# 42nd International Symposium on Forecasting

July 10-13, 2022. Oxford, UK

General Chair: John E. Boylan, Program Chair: Ivan Svetunkov

## Book of Abstracts

### Macroeconomic Forecasting Evaluation of MIDAS Models

Presenter: Nicolas Bonino-Gayoso
Co-authors: Nicolas Bonino-Gayoso;Alfredo Garcia-Hiernaux

In this paper we compare the nowcasting and forecasting performance of some MIDAS models (ADL-MIDAS, TF-MIDAS and U-MIDAS) when predicting the GDP growth of the four main Euro Area economies in the period 2011Q4-2020Q3. To do so, we consider various high frequency indicators, horizons and subperiods, each of the latter with a distinct level of uncertainty. Several meta-regressions, with an average error metric as endogenous variable, are estimated to account for potential differences in performance by country, indicator, sample period or method. The results obtained with the whole sample do not reveal any difference in the predictive performance of the models under comparison. In contrast, the analysis by subperiods shows that ADL-MIDAS and U-MIDAS are more suitable in less uncertain economic contexts, while TF-MIDAS outperforms them in more unpredictable ones. These findings are robust to the forecasting error metric used, RMSFE or MAFE.

### Testing big data in a big crisis: Nowcasting under COVID-19

Presenter: Luca Tiozzo Pezzoli
Co-authors: Luca Tiozzo Pezzoli;Luca Onorante;Luca Barbaglia;Lorenzo Frattarolo;Marco Ratto;Filippo Maria Pericoli

During the COVID-19 pandemic, economists have struggled to obtain reliable economic predictions, with standard models becoming outdated and their forecasting performance deteriorating rapidly. This paper presents two novelties that could be adopted by forecasting institutions in unconventional times. The first innovation is the construction of a large data set for macroeconomic forecasting in Europe. We collect more than a thousand time series from conventional and unconventional sources: we complement traditional macroeconomic variables with timely big data indicators, and assess their added value at nowcasting. The second novelty consists of a methodology to merge an enormous amount of non-encompassing data and a large battery of classical and more sophisticated forecasting methods in a seamlessly dynamic Bayesian framework. Specifically, we introduce an innovative "selection prior' ' that is used not as a way to influence model outcomes, but as a selecting device among competing models. This allows the search for better specifications to go hand in hand with the forecasting process.
By applying this methodology to the COVID-19 crisis, we show which variables are good predictors for nowcasting purposes and draw lessons for dealing with possible future crises.

### The Nowcasting Lab II

Presenter: Philipp Kronenberg
Co-authors: Philipp Kronenberg;Stefan Neuwirth;Heiner Mikosch

The Nowcasting Lab is an automated code-database-website environment for GDP nowcasting. It generates nowcasts and one-quarter ahead forecasts for quarterly GDP growth of currently 12 European countries and the euro area using several nowcasting models and a large set of data. The projections are updated daily and are released on a website. The website also provides detailed additional information such as the outcomes of the different models, the forecast changes from day to day, the forecast impacts of new data and past projection errors. All information about the models, methods and data is made transparent. Applied forecasters can use the website as an extended arm for their work. The daily projections and the input data are stored in a real-time database. Researchers can use the database for real-time forecasting exercises.

## The Nowcasting Lab

Presenter: Stefan Neuwirth
Co-authors: Stefan Neuwirth;Heiner Mikosch;Philipp Kronenberg

The Nowcasting Lab is an automated code-database-website environment for GDP nowcasting. It generates nowcasts and one-quarter ahead forecasts for quarterly GDP growth of currently 11 European countries and the euro area using several nowcasting models and a large set of data. The projections are updated daily and released on a website. The website also provides detailed additional information such as the outcomes of the different models, the forecast changes from day to day, the forecast impacts of new data, and past projection errors. All information about the models, methods and data is made transparent. Applied forecasters can use the website as an extended arm for their work. The daily projections and the input data are stored in a real-time database. Researchers can use the database for real-time forecasting exercises.

## Designing and Running a Biorisk Forecasting Tournament

Presenter: David Manheim
Co-authors: David Manheim;Juan Cambeiro

A number of critical policy and related objectives in biosecurity and pandemic risks have been highlighted by the ongoing COVID-19 pandemic. Unfortunately, both policy responses and actual risks are uncertain, and statistical and similar forecasting methods are ill-suited to both politics, and to rare or novel types of events like novel disease pandemics. For this reason, judgemental forecasting approaches seem particularly well suited to providing insight. Working with Metaculus, an online forecasting web site which has recently started running tournaments focused on cause-areas, we put together and operationalized a set of questions covering topics related to political and policymaker responses, technological developments, and the substantive risks of future pandemics. These are split between short and long term questions, with prizes for each intended to incentivize forecasting accuracy. In this discussion, I will describe the design and incentives created for this contest, with thoughts about how this can be applied to other similar forecasting arenas.

## Challenges in forecasting AI progress

Presenter: Florian E. Dorner
Co-authors: Florian E. Dorner;

Forecasts of progress in Artificial Intelligence (AI) can play an important role in anticipating and better understanding the potentially transformative global impacts of advanced AI technology. Such improved understanding can be vital for the effectiveness of both wider efforts at governing AI technology and technical research aimed at alleviating risks posed by advanced AI. We discuss key challenges in forecasting AI progress based on a Delphi panel of domain experts. These include the validation of forecasting efforts despite long resolution horizons as well as tensions between establishing well-defined forecasting targets for AI progress and making forecasts action-guiding. The latter is particularly challenging as, unlike in

other areas of technological forecasting such as in the energy sector, the relationship between standard performance metrics for AI systems and downstream economic impacts is not yet well understood. This is further exacerbated by discontinuous relationships between broad performance metrics and the ability to perform specific tasks that have been observed for individual AI systems at the micro-level, as well as the potential for discontinuous progress on the macro-level postulated by some scholars.

## An Interactive Calibration Training App for Improving Probabilistic Judgments

Presenter: Ross Gruetzemacher
Co-authors: Ross Gruetzemacher;David Paradice;Kang Bok Lee

An openly accessible, easy-to-use, and effective means of improving probabilistic judgments would have substantial positive impacts for a wide variety of applications. This is of particular interest in the Effective Altruism community, where decision making has significant implications on public good. We describe an exploratory study examining the viability of a novel interactive training app with this goal. Specifically, the app is designed for improving calibration and reducing overconfidence in probabilistic judgments. We evaluated this app by conducting an American college football forecasting tournament involving 153 business school students making 52 forecasts over 11 weeks. A coarsened exact matching analysis found statistical evidence that, in under thirty minutes, the more challenging training was able to reduce overconfidence, improve calibration and improve the accuracy of probabilistic judgments (measured by the Brier score). The experimental results also suggest that the app's generic training may be able to generalize across domains and that effective calibration training may be possible without expert facilitators. While no previous studies have reported similar results, we still conclude that these results should only be interpreted as a proof of concept, and that further evaluation and validation of mechanisms of the app's effect is necessary.

## Forecasting long-term treatment effects without long-term outcome data

Presenter: David Rhys Bernard
Co-authors: David Bernard;

Estimating the long-term impacts of actions and policies is important in many areas ranging from development economics to existential risk but the key common difficulty is that long-term outcomes are only observed with a long delay. I empirically test two approaches for forecasting long-term treatment effects using long-run randomized controlled trials from economics. The first approach is the surrogate index method, where a set of observed short-run outcomes are used to create a statistical surrogate for the long-run outcome. I find that the method typically suffers from attenuation bias although it performs reasonably well when there are small time gaps between the surrogates and the outcome. The second approach is asking for experts to directly provide their subjective forecasts of the long-run effects. I find that experts provide much more accurate forecasts than laypeople but previous experience with forecasting is more strongly correlated with forecasting accuracy than domain expertise. Providing people with information on short-run effects improves their long-run accuracy more than providing them with additional information on the intervention or the context. Overall, both approaches are promising for producing forecasts of treatment effects on a 5-15 year timeframe.

## How Does Economic Activity Interact with Climate? What We Learn from Global Temperature Anomaly Distributions

Presenter: Zack Miller
Co-authors: Zack Miller;Yoosoon Chang;Joon Park

We study feedbacks in the economic-climate system using a mixed autoregression, a mixture of vector autoregression and functional autoregression. Our mixed autoregression involves the global distribution of temperature anomalies as a functional variable in addition to other aggregate variables representing

economic activity, climate forcings from greenhouse gases and tropospheric aerosols resulting from global sulfur dioxide emissions, and solar and volcanic activities. We identify three types of output shocks – emissions-neutral (green), carbon-based (brown), and sulfur-based (yellow) – and a multi-dimensional natural climate variation (blue) shock, as well as solar and volcanic shocks. Anthropogenic brown shocks are historically the most important driver of global warming, while blue shocks have no eminent effect. Green shocks do not affect climate, and yellow shocks, historically well explained by policies on sulfur emissions, yield uncertain and complicated effects. By isolating the channel of shocks through climate onto economic activity, we distinguish between the direct effect (growth) and indirect effect (damage) of anthropogenic shocks. Relative to a growth rate of 2.60% per year over 170 years, we find that the growth rate would have been as high as 2.79% in the absence of carbon-based damages.

