021), and intelligent agent technologies for network optimisation in the food and beverage logistics management (Mangina & Vlachos, 2005).

We now focus on the case of forecasting the quality of beverages, and particularly coffee. Espresso coffee is among the most popular beverages, and its quality is one of the most discussed and investigated issues. Besides human-expert panels, electronic noses, and chemical techniques, forecasting the quality of espresso by means of data-driven approaches, such as association rule mining, is an emerging research topic ([Apiletti and Pastor, 2020](#), [Apiletti et al., 2020](#), [Kittichotsawat et al., 2021](#)).

The forecasting model of the espresso quality is built from a real-world dataset of espresso brewing by professional coffee-making machines. Coffee ground size, coffee ground amount, and water pressure have been selected among the most influential external variables. The ground-truth quality evaluation has been performed for each shot of coffee based on three well-known quality variables selected by domain experts and measured by specific sensors: the extraction time, the average flow rate, and the espresso volume. An exhaustive set of more than a thousand coffees has been produced to train a model able to forecast the effect of non-optimal values on the espresso quality. For each variable considered, different categorical values are considered: ground size can be coarse, optimal, or fine; ground amount can be high, optimal, or low; brewing water pressure can be high, optimal, or low. The experimental setting of categorical variables enables the application of association rule mining (see Section [2.9.2](#)), a powerful data-driven exhaustive and explainable approach ([Han et al., 2011](#), [Tan et al., 2005](#)), successfully exploited in different application contexts ([Acquaviva et al., 2015](#), [Di Corso et al., 2018](#)).

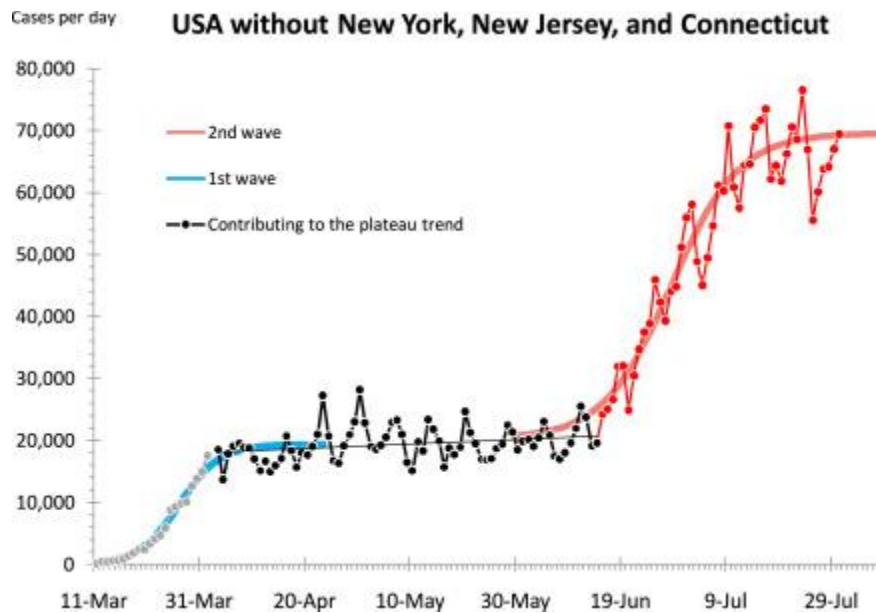
Several interesting findings emerged. If the water pressure is low, the amount of coffee ground is too high, and the grinding is fine, then we can forecast with confidence a low-quality coffee due to excessive percolation time. If the amount of coffee ground is low, the ground is coarse, and the pressure is high, then we can forecast a low-quality coffee due to excessive flow rate. Furthermore, the coarseness of coffee ground generates an excessive flow rate forecast, despite the optimal values of dosage and pressure, with very high confidence.

3.8.12. Dealing with logistic forecasts in practice¹⁶⁰

The forecaster faces three major difficulties when using the logistic equation (S curve); see also Section [2.3.19](#). A first dilemma is whether he or she should fit an S curve to the cumulative number or to the number per unit of time. Here the forecaster must exercise wise judgment. What is the “species” and what is the niche that is being filled? To the