## Analyzing Differences between Scenarios

Presenter: David Hendry
Co-authors: David Hendry;Felix Pretis

Comparisons between alternative scenarios are used in many disciplines from macroeconomics, through epidemiology to climate science to help with planning future responses. Differences between scenario paths are often interpreted as signifying likely differences between outcomes that would materialise in reality. However, even when using correctly specified statistical models of the in-sample data generation process, additional conditions are needed to sustain inferences about differences between scenario paths. We consider two questions in scenario analyses: First, does testing the difference between scenarios yield additional insight beyond simple tests conducted on the model estimated in-sample? Second, when does the estimated scenario difference yield unbiased estimates of the true difference in outcomes? Answering the first question, we show that the calculation of uncertainties around scenario differences raises difficult issues since the underlying in-sample distributions are identical for both 'potential' outcomes when the reported paths are deterministic functions. Under these circumstances, a scenario comparison adds little beyond testing for the significance of the perturbed variable in the estimated model. Resolving the second question, when models include multiple covariates, inferences about scenario differences depend on the relationships between the conditioning variables, especially their invariance to the interventions being implemented. Tests for invariance based on automatic detection of structural breaks can help identify in-sample invariance of models to evaluate likely constancy in projected scenarios. Applications of scenario analyses to impacts on the UK's wage share from unemployment and agricultural growth from climate change illustrate the concepts.

## The Expected Macroeconomic Impact of Extreme Weather Events

Presenter: Andrew Martinez
Co-authors: Andrew Martinez;

Expectations about climate change are believed to matter for macroeconomic developments and policy. Previous research has argued that this expectations channel of climate change is akin to a demand shock. We test these predictions using a new panel of high frequency macroeconomic forecast revisions over the last two decades. We identify how forecasters revise their macroeconomic expectations in response to large hurricane damage shocks and find that the expected macroeconomic impact is akin to a supply shock. We also explore how responses differ in the short vs. long-run and test for heterogeneity in attentiveness to extreme weather events across individual forecasters.

## Identification of integrated assessment models of climate change

Presenter: Sophocles Mavroeidis
Co-authors: Sophocles Mavroeidis;

I discuss the fundamental identification problem in the estimation of global integrated assessment

models using aggregate data, focusing on a stylized version of the DICE model of Nordhaus. The most significant challenge arises from the simultaneous determination (over long periods) of climate and economic activity due to bidirectional feedback mechanisms, but there are also important challenges due to measurement error and missing data. I critically evaluate the identifying assumptions used in the literature via short and long-run restrictions and discuss their implications for the estimation of the social cost of carbon and for long-run projections under alternative mitigation strategies

## Forecast Evaluation and Selection in Unstable Environments

Presenter: Katja Smetanina
Co-authors: Katja Smetanina;Stefan Richter

Out-of-sample tests are widely used for evaluating and selecting between models' forecasts in economics and finance. Although widely used, underlying these tests is often the assumption of constant relative performance between competing models which is invalid for many practical applications. We propose a new two-step methodology designed specifically for Forecast Evaluation and selection in a world of changing relative performance. In the first step we estimate the time-varying mean and variance of the series for forecast loss differences, and in the second step we use these estimates to compare and rank models in a changing world. We show that our tests have high power against a variety of fixed and local alternatives.

## On the asymptotic behavior of bubble date estimators

Presenter: Anton Skrobotov
Co-authors: Anton Skrobotov;Eiji Kurozumi

In this study, we extend the three-regime bubble model of Pang et al. (2021) to allow the forth regime followed by the unit root process after recovery. We provide the asymptotic and finite sample justification of the consistency of the collapse date estimator in the two-regime AR(1) model. The consistency allows us to split the sample before and after the date of collapse and to consider the estimation of the date of exuberation and date of recovery separately. We have also found that the limiting behavior of the recovery date varies depending on the extent of explosiveness and recovering.

## FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series

Presenter: Matteo Barigozzi
Co-authors: Matteo Barigozzi;Haeran Cho;Dom Owens

We propose fnets, a methodology for network estimation and forecasting of high-dimensional time series exhibiting strong serial- and cross-sectional correlations. We operate under a factor-adjusted vector autoregressive (VAR) model where, after controlling for common factors accounting for pervasive co-movements of the variables, the remaining idiosyncratic dependence between the variables is modelled by a sparse VAR process. Network estimation of fnets consists of three steps: (i) factor-adjustment via dynamic principal component analysis, (ii) estimation of the parameters of the latent VAR process by means of 1-regularised Yule-Walker estimators, and (iii) estimation of partial correlation and long-run partial correlation matrices. In doing so, we learn three networks underpinning the latent VAR process, namely a directed network representing the Granger causal linkages between the variables, an undirected one embedding their contemporaneous relationships and finally, an undirected network that summarises both lead-lag and contemporaneous linkages. In addition, fnets provides a suite of methods for separately forecasting the factor-driven and the VAR processes. Under general conditions permitting heavy tails and weak factors, we derive the consistency of fnets in both network estimation and forecasting. Simulation studies and real data applications confirm the good performance of fnets.

## A Machine Learning Attack on Illegal Trading

Presenter: Artem Prokhorov
Co-authors: Artem Prokhorov;Robert James;henry leung

We design an adaptive framework for the detection of illegal trading behavior. Its key component is an extension of a pattern recognition tool, originating from the field of signal processing and adapted to modern electronic systems of securities trading. The new method combines the flexibility of dynamic time warping with contemporary approaches from extreme value theory to explore large-scale transaction data and accurately identify illegal trading patterns. Importantly, our method does not need access to any confirmed illegal transactions for training. We use a high-frequency order book dataset provided by an international investment firm to show that the method achieves remarkable improvements over alternative approaches in the identification of suspected illegal insider trading cases.

## Predictability of Jumps

Presenter: Jian Chen
Co-authors: Jian Chen;Michael Clements;Andrew Urquhart

We assess the predictability of jumps in asset prices using a self-exciting process embedded in a stochastic volatility model. We show the jump clustering feature can vary from frequencies of data. In the out-of-sample period, we use a particle filter to sample latent states and forecast one-step-ahead underlying intensity (probability) of jumps. We find our approach successfully predict directions of future returns. We study the discrepancies between probabilities of positive and negative jumps. In a regression analysis of empirical studies on Bitcoin data, we find returns can be explained by distribution's tails of the probability discrepancies, and coefficients are statistically significantly. We propose a trading strategy based on the jump clustering feature. The trading portfolio can yield a Sharpe ratio of 4.16 with transaction costs. We also show that the jump clustering feature diminishes, and that the trading strategy becomes less effective when the data frequency is lower than 45 minutes.

## Revisiting the volatility prediction with a global integration index

Presenter: Shiyao Qin
Co-authors: Shiyao Qin;Cindy S.H. Wang

This paper comprehensively examines the stock index volatility predictions of our sample of 12 economies with active trading stock index options. We show that (i) a family of new proposed global market integration indices based on various trading strategies from the option markets are useful in improving the prediction of global stock volatilities; (ii) incorporating the new indices into the heterogeneous autoregression model with implied volatility (Jeon, Seo & Kim, 2020 1 ) by pooled estimation could outperform the current volatility prediction model.

## The Most Predictable Aspects of Time Series

Presenter: Tommaso Proietti
Co-authors: Tommaso Proietti;

The paper deals with establishing what are the most predictable aspects of a univariate time series. Assuming the mutual information between past and future as the measure of predictability, we consider semiparametric classes of transformations, depending on a set of basis functions, that are combined linearly with coefficients chosen so as to maximise the mutual information between past and future. For financial time series it turns out that the most predictable aspect is volatility, whereas for macro time series it is a robustfied measure of growth. Also, nonlinear dependencies can be discovered.

## Forecasting extended CoVaR for crypto and non-crypto assets in times of the COVID-19 crisis: An asymmetric model approach

Presenter: Arief Hakim
Co-authors: Arief Hakim;Khreshna Syuhada

Cryptocurrencies have been viewed as new speculative investment assets with evidence of bubble behavior that might endanger the stability and functioning of the global financial system. In this paper, we aim to forecast an extension of conditional value-at-risk (CoVaR), namely, multiple CoVaR (MCoVaR), to assess systemic risk in the markets for cryptocurrencies and other global assets amidst the ongoing COVID-19 crisis. The MCoVaR forecast is computed for each asset by incorporating the joint conditioning event of the remaining assets at the same time. A multivariate Johnson's SU distribution with a smooth-transition GARCH (ST-GARCH) specification for the volatility equation is constructed to capture asymmetry in the conditional distribution as well as in volatility of their return processes. This approach provides the MCoVaR forecast with an explicit expression. The empirical results could bring practical implications for policy designers in making decisions to maintain financial stability in times of crisis due to COVID-19.

## How to Limit the Spillover from the 2021 Inflation Surge to Inflation Expectations?

Presenter: Michael Lamla
Co-authors: Michael Lamla;Lena Dräger;Damjan Pfajfar

Many central banks currently face inflation well above their targets and with that the challenge to prevent spillovers on inflation expectations. We study the effect of different communication about the 2021 inflation surge on German consumers' inflation expectations using a randomized control trial. We show that information about rising inflation increases short- and long-term inflation expectations. This initial increase in expectations can be mitigated using information about inflation projections, where numerical information about professional forecasters' projections seems to reduce inflation expectations by more than policymaker's characterization of inflation as a temporary phenomenon.

## Survey respondents' inflation forecasts and unprecedented events

Presenter: Mike Clements
Co-authors: Michael Clements;

How do professionals forecast in uncertain times, when the relationships between variables that held in the past may no longer be useful for forecasting the future? For Inflation Forecasting, we answer this question by measuring survey respondents' adherence to their pre-Covid Phillips curve models during the pandemic. We also ask whether professionals ought to have put their trust in their Phillips curve models over the Covid period. We address these questions allowing for heterogeneity in respondents' forecasts, and in their perceptions of the Phillips curve relationship.

## Heterogeneity in survey-based density forecasts: A compositional data approach

Presenter: Geoff Kenny
Co-authors: Geoff Kenny;Alexander Glass;Jonas Dovern

We propose to treat survey-based density expectations as compositional data when testing either for changes in density forecasts over time or for heterogeneity across different density forecasters. Monte Carlo simulations show that our test has more power relative to both a bootstrap approach based on the KLIC and an approach, which involves multiple testing for differences of individual parts of the density. In addition, the test is much faster than the KLIC-based one since it does not rely on simulations. Using density expectations from the ECB Survey of Professional Forecasters and the Survey of Consumer Expectations, we

show the usefulness of the test in detecting possible changes in inflation density forecasts and in identifying heterogeneity across different types of inflation density forecasts.

## The Predictive Content of Inflation Expectations Measures

Presenter: James Mitchell
Co-authors: James Mitchell;Saeed Zaman

We empirically investigate the predictive relationship between the full distribution of future inflation and current measures of inflationary expectations from households, professional forecasters, and financial markets. We find that households' expectations of inflation better explain the upper and lower quantiles of future inflation, whereas professional economists and market-based measures have more predictive power in the middle of the distribution. Our findings suggest that in the current high inflation environment, households' inflationary expectations provide a more accurate assessment about the future trajectory of inflation. Conditional predictive densities for inflation exhibit marked departures from Gaussianity, demonstrating the importance of allowing for nonlinearities and a diversity of views when using expectations data to gauge inflationary pressures.

## Generalized additive models for residential electricity demand forecasting

Presenter: Matteo Fasiolo
Co-authors: Matteo Fasiolo;

Future grid management systems will coordinate distributed production and storage resources to manage, in a cost-effective fashion, the increased load and variability brought by the electrification of transportation and by a higher share of weather-dependent production. Electricity demand forecasts at a low level of aggregation will be key inputs for such systems. In this talk, I will focus on forecasting demand at the individual household level, which is more challenging than forecasting aggregate demand, due to the lower signal-to-noise ratio and to the heterogeneity of consumption patterns across households. I will describe a new ensemble method for Probabilistic Forecasting, which borrows strength across the households while accommodating their individual idiosyncrasies. The first step consists in designing a set of models or 'experts' which capture different demand dynamics and fitting each of them to the data from each household. Then the idea is to construct an aggregation of experts where the ensemble weights are estimated on the whole data set, the main innovation being that we let the weights vary with the covariates by adopting an additive model structure. In particular, the proposed aggregation method is an extension of regression stacking where the mixture weights are modelled using linear combinations of parametric, smooth or random effects. The methods for building and fitting additive stacking models are implemented by the gamFactory R package, available at https://github.com/mfasiolo/gamFactory

## The Effect of Weather on Probabilistic Smart Meter Load Forecasting

Presenter: Nicole Ludwig
Co-authors: Nicole Ludwig;Thomas Kittel;Stephen Haben

The optimal operation of future sustainable energy systems requires accurate probabilistic forecasts of electricity demand on various hierarchy levels of the energy system — ranging from individual households to cities and entire countries. Weather variables, such as temperature and cooling power of wind, are strongly correlated with electricity demand on higher levels in the hierarchy (e.g. city level). Still, recent research suggests a limited effect on lower levels (e.g. household level). In this work, we investigate the impact of weather on low-voltage electricity demand forecasting on different hierarchies. We consider four Probabilistic Forecasting models, adjust the models to utilize weather data and compare the effect of weather variables on the forecasting accuracy. Our results indicate that both actual weather data and weather forecasts can improve the forecasting performance on most hierarchy levels. Furthermore, the weather effect is strongly related to the number of aggregated smart meters.

## Review of low voltage load forecasting: Methods, applications, and recommendations

Presenter: Siddharth Arora
Co-authors: Siddharth Arora;Stephen Haben;Georgios Giasemidis;Marcus Voss;Danica Vukadinovic Greetham

This talk presents a literature review on the topic of Low Voltage (LV) load forecasting. We focus on the current approaches used for LV forecasting, along with applications, available datasets, current trends in the field, and common challenges. The focus of this study is on the hierarchy between household-level and system-level load forecasting. LV systems present unique challenges that are not as prevalent in national or system level demand, such as increased volatility due to lower aggregation of demand, more varied demand as different feeders make up different numbers and types of customers, and less well understood explanatory variables, some of which are discussed in this talk. We also provide a set of recommendations, to help facilitate creation and adoption of a set of guidelines or best practises in the field. As part of this study, we establish an open, community-driven list of the known low voltage level open datasets to encourage further research and development.

## Probabilistic LV load forecasting: forecast fusion and daily peaks

Presenter: Jethro Browell
Co-authors: Jethro Browell;Ciaran Gilbert;Bruce Stephen

Short-term forecasts of energy consumption are invaluable for operation of energy systems, including low voltage electricity networks. However, network loads are challenging to predict when highly desegregated to small numbers of customers, which may be dominated by individual behaviours rather than the smooth profiles associated with aggregate consumption. Furthermore, distribution networks are challenged almost entirely by peak loads, and tasks such as scheduling storage and/or demand flexibility maybe be driven by predicted peak demand, a feature that is often poorly characterised by general-purpose forecasting methods. Here we propose an approach to predict the timing and level of daily peak demand, and a data fusion procedure for combining conventional and peak forecasts to produce a general-purpose probabilistic forecast with improved performance during peaks. The timing and intensity of peaks are modelled using time-to-event analysis and distributional regression in the framework of Generalised Additive Models for Location Scale and Shape (GAMLSS). Conventional forecasts are produced in the same framework, with the normal distribution sufficient for aggregated loads but not individual households. Instead, we propose a regression using the four-parameter Generalised Beta Prime distribution and demonstrate its utility for this application. The proposed approach is demonstrated at smart meter, feeders, secondary and primary substations levels, using real smart meter data and a hypothetical low voltage network hierarchy. Fusing state-of-the-art probabilistic load forecasts with peak forecasts is found to improve performance overall, particularly at smart-meter and feeder levels and during peak hours, where improvement exceeds 10% in terms of CRPS compared to a competitive benchmark.

## Interval Prediction with Deep Learning Models

Presenter: Ghulam Qadir
Co-authors: Ghulam Abdul Qadir;Tilmann Gneiting

Statistical analysis for the purpose of prediction is preferably accompanied by uncertainty quantification, often in the form of prediction intervals. In many applications, deep learning approaches have been extensively shown to provide accurate point predictions. However, the generation of prediction intervals within conventional deep learning models calls for diligent adaptation, development, testing, and implementation. To this end, we propose a novel deep learning model which provides both accurate point predictions and prediction intervals, through empirical score minimization of proper scoring rules for interval forecasts. In simulation studies and real data applications, we demonstrate the efficacy of the novel deep learning model against traditionally used methods.

## GARCH Models Structure of Neural Network and Applications

Presenter: Daniel Adetona Adeniji
Co-authors: Daniel Adetona Adeniji;AYOIGBALA JOSHUA Akanbi

A neural network architect of the Generalized Autoregressive Heteroskedasticity (GARCH) model is developed from a recurrent family of Neural Networks. A comparative study of conventional volatility models and Neural Networks was carried out with visualization analysis using Standard and Poor's 500 index ( S & P 500) . It is was discovered that the remodified neutral network performed better than the conventional volatility models due to the fact that there is no formal systematic model building approach for Neural Networks. Selection criteria such as Akaike Information Criterion (AIC), Bayesian Information criterion and information-based in-sample model selection criteria in selecting Neural Networks for financial time series forecasting. We also report the in-sample and out out- sample forecasting performance.

## An empirical study to compare different types of encoder decoder models for forecasting network traffics

Presenter: Arpita Mukherjee
Co-authors: Arpita Mukherjee;

Building a reliable and efficient network heavily relies on accurate probabilistic forecasts of future traffic demand. Network traffics datasets are multivariate time series and the forecasts are required to be obtained at multiple horizon points. Most of the recent developments of deep learning models in the area of multi-horizon forecasting of multivariate time series are built upon the encoder-decoder architecture, that starts with an Encoder network, which learns a representation of the input sequence and creates a context vector, which is then passed to the Decoder network. We explore a class of deep learning models that uses dilated RNN (recurrent neural network) stacks as both encoder and decoder network. This work is intended to be an exposition to different variants of encode-decoder design that feeds different pieces of information to the decoder. The versions we consider are various combinations of the following four cases, namely — passing the forecasted value from the previous time step, passing the final state of the encoder, passing the final output from the encoder and passing the actual value from the previous step. For the first case the forecasted value from previous time step along with the metadata are used as the input to the current step of the decoder. If the forecast is poor at a certain step then there is a chance that it will affect the subsequent steps of the model. In the second and third approach the only data that is passed along with the metadata is the context vector or output vector. At every time-step this input remains unchanged. Thus if a forecast is poor at a certain stage it is less likely to impact the subsequent steps. The fourth case is 'teacher-forcing', where we use the metadata and the actual value at the previous time-step as the input to the decoder during training. We consider cases where one or more of these inputs are passed to the decoder along with the metadata and compare the model performances on a variety of network traffic datasets.