frustration of business people there is no universal answer. When forecasting the sales of a new product it is often clear that one should fit the cumulative sales because the product's market niche is expected to eventually fill up. But if we are dealing with something that is going to stay with us for a long time (for example, the Internet or a smoking habit), then one should not fit cumulative numbers. At times this distinction may not be so obvious. For example, when COVID-19 first appeared many people (often amateurs) began fitting S curves to the cumulative number of infections (for other attempts on forecasting COVID-19, see Section 3.6.2). Some of them were rewarded because indeed the diffusion of the virus in some countries behaved accordingly (Debecker & Modis, 1994). But many were frustrated and tried to “fix” the logistic equation by introducing more parameters, or simply gave up on trying to use logistics with COVID 19. And yet, many cases (e.g., the US) can be illuminated by logistic fits but on the daily number of infections, not on the cumulative number. As of August 1, 2020, leaving out the three eastern states that had gotten things under control, the rest of the US displayed two classic S curve steps followed by plateaus (see Fig. 18). The two plateaus reflect the number of infections that American society was willing to tolerate at the time, as the price to pay for not applying measures to restrict the virus diffusion.

The second difficulty in using the logistic equation has to do with its ability to predict from relatively early measurements the final ceiling. The crucial question is how early can the final ceiling be determined and with what accuracy. Some people claim that before the midpoint no determination of a final level is trustworthy (Marinakis & Walsh, 2021). Forecasters usually abstain from assigning quantitative uncertainties on the parameters of their S curve forecasts mostly because there is no theory behind it. However, there is a unique study by Debecker and Modis (2021) that quantifies the uncertainties on the parameters determined by logistic fits. The study was based on 35,000 S curve fits on simulated data, smeared by random noise and covering a variety of conditions. The fits were carried out via a minimisation technique. The study produced lookup tables and graphs for determining the uncertainties expected on the three parameters of the logistic equation as a function of the range of the S curve populated by data, the error per data point, and the confidence level required.



1. [Download: Download high-res image \(247KB\)](#)
2. [Download: Download full-size image](#)

Fig. 18. Two logistic-growth steps during the early diffusion of COVID-19 in America (March to July, 2020).

The third difficulty using the logistic equation comes from the fact that no matter what fitting program one uses, the fitted S curve will flatten toward a ceiling *as early and as low* as it is allowed by the constraints of the procedure. As a consequence fitting programs may yield logistic fits that are often biased toward a low ceiling. Bigger errors on the data points accentuate this bias by permitting larger margins for the determination of the S curve parameters. To compensate for this bias the user must explore several fits with different weights on the data points during the calculation of the . He or she should then favour the answer that gives the highest ceiling for the S curve (most often obtained by weighting more heavily the recent historical data points). Of course, this must be done with good justification; here again the forecaster must exercise wise judgment.

3.9. The future of forecasting practice¹⁶¹

Jean-Baptiste Karr (1849)

It would be a more straightforward task to make predictions about the future of forecasting practice if we had a better grasp of the present state of forecasting practice. For that matter, we lack even a common definition of forecasting practice. In a recent article, [Makridakis, Bonneli et al. \(2020\)](#) lamented the failure of truly notable advances

in forecasting methodologies, systems, and processes during the past decades to convince many businesses to adopt systematic forecasting procedures, leaving a wide swath of commerce under the guidance of ad hoc judgment and intuition. At the other extreme, we see companies with implementations that combine state-of-the-art methodology with sophisticated accommodations of computing time and costs, as well as consideration of the requirements and capabilities of a diverse group of stakeholders (Yelland, Baz, & Serafini, 2019). So, it is not hyperbole to state that business forecasting practices are all over the place. What surely is hyperbole, however, are the ubiquitous claims of software providers about their products accurately forecasting sales, reducing costs, integrating functions, and elevating the bottom line (Makridakis, Bonneli et al., 2020, Sorensen, 2020). For this section, we grilled a dozen practitioners and thought leaders (“the Group”) about developments playing out in the next decade of forecasting practice, and have categorised their responses:

- •
Nature of forecasting challenges;
- •
Changes in the forecasting toolbox;
- •
Evolution in forecasting processes such as integration of planning functions;
- •
Expectations of forecasters; and
- •
Scepticism about real change.

Forecasting Challenges: Focusing on operations, the Group sees demand forecasting becoming ever more difficult due to product/channel proliferation, shorter lead times, shorter product histories, and spikes in major disruptions.

- •
Operational forecasts will have shorter forecast horizons to increase strategic agility required by business to compete, sustain, and survive.
- •
New models will need to incorporate supply-chain disruption. Demand chains will need to be restarted, shortening historical data sets and making traditional models less viable due to limited history.

-
-

Lead times will decrease as companies see the problems in having distant suppliers. Longer lead times make accurate forecasting more difficult.