## Forecasting Tourism in the EU after the COVID-19 Crisis

Presenter: Bozana Zekan
Co-authors: Bozana Zekan;Ulrich Gunter;Egon Smeral

The COVID-19 pandemic led to a severe economic downturn and triggered significant negative impacts on the tourism industry. In terms of tourism forecasting, we had never been confronted with a situation of such a bottomless fall and such strong demand fluctuations as a reaction of the stop and go policy caused by the different infection waves. Presently, the medium-term outlook is very uncertain and studies looking some years ahead covering the development of the tourism and leisure industries are still rare. This is where the present study comes in: we attempt to shed more light about the potential impact on tourism by forecasting the demand in terms of the tourism imports of selected EU countries. For developing a forecast model, we employed a panel pooled Fully Modified Ordinary Least Squares (FMOLS) approach with the aim to develop a modest optimistic "baseline" scenario and a pessimistic "downside" scenario to project demand for foreign travel until 2025. To the best of our knowledge, this is the first study using a

panel pooled FMOLS approach for analyzing and forecasting tourism demand in the medium term based on an aggregate economic framework. Additionally, the results highlight important managerial implications as they point towards the individual countries that could be the fastest to recover and are therefore also of interest to destination managers and other tourism stakeholders.

## How to forecast the tourism business transformation configuration after the pandemic? – by analysing the digital business model transformation of SMEs in the UK

Presenter: Shujun Xiao
Co-authors: Shujun Xiao;Kostas Nikolopoulos;Sarah Hong Xiao

As the Covid-19 pandemic presented extraordinary challenges and opportunities for the tourism industry, the orientation of tourism business model transformation should enable the repair of the wounds suffered from the pandemic in the short term and provide sustainable solutions for long-term development. For the purposes of after-pandemic recovery, the need for forecasting in the tourism industry has also become more urgent than ever. Because as consumer behaviour changes and continuous policies adaptation, the ability of tourism companies to develop new strategies and business model transformation through accurate forecasting determines the company's prosperity. Even before the pandemic, the collapse of Thomas Cook confirmed the importance of forecasting for business transformation. This article aims to develop a model, starting with the RCOV (resource, competences, organisation, and value) framework, and to analyse how small and medium-sized tourism companies in the UK are building social networks and allocating resources through digitalisation after the pandemic. Based on this transformation model, tourism companies can respond more positively to external changes and thus improve the forecast accuracy of tourism demand and policy changes.

## Forecasting Country-Level Airbnb Prices While Respecting the Endogeneity of Demand as Instrumented by a Continuous Treatment

Presenter: Ulrich Gunter
Co-authors: Ulrich Gunter;Francesco Luigi Milone;Bozana Zekan

The proposed study uses monthly data on Airbnb sourced from AirDNA for 44 European countries covering the period from January 2017 to December 2021. This results in a panel dataset comprising 2,640 observations. The forecast variable of interest is average monthly price. Employing time-variant price and demand determinants and respecting the endogeneity and simultaneous determination of both price and demand, the latter is instrumented within a 2SLS-DID framework by a continuous treatment: the COVID-19 Stringency Index. The panel model is estimated using the period January from 2017 to December 2020 as the initial training set. All variables are seasonally adjusted and transformed into natural logarithms before they enter the estimation. The period from January 2021 to December 2021 is used as the test set for which Forecast Evaluation is carried out in terms of typical accuracy measures and based on expanding estimation windows (i.e., recursive forecasting). All covariates – including the continuous treatment – are forecast using appropriate univariate forecast models to ensure the pseudo-out-of-sample Forecast Evaluation to be ex-ante. Naturally, the forecast accuracy of the 2SLS-DID framework is benchmarked against suitable competitors from different univariate and multivariate forecast model classes. The expected result of the proposed study is that the theoretically founded 2SLS-DID framework is able to (statistically significantly) outperform its competitors in a considerable number of cases across accuracy measures and horizons. Finally, the ex-ante evaluation is contrasted with a more conventional ex-post evaluation.

## Forecasting tourism demand amid COVID-19: A time-varying parameter perspective

Presenter: H Song
Co-authors: Haiyan Song;Ying Liu;Long Wen;Han Liu

Background: The COVID-19 is still spreading and the duration of its effects is unclear, tourism demand forecasts are of great important to policymakers, industry professionals and stakeholders. Purpose: This study aims to forecast tourism demand for Hainan province using the daily Baidu Index and monthly economic data, and to delineate their relationship with tourism demand over the period of the COVID-19. Methodology: A time-varying parameter mixed-data sampling (TVP-MIDAS) model with a novel iterated filtering process is used to overcome the limitations of the traditional econometric models. Not only does this new model utilize high-frequency information provided by search query data, but also captures the dynamic relationships between the dependent and independent variables by allowing the model parameters to vary over time. Findings: The simulation results with the time-varying data-generating processes (DGPs) demonstrate the superiority of the iterated filtering algorithm even if the DGP varies over time. The empirical results show that the forecasts of the TVP-MIDAS model significantly outperforms the benchmark models.

## Forecasting the geolocalized imbalance between supply and demand in last-mile logistics

Presenter: Pablo Perez Piskunow
Co-authors: Benjamin Wolter;

One of the biggest challenges for business operations in the last-mile logistics industry is to optimize three competing conditions: improving the couriers' pay, reducing the price tag for our clients and increasing the gross margin for the delivery company - in our case Stuart. Variables (or KPIs) that represent this problem statement, amongst others, are the cost-per-order (CPO), the distance to pickup for a courier, on-time rates, the courier busy rate or the most holistic one of them being deliveries per hour per courier. The majority of tools and features of our product that influence and/or improve these KPIs need to be fed by an accurate, reliable and scalable forecast of the imbalance between supply (S), being the amount of couriers (' availability), and demand (D), being the amount of packages in the platform, as a function of geo location and time bin.

In this session, we present the approaches we take in the Data Science team at Stuart on forecasting the S&D imbalance. Moreover, we'll give an overview of the recommendations and conclusions we derive from that imbalance forecast in order for our business units to tune a multitude of levers that then again influence our couriers' (supply) and our clients' (demand) future behavior. These levers can span from one of the operation's most commonly applied incentives such as the Multiplier, i.e. increase the courier pay temporarily in a specific zone by e.g. 1.5x, to a more drastic one such as e.g. zone closings. Each one of these levers is costly and impacts the platform's performance in a different way making it challenging to measure the effective error of our forecasts.

## Online changepoint detection using forecasts

Presenter: Rebecca Killick
Co-authors: Rebecca Killick;Tom Grundy;Ivan Svetunkov

In many organisations, accurate forecasts are essential for making adequate informed decisions for a variety of applications from inventory management to staffing optimization. Whatever forecasting model is used, changes in the underlying process can lead to inaccurate forecasts, which will be damaging for decision making. At the same time, forecasting models are becoming increasingly complex and identifying change through direct modelling is problematic. We present a novel framework for monitoring forecasts to ensure they remain accurate. By utilizing sequential changepoint techniques on the forecast errors, our framework allows for the real-time identification of potential changes in the process caused by various external factors. We show theoretically that some common changes in the underlying process will manifest in the forecast errors and will therefore be identified faster by identifying changes in the forecast errors than within the original modelling framework. Moreover, we demonstrate the effectiveness of this framework on numerous forecasting frameworks through simulations and show its effectiveness over alternative approaches. Finally,

we present two concrete examples, one from Royal Mail parcel delivery volumes and one from NHS A&E admissions relating to gallstones.

## Forecast Evaluation

Presenter: Jeroen Rombouts
Co-authors: Jeroen Rombouts;Ines Wilms

Optimizing on-demand deliveries is of existential importance to platform businesses, which, given their size, requires planning ahead in real time using automated forecast procedures. Platform businesses that can swiftly adapt to newly incoming demand data streams which exhibit strong seasonality, irregular growth patterns and nonlinear dynamics can obtain a competitive advantage. We propose a fast streaming demand forecast approach tailored towards these characteristic features of platform data and build in a dynamic adaptation of the model when forecast performance starts to worsen due to a rapidly changing environment. We empirically demonstrate the benefits of such an approach on all UK delivery areas of an on-demand last mile delivery platform. We find strong performance gains against several industry benchmarks across different forecast error loss functions. We study managerial implications for the platform business by computing financial metrics based on economic loss, by computing operating costs, and by establishing relationships between size of the delivery areas, their location, market growth, and forecast performance.

## A robust autonomous method for blood demand forecasting

Presenter: Esa Turkulainen
Co-authors: Esa Turkulainen;Merel Wemelsfelder;Mart Janssen;Mikko Arvas

BACKGROUND Blood Supply Chain reliability is largely dependent on the accuracy of the demand estimates. Shortages in fresh blood translate directly to potential loss of life, while oversupply means that blood and its procurement costs go to waste. Improving analytical demand forecasts is a requirement for the automation of blood Supply Chain operations and is expected to result in fewer shortages and outdated blood products.

AIMS In this study, we aim to review the current literature on blood demand forecasting, discuss and determine metrics for evaluating forecasts for the blood Supply Chain, study the trends and changes in demand in Finland, compare method performances over time, devise an autonomous method selection system and test it over a 7 year time span of demand in Finland and the Netherlands.

METHODS We limit ourselves in this study to autoregressive forecasting to enable a broad applicability of the results. We identify a period of structural change in the Finnish weekly demand history, artificially extend it, and test how this change affects method performances using mean absolute percentage errors. We then devise two different method selection heuristics for an autonomous forecasting system and test its ability to change used methods in response to changes in the behavior of the demand. Finally, we compare the autonomous system against individual methods using real, unaltered demand data from Finland and the Netherlands.

RESULTS We find that a shift in the demand signal behavior from more stochastic to seasonal affects the relative performances of the methods chosen for this study. Our autonomous system outperforms all examined individual methods when forecasting the using the altered demand, exhibiting meaningful robustness. When using historical demand in Finland and the Netherlands, the most accurate methods are the autonomous system and the method average, respectively.

SUMMARY The heuristics tested here, while not exhaustive, were able to adjust for changes in the demand while maintaining sufficient forecasting accuracy in the absence of such changes, indicating the viability of an autonomous forecasting solution for blood centers and blood products in general.

## Asthma Monitoring and Prediction

Presenter: Jooyoung Jeon
Co-authors: Jooyoung Jeon;Sehjeong Kim;So-Yeon Kim

Asthma is a non-curable chronic disease with the symptoms such as recurrent attacks of breathlessness and wheezing due to inflammation and narrowing of airways in the lungs. The causes of asthma have not been clearly known and highly vary individual to individual. As many as 334 million people currently suffer from asthma in the world. Since asthma is non-curable, the only way out is managing triggers for asthma attacks under control. Due to high individual uncertainties in the causes of asthma attack, we analyze individual causes from patient's respiratory data, indoor/outdoor air pollution and weather data. Then, we produce forecasts for the likelihood of having abrupt asthmatic events and daily deteriorating respiratory condition for the following day. Eventually, the model we built is expected to provide personalized care and risk factor investigation for asthma patients and the probability for potential asthmatic events, which will reduce health care cost and time spent in hospitals.

## Impact of the COVID-19 pandemic on births' trend in Mexico

Presenter: Eliud Silva
Co-authors: Eliud Silva;Alejandro Aguirre

The COVID-19 pandemic has represented a huge challenge in several aspects such as health, economy and demography, both, at the global and national levels Two of the main demographic variables, fertility and mortality, have received the most attention in research papers. Assessing the pandemic impact on fertility, both in the short-time and med-term, has received less attention in research papers, particularly for developing countries such as the Mexican case. To approximate this impact, we decide to work with the births instead of some standard fertility rate, due to the fact that there are no official yearly figures about population size adjusted by the results coming from the last 2020 Population Census. Given that in Mexico the births trend before the pandemic was decreasing, we have formulated the following questions: Is there any impact such that the downward fertility trend is accelerated? or Does the trend remain without changes?; regarding deaths, Is there any crossover between them and births series due to excess mortality? Is this a common worldwide phenomenon? That is why, the objective of this paper is to answer these questions taking into account the Mexican mother's age and making a multivariate time series model (VAR(p)) to forecast and compare our results, firstly with the observed in 2020 and secondly with the deaths' trend. Our results suggest that for some cases it has presented a downward fertility trend acceleration for specific ages and unsual crossover among deaths and births is ocurring, that it is not a typical behavior in other countries.

## Point and probabilistic forecast reconciliation for general linearly constrained multiple time series

Presenter: Daniele Girolimetto
Co-authors: Daniele Girolimetto;Tommaso Di Fonzo

Hierarchical forecast reconciliation is the post-forecasting process aimed to revise a set of incoherent base forecasts into coherent forecasts in line with cross-sectional/temporal/cross-temporal data structure. In both theoretical and empirical frameworks, most of the point and probabilistic hierarchical forecast reconciliation results move from the classic reconciliation formula valid for the structural representation of a hierarchical time series (Hyndman and Athanasopoulos, 2021, ch. 11). However, this formula holds for genuine hierarchical/grouped time series, sharing both the top and the bottom level variables. When a general linearly constrained multiple time series is considered, the projection approach reconciliation formula (van Erven and Cugliari, 2015) gives a general solution. While it is well known that the classic structural reconciliation formula is equivalent to its projection approach counterpart, it is not obvious if and how a structural-like reconciliation formula may be derived for a general, not genuinely hierarchical time series. Such an expression would permit to extend definitions, theorems and results found by Panagiotelis

et al. (2020) for probabilistic forecast reconciliation in a rather straightforward manner. In this paper, we show that even for general linearly constrained multiple time series it is possible to express the reconciliation formula according to a structural approach that keeps distinct free and basic, instead of bottom and upper (aggregated), variables. Then, we extend the definition of probabilistic forecast reconciliation to a general linearly constrained multiple time series. Finally, we apply the results to obtain a 'one number forecast' for the Australian GDP from Income and Expenditure Sides, in both point and probabilistic settings.

References Hyndman, R.J., Athanasopoulos, G. (2021), Forecasting: principles and practice (3rd ed.), Melbourne, OTexts. Panagiotelis, A., Gamakumara, P., Athanasopoulos, G., Hyndman, R.J. (2020), Probabilistic forecast reconciliation: properties, evaluation and score optimisation, Monash University, Department of Econometrics and Business Statistics, Working Paper 26/20. Van Erven, T., Cugliari, J. (2015), Game-theoretically Optimal Reconciliation of Contemporaneous Hierarchical Time Series Forecasts, in Antoniadis, A., Poggi, J.M., Brossat, X. (eds.), Modeling and Stochastic Learning for Forecasting in High Dimensions, Berlin, Springer, 297-317.

## Hierarchical Mortality Forecasting with EVT Tails: An Application to Solvency Capital Requirement

Presenter: Han Li
Co-authors: Han Li;Hua Chen

In this paper, we propose a new framework to coherently produce probabilistic mortality forecasts by exploiting techniques in extreme value theory (EVT) and hierarchical forecast reconciliation. We are amongst the first to model and analyze U.S. monthly death counts data during the period of 1968-2019 to explore the seasonality and the age-gender dependence structure of mortality. Our results indicate that incorporating EVT and trace minimization forecast reconciliation greatly improves the overall mortality forecast accuracy, which has important implications for life insurers in terms of rate making, reserve setting, and capital adequacy compliance. Using the solvency capital requirement (SCR) under Solvency II as an example, we show that the SCR calculated by our approach is much higher than those calculated by alternative models, suggesting that failing to account for extreme mortality risk and mortality dependence can result in significant underfunded problems for life insurers.

## Cross-temporal reconciliation of solar forecasts

Presenter: Tommaso Di Fonzo
Co-authors: Tommaso Di Fonzo;Daniele Girolimetto

In previous works by Yang et al. (2017a,b), and Yagli et al. (2019), cross-sectional, temporal, and sequential deterministic (point) reconciliation of hierarchical photovoltaic (PV) power generation have been considered for a simulated PV dataset in California. In this paper, (i) the cross-temporal point forecast reconciliation approach is applied to generate fully coherent (both in space and time) day-ahead and hour-ahead forecasts, (ii) some useful relationships between two-step, iterative and simultaneous cross-temporal reconciliation procedures are established, and (iii) non-negativity issues of the final reconciled forecasts are discussed and correctly dealt with in a simple and effective way. In the empiricaal application, the iterative and simultaneous approaches by Di Fonzo and Girolimetto (2021), and the heuristic cross-temporal procedure proposed by Kourentzes and Athanasopoulos (2019) are applied to the forecasts of a hierarchy consisting of 324 hourly series along 3 levels. The results show that the cross-temporal forecast reconciliation approach significantly improves on the standard uni-dimensional reconciliation procedures, while assuring both spatial and temporal coherence, and non-negativity of the reconciled forecasts, at any cross-sectional level of the hierarchy and for any temporal aggregation order.

References Di Fonzo, T., Girolimetto, D. (2021), Cross-Temporal Forecast Reconciliation: Optimal Combination Method and Heuristic Alternatives, International Journal of Forecasting (in press). Kourentzes, N., Athanasopoulos, G. (2019), Cross-temporal coherent forecasts for Australian tourism, Annals of Tourism Research, 75, 393-409. Yagli, G.M., Yang, D., Srinivasan, D. (2019), Reconciling solar forecasts: Sequential

reconciliation, Solar Energy, 179, 391-397. Yang, D., Quan, H., Disfani, V.R., Liu, L. (2017a), Reconciling solar forecasts: Geographical hierarchy, Solar Energy, 146, 276-286. Yang, D., Quan, H., Disfani, Rodriguez-Gallego, C.D. (2017b), Reconciling solar forecasts: Temporal hierarchy, Solar Energy, 158, 332-346.

## What's new in Retail Demand Forecasting?

Presenter: Robert Fildes
Co-authors: Robert Fildes;Stephan Kolassa;Shaohui Ma

Our 2019 review article "Retail forecasting: Research and practice" aimed to summarize the state of research and practice at the time. But time moves on and in retailing there have been rapid developments. These include the COVID-19 pandemic and the substantial new research on machine learning algorithms, when applied to retail. This introduction to the invited sessions on retailing offers new conclusions and challenges for both research and practice in Retail Demand Forecasting

## Evaluating Human Behaviour in Response to AI Recommendations for Judgemental Forecasting

Presenter: Naghmeh Khosrowabadi
Co-authors: Naghmeh Khosrowabadi;Kai Hoberg;Christina Imdahl

Various advanced systems deploy artificial intelligence (AI) and machine learning (ML) to improve demand forecasting. Supply Chain planners need to become familiar with these systems and trust them, considering real-world complexities and challenges the systems are exposed to. However, planners have the opportunity to intervene based on their experience or information that the systems may not capture. In this context, we study planners' adjustments to AI-generated demand forecasts. We collect a large amount of data from a leading AI provider and a large European retailer. Our dataset contains 30 million forecasts at the SKU-store-day level for 2019, plus variables related to products, weather, and holidays. In our two-phase analysis, we aim to understand the adjustments made by planners and the quality of these adjustments. Within each phase, we first identify the drivers of adjustments and their quality using random forest, a well-known ML algorithm. Next, we investigate the collective effects of the different drivers on the occurrence and the quality of the adjustments using a decision tree approach. We find that product characteristics such as price, freshness, and discounts are important factors when making adjustments. Large positive adjustments occur more frequently but are often inaccurate, while large negative adjustments are generally more accurate but fewer in number. Thus, planners do not contribute to accuracy on average. Our findings provide insights for the better use of human knowledge in judgemental forecasting.