Forecasting Tool Box: Unsurprisingly, this category received most of the Group’s attention. All predict greater reliance on AI/ML for automating supply-and-demand planning tasks and for reconciling discrepancies in hierarchical forecasting. Longer-horizon causal forecasting models will be facilitated by big data, social media, and algorithmic improvements by quantum computing. Post-COVID, we will see a greater focus on risk management/mitigation. The Cloud will end the era of desktop solutions.

-
-

Quantum computers will improve algorithms used in areas like financial forecasting (e.g., Monte Carlo simulations), and will change our thinking about forecasting and uncertainty.

-
-

Although social media is a tool for “what’s trending now”, new models will be developed to use social-media data to predict longer-term behaviour. Step aside Brown (exponential smoothing) and Bass (diffusion).

-
-

Greater automation of routine tasks (data loading, scrubbing, forecast generation and tuning, etc.) through AI/ML-powered workflow, configurable limits, and active alerts. More black box under the hood, but more clarity on the dashboard.

-
-

Greater focus on risk management/mitigation through what-if scenarios, simulations, and probabilistic forecasting.

Forecasting Processes and Functional Integration: Systems will become more integrated, promoting greater collaboration across functional areas and coordination between forecast teams and those who rely upon them. Achieving supply-chain resilience will become as important as production efficiency, and new technology such as Alert and Root Cause Analysis systems will mitigate disruptions.

-
-

S&OP will expand from its home in operations to more fully integrate with other functions such as finance and performance management, especially in larger multinationals.

- •

The pandemic has forced firms to consider upping supply-chain resilience. Firms are building in capacity, inventory, redundancy into operations—somewhat antithetical to the efficiency plays that forecasting brings to the table.

- •

Forecasting will be more closely tied to Alert and Root Cause Analysis systems, which identify breakdowns in processes/systems contributing to adverse events, and prevent their recurrence.

Expectations of Forecasters: Agreement was universal that the forecaster's job description will broaden and become more demanding, but that technology will allow some redirection of effort from producing forecasts to communicating forecasting insights.

- •

The interest around disease models increases our awareness of the strengths and weaknesses of mathematical models. Forecasters may need to become more measured in their claims, or do more to resist their models being exploited.

- •

We will see a transformation from demand planner to demand analyst, requiring additional skill sets including advanced decision making, data and risk analysis, communication, and negotiation.

- •

Professional forecasters will be rare except in companies where this expertise is valued. Fewer students are now educated or interested in statistical modelling, and time is not generally available for training.

- •

Forecasters will learn the same lesson as optimisation folks in the 1990s and 2000s: the importance of understanding the application area—community intelligence.

Scepticism: Many were sceptical about the current enthusiasm for AI/ML methods; disappointed about the slow adoption of promising new methods into software systems and, in turn, by companies that use these systems; and pessimistic about the respect given to and influence of forecasters in the company's decision making.

- •

While AI/ML are important additions to the forecaster's toolbox, they will not automatically solve forecasting issues. Problems include data hunger, capacity brittleness, dubious input data, fickle trust by users (Kolassa, 2020c), and model bias.

- •
Practices in the next decade will look very similar to the present. Not that much has changed in the last decade, and academic developments are slow to be translated into practice.
- •
Politics, gaming, and the low priority given to forecasting are the prime drivers of practice, thus limiting interest in adopting new methodologies.
- •
None of the topical items (AI/ML, big data, demand sensing, new forecasting applications) will have much of an impact on forecasting practice. Forecasting departments hop from one trend to the other without making much progress towards better forecasting accuracy.
- •
Software companies will struggle, despite good offerings. Most companies do not want to invest in excellent forecasting engines; whatever came with their ERP system is "good enough".
- •
Forecasting will continue to suffer from neglect by higher levels of management, particularly when forecasts are inconveniently contrary to the messages management hopes to convey.

Note finally that the COVID-19 pandemic has elevated practitioner concerns about disruptions to normal patterns as well as the fear of an increasingly volatile environment in which forecasts must be made. There are indications that companies will place more stress on judgmental scenarios, likely in conjunction with statistical/ML methods.

4. Forecasting: benefits, practices, value, and limitations¹⁶²

The purpose of this unique article is to provide an encyclopedic knowledge about the various aspects of forecasting. In this article, there are more than 140 sections and subsections, with more than 2,100 references, written by 80 of some of the best-known forecasting researchers and practitioners in the world, making it into a selective,

encyclopedic piece covering, into a single source, a great deal of the available knowledge about the theory and practice of forecasting. We hope that this article will serve as an easy-to-use reference source. We aim to convert it into an online resource that will be regularly updated as new information becomes available.