## Automatic determination of calendar effects on daily retail time series

Presenter: Philipp Reinhard
Co-authors: Philipp Reinhard;Ulrich Küsters

Demand forecasting based on daily time series allows a faster adoption to changing operational situations. In contrast to weekly data, forecasting daily data requires the inclusion of a large number of day specific influences. These effects are less severe in weekly data due to the aggregation. However, day specific events like public holidays, special days (e.g. Valentine's Day) and payment day effects influence adjacent days. Therefore, demand forecasting for a huge number of series requires the determination of leads and lags. These can be set either manually or as a fixed parameter. We developed a semi-automatic algorithm, which automatically determines the length of lead and lag effects for each calendar effect to accommodate time series specific data generation processes. The algorithm is based on an AIC-search procedure which can be applied to all kinds of forecasting models that incorporate calendar effects by regression components via distributed lags (e.g. dynamic regression and SARIMAX). The algorithm requires only a list containing all relevant events and dates.

## Forecasting metals network synchronization with commodity currencies

Presenter: Nicolas Magner
Co-authors: Nicolas Magner;Nicolas Hardy

We illustrate our algorithm by a small number of time series taken from a large grocery chain where each series covers more than ten years of daily data.

## Impact of Covid-19 on Oil prices volatility and their recovery

Presenter: Nida Siddiqui
Co-authors: Nida Siddiqui;Haslifa Hasim

In financial markets, volatility is a statistical measure of the dispersion of returns around the mean price. The most widely used measure of financial risk is volatility. Crude oil prices play a crucial role in world economics and a high level of volatility in these prices directly impacts the economic stability of importing and exporting nations. Covid-19 has had a major impact on oil prices around the world. In this research we model and forecast crude oil prices volatility of the cash prices of Das Blend UAE, which is a blend of one of the worlds benchmark crudes, Dubai Crude. The research aims to estimate the time it will take for the oil prices to recover from the shock of Covid-19 and reach their historical peak in the absence of any other shocks. The data covers a time period of two years starting from the 1st of January, 2018 till 31st January, 2022. Time series analysis is a very reliable approach for modelling and forecasting volatility. Traditional Time Series Models assume a normal distribution of returns, and hence constant variance of the error terms. But empirical data for crude oil prices does not support this assumption. The returns are not normally distributed (leptokurtic) and hence variance is not constant, instead it keeps on evolving. In this research, we use the GARCH model and its variations from time series, to capture the non-constant variance and produce formidable forecasts. The analysis was carried out using the R studio package, rugarch, which is an R package for univariate GARCH models. Several GARCH models are analysed assuming normal distribution as well as student t distribution of errors. The AIC and BIC criteria are used to choose the best fit model. Applying simulation techniques to the selected model, we try to work out the probabilities of the crude oil prices reaching their historical peak.

## Trading volume as a predictor of risk and return of the Swiss SMI: A quantile regression approach

Presenter: Erhan Uluceviz
Co-authors: Erhan Uluceviz;Harald Schmidbauer;Angi Roesch

Core concepts used to explain non-normal behavior of stock price changes (returns) refer to the varying amount of information flow into the markets, often proxied by the trading volume. This idea is referred to as the mixture of distributions hypothesis (MDH). On the other side, the market efficiency hypothesis suggests that trading volume information fails to predict future price changes. In our research, inspired by MDH's main line of arguments, we investigate whether trading volume information has any predictive content for assessing tomorrow's distributions of return. To this end, we link daily volume and return data using quantile regression methodology. Our subject of study is the Swiss Stock Exchange blue chip market, represented by its index SMI. Our results suggest that, within the quantile regression domain, trading volume information can predict future return behavior, especially the tails of the return distribution. The predictive power of volume is strongest when filtered, that is, when it reflects long-term information flow.

## Disentangling supply and demand effects on business expectations during the pandemic

Presenter: Xuguang Simon Sheng
Co-authors: Xuguang Simon Sheng;Brent Meyer;Brian Prescott

Utilizing the monthly Federal Reserve Bank of Atlanta's Business Inflation Expectations (BIE) survey, which has been continuously collecting subjective probability distributions over own-firm future unit costs, we build a proxy for the inflation expectations of businesses. We document two facts about firms' inflation expectations and uncertainty during the COVID-19 pandemic. First, in the early months of the pandemic, firms, on net, saw COVID-19 as largely a demand shock and lowered their one-year ahead inflation expectations. However, as the pandemic wore on, firms' one-year ahead inflation expectations appear to have risen sharply alongside their views on Supply Chain and operating capacity disruption. Second, firms appear to have been fairly confident about the likely impact of the pandemic on their future unit costs. During the course of the pandemic firms' view flipped from anticipating low unit cost outcomes to increasingly higher outcomes, leaving unit cost uncertainty largely unchanged.

## A Neural Phillips Curve and a Deep Output Gap

Presenter: Philippe Goulet Coulombe
Co-authors: Philippe Goulet Coulombe;

Many problems plague the estimation of Phillips curves. Among them is the hurdle that the two key components, inflation expectations and the output gap, are both unobserved. Traditional remedies include creating reasonable proxies for the notable absentees or extracting them via some form of assumptions-heavy filtering procedure. I propose an alternative route: a Hemisphere Neural Network (HNN) whose peculiar architecture yields a final layer where components can be interpreted as latent states within a Neural Phillips Curve. There are benefits. First, HNN conducts the supervised estimation of nonlinearities that arise when translating a high-dimensional set of observed regressors into latent states. Second, computations are fast. Third, forecasts are economically interpretable. Fourth, inflation volatility can also be predicted by merely adding a hemisphere to the model. Among other findings, the contribution of real activity to inflation appears severely underestimated in traditional econometric specifications. Also, HNN captures out-of-sample the 2021 upswing in inflation and attributes it first to an abrupt and sizable disanchoring of the expectations component, followed by a wildly positive gap starting from late 2020. HNN's gap unique path comes from dispensing with unemployment and GDP in favor of an amalgam of nonlinearly processed alternative tightness indicators – some of which are skyrocketing as of early 2022.

## Commodity jump tail risk as predictor of inflation

Presenter: Laurent Ferrara
Co-authors: Laurent Ferrara;Athanasios Triantafyllou;Katerina Karadimitropoulou

We construct a commodity jump tail risk index capturing comovement in the price jump tail risk of major agricultural commodity futures markets. We show that our commodity jump tail risk index contains significant in-sample and out-of-sample forecasting power on US inflation when controlling for major well-established determinants of inflation like the Philips curve, oil prices and exchange rates. Surprisingly, while the findings in the literature have extensively shown the inflationary effect of various types of commodity price shocks on the US economy, we find that the rising commodity jump tail risk predicts subsequent drops in US inflation. Our results implicitly reveal, for the first time in the literature, that the increased synchronization of commodity price jumps is not linked with largely idiosyncratic commodity supply shocks but with shocks related with aggregate demand (for agricultural commodities). Our commodity jump tail risk factor outperforms autoregressive and Philips curve models when used for forecasting US core and headline inflation for medium and long-term forecasting horizon. Moreover, our commodity jump tail risk forecasting model has superior Inflation Forecasting power when compared to a battery of bivariate and multivariate Inflation Forecasting models suggested in the literature. Finally, the Bayesian Dynamic Model Averaging approach for forecasting inflation shows that best performing model is the one which includes our commodity jump tail risk factor as inflation predictor for most of the time period under examination, covering January 1987 to December 2017, with its time varying predictive power increasing significantly (when compared to that of well-known inflation determinants) after the 2008 Great recession.

## Environmental surveillance as a complement to clinical case data during COVID-19 and the future of wastewater-based epidemiology

Presenter: Stephanie Brennhofer; Mami Taniuchi
Co-authors: Stephanie Brennhofer;Mami Taniuchi

Environmental surveillance (ES), commonly referred to as wastewater-based epidemiology (WBE), became an increasingly popular way to track SARS-CoV-2 throughout populations during the pandemic. However, the process and challenges of carrying out ES in low- and middle-income countries differ substantially from high-income countries. We will provide an in-depth look at the analyses we have conducted for our ES study in Dhaka, Bangladesh and the model we developed to forecast clinical cases based on prior ES and clinical case data. We will discuss how ES can be used as a complementary tool to clinical case data, allowing for comprehensive and timely pathogen surveillance. Additionally, we will look beyond ES for the current pandemic and how it can be used as long-term surveillance for pathogens we aim to eliminate (e.g. poliovirus) and contain (e.g. Salmonella typhi and Vibrio cholera).

## Model-based assessment of Covid-19 epidemic dynamics by wastewater analysis

Presenter: Atte Aalto
Co-authors: Atte Aalto;

Continuous surveillance of Covid-19 diffusion remains crucial to control its spread and to anticipate infection waves. Detecting viral RNA load in wastewater has been suggested as an effective approach for epidemic monitoring. However, the quantitative link between wastewater measurements and the epidemic status is still elusive. Modelling is thus crucial to address this challenge. I present a novel mechanistic model-based approach to reconstruct the complete epidemic dynamics from SARS-CoV-2 viral load in wastewater. The approach integrates noisy wastewater data and daily case numbers into a dynamical epidemiological model by a Kalman filter. I will also discuss the role of our forecasts in Luxembourg's fight against the pandemic.

## Identifying opportunities for environmental surveillance in areas without centralized sewer systems

Presenter: Jillian Gauld
Co-authors: Jillian Gauld;

Environmental surveillance in many locations involves sampling from centralized sewer systems. In low- and middle- income countries, centralized sewer systems may not exist, or they may exclude large portions of the population. In these locations, there is a challenge in identifying what environmental sources should even be sampled. This talk will walk through an example in Blantyre, Malawi, where typhoid fever was transmitting widely. Through multiple methodological approaches, river systems were identified as a source of exposure, leading to the need to confirm the presence of S. Typhi in rivers through environmental surveillance. These river systems have been valuable as a source of wastewater for identifying S. Typhi and more recently, monitoring SARS-Cov-2.

## Integrating text analysis in electricity load forecasting: evidence from UK electricity load

Presenter: Yun Bai
Co-authors: Yun Bai;

The relationship between electricity load and weather has been established since long time and is one of the cornerstones in load prediction for operation and planning, along with behavioural and social aspects such as holidays or major events. This work presents the initial findings of research into the use of

Natural Language Processing (NLP) for extracting previously unused information embedded into news or other forms of unstructured textual data for improving electricity demand forecasts.

In this work, valuable sentiment and topic information is extracted from news texts with methods such as TextBlob and Latent Dirichlet Allocation (LDA) respectively. Multivariate time series are constructed by combining both text information and relevant numerical features (e.g. the number of news items per day). Granger causality test assists in finding which external features contribute to the forecasting of electricity load. At this point different time series forecasting approaches such as Autoregressive Integrated Moving Average Model, Support Vector Regression or Dilated Convolutional Neural Network are benchmarked.

These methods are applied on a news dataset collecting from the BBC over the period between Jul 2015 and Dec 2021, and electricity load in the UK from the ENTSOe transparency platform.

The results of this analysis are fed into a state-of-the-art load prediction model based on randomised decision trees. The performance over several horizons from days to months ahead is analysed according to a series of metrics for deterministic, probabilistic and ensemble forecasting. A comparison is made between two series of models, using or not using the textual features identified in the previous step. The objective is to understand under which conditions the proposed approach can improve the performance of the predictions.

## Deep learning for VAR modelling and forecasting

Presenter: Xixi Li
Co-authors: Xixi Li;Jingsong Yuan

The vector autoregressive (VAR) model has been used to model dependence within and across several time series. In this study, we propose a class of new approaches that employ deep learning methodology to generate the parameters and add a trend to the VAR model. A Long Short-Term Memory (LSTM) network is used for this purpose. To ensure the stability of the model, we enforce the causality condition on the autoregressive coefficients using the transformation of Ansley and Newbold (1979). We provide simulation studies of the estimation ability of the proposed approaches using realistic parameters and trends generated from data. To demonstrate the practical value of our approach, we compare its forecasting performance with existing methods using some real data sets.

## Predicting the solutions of optimisation models via machine learning

Presenter: Mahdi Abolghasemi
Co-authors: Mahdi Abolghasemi;

In this talk, I will explore the applications of machine learning in predicting the solutions of constraint optimization problems. I will discuss the role of loss functions in machine learning algorithms for optimising various decisions and show that by a good choice of loss function we can train machine learning models that not only provide near-optimal solutions but also take into account the constraints of the models to large extent. We implement our approach on a blood Supply Chain case study where a network of hospitals are required to transfer blood units to meet their demand and improve their resiliency. Our results show that while using a light gradient boosting machine model with the mean absolute deviation criterion as loss function provides near-optimal solutions with only 2% higher total costs than a mathematical model, it reduces transhipment and shortage costs by 27% and 6%, respectively.

## The UFO Project

Presenter: Len Tashman
Co-authors: Len Tashman;Jim Hoover

While our methods of business forecasting have advanced dramatically during the past several decades, their value has been poorly understood by many organizations and consequently their substantial benefits are underutilized. Evidence for this has mainly been anecdotal. The UFO project was initiated to learn how

widespread the neglect of systematic forecasting methods (SFM) is, what the barriers to implementation of SFM have been and how these can best be overcome. Many firms, of all sizes according our survey results, persist in relying on ad hoc, judgmental forecasting approaches., despite the demonstrated superiority of systematic methods. How do we extend these methodological benefits and collateral know-how to such organizations? For the UFO project to date, we have undertaken two surveys and held 20 interviews held with forecasting practitioners and consultants. Jim Hoover will highlight the survey results and Paul Goodwin the insights from the interviews. Here I'll briefly mention what I consider from my perspective as Foresight Editor for many years to be the advances made in forecasting methods and processes that are of most practical value to the business world.

## Findings from the Foresight Survey and Implications

Presenter: Jim Hoover
Co-authors: Jim Hoover;

Recently, Foresight, The International Journal of Applied Forecasting, completed analyses of two surveys of forecasting practitioners. These surveys were developed to understand the extent of use of systematic forecasting methods by forecasters across a range of organizations. This discussion will present some key findings from those surveys and discuss some important conclusions for the future of the forecasting practice.

## Forecasting in Organizations: Findings from Semi-structured Interviews.

Presenter: Paul Goodwin
Co-authors: Paul Goodwin;Fotios Petropoulos;Spyros Makridakis;George Athanasopoulos

The Understanding Forecasting in Organizations (UFO) project aims to discover how forecasting is conducted in organizations, and the opportunities there are for achieving greater accuracy and its associated benefits. As part of this project, in-depth semi-structured interviews were conducted with thirteen forecasters and seven consultants involved in forecasting in organizations in North and South America and Europe. The interviews revealed that forecasting serves multiple purposes, including production and capacity planning, asset valuation and meeting the requirements of regulators or stock markets. However, many organizations do not measure accuracy, and in some cases, they do not regard it as the most important objective. Nevertheless, many interviewees identified benefits that could be achieved through enhanced accuracy, though it is usually difficult to put a monetary value on them. Several people reported that end-users often misunderstand what forecasting can achieve, and 'glitzy' pitches by software salespeople can sometimes distort their expectations. Many organizations rely on spreadsheets to produce forecasts or employ thirty-year-old legacy software. Judgmental inputs to forecasts are ubiquitous, but uncertainty assessment is rare. Barriers to improved accuracy include organizational politics, lack of available information, persistent use of less reliable data, untrained or unmotivated forecasters, lack of belief in the benefits of forecasting, continued use of inappropriate software, misuse of software, and lack of senior management support.

## Are Forecasters All Alike? Evidence from Three Decades of Growth Forecasts, 1990 to 2020

Presenter: Martin McCarthy
Co-authors: Martin McCarthy;Zidong An;Umberto Collodel;Prakash Loungani

We examine the quality of growth forecasts from the private sector (Consensus Forecasts) and major public institutions (IMF, World Bank, European Commission) for a large sample of countries over the last three decades. We find that forecasts from these sources exhibit a very high degree of collinearity. In particular, the evolution of forecasts during recession years is fairly similar across sources; over the period 1990 to 2019, there is strong serial correlation in forecast revisions during recession years. Private and

public forecasters both appear reluctant to recognize the magnitude of recessions, and forecasts approach the outcomes only toward the end of the recession year. This pattern, however, was broken during the COVID-19 pandemic, with Consensus and IMF mid-year 2020 forecasts 'overshooting' the outcome and then backtracking. We show that the bias in growth forecasts comes from the slow revision of forecasts during recession years; in normal times, forecasts are unbiased. We also investigate claims that IMF growth forecasts exhibit optimism when countries are in IMF loan programs, reflecting pressures to show that the programs will not hurt growth. We do find that IMF forecasts tend to be too optimistic for large programs; however Consensus Forecasts in these cases also exhibit optimism, which is puzzling since the private sector does not face the same pressures.

## A General Procedure for Localizing Strictly Proper Scoring Rules

Presenter: Ramon de Punder
Co-authors: Ramon de Punder;Cees Diks;Roger Laeven;Dick van Dijk

Forecasters are typically not equally interested in all possible realizations of a random variable under scrutiny. Financial risk managers, for instance, usually put relatively more weight on regions of extreme losses. In density forecast comparison, it is common practice to use strictly proper scoring rules to rank a collection of candidate predictive distributions (Gneiting and Raftery, 2007, Journal of the American Statistical Association 102(477), 359-378). When focusing on a region of interest, however, weighted scoring rules obtained via conditioning are no longer strictly proper (Holzmann and Klar, 2017, Annals of Applied Statistics 11(4), 2404-2431). We develop a general procedure for focusing, i.e., localizing, scoring rules in a way that preserves their strict propriety.

A critical insight of our paper is that censoring observations outside the region of interest, as opposed to conditioning, retains just enough information about the original distribution to maintain strict propriety. Our procedure provides a myriad of strictly locally proper scoring rules beyond the censored likelihood score (Diks, Panchenko and Van Dijk, 2011, Journal of Econometrics 163(2), 215-230). Based on a collection of popular scoring rules, including the Logarithmic, Spherical, Quadratic and Continuously Ranked Probability Score (CRPS), Monte Carlo simulations align with the intuition that censoring is power-enhancing, especially if the number of expected tail observations is small.

## The Anatomy of Out-of-Sample Forecasting Accuracy

Presenter: David Rapach
Co-authors: David Rapach;Daniel Borup;Philippe Coulombe;Christian Montes Schütte;Sander Schwenk-Nebbe

The relevance of individual predictors in forecasting models—especially "black box" models such as Neural Networks or random forests from machine learning—is typically analyzed using a variable importance measure applied to a fitted model based on data from the in-sample period (or training sample). However, the importance of individual predictors in the fitted model over the in-sample period does not necessarily correspond to their importance with regard to out-of-sample forecasting accuracy. We develop a performance-based Shapley value (PBSV) metric for measuring the contributions of individual predictors in any fitted model to the out-of-sample loss, thereby allowing us to identify the predictors that are most responsible for a model's out-of-sample performance (whether good or bad), which, from a forecasting perspective, is more informative than the in-sample variable importance measure. In empirical applications involving forecasting inflation and the aggregate stock market excess return, we find substantial discrepancies between individual predictor relevance according to the conventional in-sample variable importance measure and our new PBSV metric. We use simulations to analyze potential sources of differences across in-sample and out-of-sample variable importance, including benign overfitting, evolving volatilities, and structural breaks.