But some people argue if there is any value in attempting to predict the future and if forecasting is any different than fortune telling, given the large numbers of mistaken forecasts made in the past, including our inability to accurately predict the progression of COVID-19 and its economic and human consequences? What is, therefore, the usefulness of a paper like the present one when crystal balling is not possible, and uncertainty reigns? It is the aim of this concluding article to set the record straight, explaining the benefits and practical value of forecasting while reporting its limitations too.

The Myriad of Forecasts: All planning and the great majority of decisions we make require forecasting. Deciding what time to get up in the morning, not to be late for work implies a correct prediction of the commuting time to go to the office. Determining what to study is another decision requiring elaborate predictions about the demand for future jobs decades away. In the business world, firms must decide/forecast how many products to manufacture, the price they should be sold, how much money to spend on advertising and promotion, how much and in what type of new technologies to invests and a plethora of other future-oriented decisions requiring both predictions and assessing their inevitable uncertainty. Whether we like it or not, we have no choice but making these forecasts to benefit as much as possible from their value, knowing perfectly well that all predictions are uncertain while some may turn out to be wrong.

The Pervasiveness of Uncertainty: Apart from some areas of hard sciences, all other forecasts are uncertain and must be accompanied with a measure of its magnitude, expressed as a prediction interval, or as a probability distribution around the most likely forecast. Although the value and usage of forecasts is clear, that of uncertainty is not. Worse, it becomes an unwelcome source of anxiety whose usefulness is misunderstood. Executives want to know the exact sales of their firm for next month to set up their production schedule. Instead, they are given prediction intervals (PIs) around such forecast and told that most of the time, sales will be within this interval, assuming the fluctuations follow some distributional assumptions. They argue that forecasting must decrease, not amplify, uncertainty and that the PIs are too wide and 'uninformative' to

be used for making practical business decisions. The trouble is that these PIs are based on past fluctuations and present the best estimation of future uncertainty, even if they seem too wide. Worse, empirical research has shown that they are too narrow, underestimating uncertainty often considerably.

Assessing Uncertainty and Dealing with its Implied Risks: Uncertainty entails risks, requiring action to minimise their negative consequences. There are two kinds of uncertainty that can be illustrated by a commuting example. The first relates to fluctuations in the commuting time under normal driving conditions when there are no serious accidents, road works or major snowstorms. Such fluctuations are small and can be captured by a normal curve that allows to balance the risk of arriving earlier or later than the desired time. In the opposite case, uncertainty is fat-tailed and hard to estimate, as delays can be substantial depending upon the seriousness of the accident or that of the snowstorm while the risk of being early to work is smaller than being late. Moreover, such risk is substantially different when going to the airport to catch a flight, requiring starting much earlier than the average time it takes to go to the airport to minimise the risk of missing the flight.

More Accurate Ways of Forecasting and Assessing Uncertainty: Extensive empirical research, including forecasting competitions, has shown that systematic approaches improve the accuracy of forecasting and the correct assessment of uncertainty resulting in substantial benefits when compared to ad-hoc judgmental alternatives ([Makridakis, Bonneli et al., 2020](#)). The biggest advantage of such approaches is their ability to identify and estimate, in a mathematically optimal manner, past patterns and relationships that are subsequently extrapolated to predict their continuation, avoiding the over optimism and wishful thinking associated with judgmental approaches. At the same time, it must be clear that the accuracy of the forecasts and the correctness of uncertainty will depend on the established patterns/relationship not changing much during the forecasting period.

Using Benchmarks to Evaluate the Value of Forecasting: The accuracy of the forecasts and the correct assessment of uncertainty must be judged not on their own but in comparison to some simple, readily available benchmarks. In stock market forecasts, for instance, the accuracy of predictions is compared to that of today's price used as the forecast for future periods. Empirical comparisons have shown that such a benchmark beats the great majority of professional forecasters, hence Buffet's advice in the epigram

for his wife to invest in a low-cost index fund that selects stocks randomly. In weather forecasting, meteorologists are judged by the improvement of their forecasts over the naive prediction that tomorrow's weather will be the same as today.

Concluding remark: Accepting the advantages and limitations of systematic forecasting methods and most importantly avoiding any exaggerated expectations of what it can achieve is critical. Such methods do not possess any prophetic powers, they simply extrapolate established patterns and relationships to predict the future and assess its uncertainty. Their biggest advantage is their objectivity and ability for optimal extrapolation. Their biggest disadvantages are: (i) the patterns and the relationships must remain fairly constant during the forecasting phase for the forecasts to be accurate, and (ii) uncertainty must not be fat-tailed so it can be measured quantitatively.