## Revisiting Forecasting for Emergency Department Staffing

Presenter: Shixuan Wang
Co-authors: Shixuan Wang;Bahman ROSTAMI-TABAR;Stephen Disney

Demand forecasting of healthcare services represents a vital component of healthcare planning. The ability to accurately forecast the demand for a service allows for more accurate staffing schedules on any given shift. Accurate staffing levels can reduce congestion in the service, improves the quality of provision for patients, reduces pressure on staff, decreases waiting time and reduces costs for the health system. The uniqueness of our forecasting method is that we aim to provide a forecasted range, rather than a precise number. This is strongly motivated by the ultimate purpose of this study which is to support the planning of the emergency department. In practice, service leads typically know what cut-off point does the changing demand alter the decision on staffing allocation, and the planning is not going to change within a specific range. Therefore, we first transform a given time series of hospital admissions into a categorial time series based on a given ratio, then we propose a Probabilistic Forecasting model that considers various predictors (including historical categories, lag and lead categories, holidays, shift of the day, day of week, month of year and the interaction between holidays, shift and a lead and lag effect) to accurately forecast the outcome category and their related probabilities. Then we benchmark the accuracy of the proposed model against Naïve, Exponential Smoothing State Space (ETS), Autoregressive Integrated Moving Average (ARIMA) and Exponential smoothing state space model with Box-Cox transformation, ARMA errors and Trend and Seasonal components (TBATS) models. Based on our data, our proposed method has superior predictive power and outperform the four benchmark methods in terms of the Brier Score and the ranked probability score. We further provide tangible suggestions for healthcare planning and staffing based on our Probabilistic Forecasting model.

## Probabilistic Forecasting of hourly Emergency Department arrivals

Presenter: Bahman Rostami-Tabar
Co-authors: Bahman Rostami-Tabar;Jethro Browell;Ivan Svetunkov

An accurate forecast of Emergency Department (ED) arrivals by an hour of the day is critical to meet patients' demand. It enables planners to match ED staff to the number of arrivals, redeploy staff, and reconfigure units. This can have many advantages for healthcare staff and the quality of care delivered to patients. In this study, we develop an innovative model based on Generalised Additive Models and an advanced dynamic model based on exponential smoothing to generate an hourly probabilistic forecast of ED arrivals. We compare the forecast accuracy of these models against appropriate benchmarks, including TBATS, Poisson Regression, Prophet, and simple empirical distribution. We use Root Mean Squared Error (RMSE) to examine the point forecast accuracy and assess the forecast distribution accuracy using Quantile Bias, PinBall Score and Pinball Skill Score. Our results indicate that the proposed models outperform their benchmarks for point and probabilistic forecasts. Our developed models can also be generalised to forecast hourly arrivals in other services such as hospitals, ambulances, or clinical desk services.

## Using Bayesian Network Model to Explore the Influencing Factors of Incident Reporting in Healthcare

Presenter: Salma Albreiki
Co-authors: Mecit Can Emre Simsekler;Salma Albreiki;Abroon Qazi;Ali Bouabid

Several organizational factors may impact the incident reporting practice in healthcare. However, limited research is available on probabilistic interdependencies between organizational factors and reporting practice. To explore this, we adopt a data-driven Bayesian Belief Network (BBN) model to represent a class of probabilistic models, using the hospital-level aggregate survey data from U.K. hospitals. Leveraging probabilistic dependence models and visual features in the BBN model, the results shed new light on relationships among various organizational factors and incident reporting. This study provides a new application to forecasting and prediction in the incident reporting context in healthcare. Further, it offers

significant insights to understand organizational factors' role and their relative importance in supporting decision-making and safety improvements in healthcare.

## The value of hierarchically aligned forecasts for staff scheduling

Presenter: Benedikt Sonnleitner
Co-authors: Benedikt Sonnleitner;Simon Hoelck

Hierarchical forecasting provides value for decision-making by improving accuracy and guaranteeing aligned forecasts (i.e., the sum-constraints in the hierarchy are fulfilled). We examine temporal hierarchical forecasting for staff scheduling in cross-docking and offer a threefold contribution: (i) We quantify the value of aligned forecasts for decision-makers on multiple planning levels. (ii) We investigate the value of accuracy improvements for decision-making. (iii) We offer a new prescriptive solution, by integrating freight demand forecasting with staff scheduling for cross-docking. The analysis is performed using real data from a German carrier.

## An Importance Sampling algorithm for probabilistic reconciliation

Presenter: Lorenzo Zambon
Co-authors: Lorenzo Zambon;Giorgio Corani

Time series are often organized in hierarchies or temporal hierarchies. In both cases, the forecasts should be coherent, that is, they should satisfy the constraints given by the hierarchy. The base forecasts computed independently on each time series are however incoherent. The problem of finding coherent predictions starting from the base forecasts is called reconciliation. We consider the probabilistic case, in which forecasts are in the form of probability distributions. The reconciled distribution is obtained by conditioning on the set of constraints. We use Importance Sampling (IS) to sample from the reconciled distribution. We use the base distribution on the bottom level as a proposal distribution for IS, while we compute the importance weights based on the distribution on the upper levels. However, the standard IS is not usable when the size of the hierarchy grows, or the incoherence gets large. In this case, the effective sample size drops dramatically and we cannot recover reliably the reconciled distribution. To overcome the problem, we propose a new algorithm called Bottom-Up Importance Sampling. Assuming independence between base forecasts, reconciliation is performed by iteratively conditioning on each upper observation, from the bottom to the top. In this way, a (possibly) large importance sampling task is split into several one-dimensional tasks. Our algorithm can reconcile base forecasts of any type: Gaussians (in which case the reconciled distribution has the same mean and variance of MinT), negative binomials (in the case of count time series), etc. Moreover, it can reconcile base forecasts that are only available in the form of samples. The algorithm is very fast and able to efficiently sample from the true reconciled distribution even in the case of big hierarchies or large incoherence levels.

## Hierarchies Everywhere - Bayesian Hierarchical Forecasting

Presenter: Ross Hollyman
Co-authors: Ross Hollyman;Fotios Petropoulos;Mike Tipping

The topic of hierarchical forecasting has generated much recent academic interest, and researchers have demonstrated its empirical validity in many settings. Several authors have extended these techniques to the probabilistic realm, although most ignore the 'Elephant in the Room' - the significant parameter uncertainty inherent in existing techniques. Bayesian approaches have been successfully applied in other parameter-rich time series settings such as vector auto-regressions, and are a natural way to model hierarchical time series. In this paper we propose a Bayesian solution to the reconciliation problem which uses valuable prior information embedded in the hierarchical structure to shape the reconciled forecasts.

## Development of censored ExponenTial Smoothing models for lost-sales demand forecasting

Presenter: Diego J. Pedregal
Co-authors: Diego J. Pedregal;Juan R. Trapero;Enrique Holgado

Supply Chain management efficiency depends crucially on the accuracy of demand forecasts. Very often sales data are used as a good approximation to the true demand, even though it is well-known that such approximation is not as good as it may seem in lost-sales contexts. In these cases, crucial elements for the inventory management, such as safety stocks or reorder points, are miscalculated. The main drawback in this situation is that sales underestimate the actual demand and thus bias all estimates that do not take into account the level of censoring. The literature about censored demand forecasting is scarce, with some solutions that only deals with rather limited versions of the problem. In this paper, a censored general ExponenTial Smoothing algorithm is developed to estimate correctly demand when only sales are available in contexts where lost sales are known to occur. The algorithm is based on a general dynamic linear innovations state space system with censoring levels at the output. Once the model is set up, estimation of parameters and initial conditions are estimated as usual by Maximum Likelihood. Such generality provides a solution with several advantages. First, the solution is general enough to be able to implement other methodologies straightforwardly, like time-varying regression or ARIMA models with censoring levels. Second, time-varying censoring levels may be considered directly in the solution as is. This is actually the usual situation in real life, where censoring levels depend on forecasts themselves and, consequently, change over time. Finally, models with all sort of components (like trend, seasonal, coloured noise, exogenous variables) can be used. The elegant solution provided overcomes certain limitations of previous contributions, which tend to focus on simple assumptions about demand components and constant censoring. Such solutions imply analytical developments for each model, whereas the proposed solution is broadly general. The methodology is shown working on simulations, as well as artificial data and real examples. The results demonstrate the suitability of the methodology even in cases where severe censoring levels occur, e.g., below average demand.

## A data-driven stock control optimization framework

Presenter: Evangelos Theodorou
Co-authors: Evangelos Theodorou;Evangelos Spiliotis;Vassilios Asimakopoulos

Stock control is a key aspect in retail operations. Deciding for each item when an order should be placed, how many units should be ordered, and what the target service level should be, are just some of the parameters that retailers must define on a regular basis. Although shorter review periods and larger orders typically increase product availability and improve customer experience, they also have a negative impact on inventory costs (holding and ordering). In that end, a challenging trade-off emerges, requiring inventory policies that carefully balance the parameters described above. Various numerical approaches of optimal control theory have been proposed in the literature to define such parameters. Although theoretically effective, these approaches can become extremely time consuming when the dimensions of the optimization problem grow. A promising alternative is the incorporation of machine learning algorithms that are capable of learning from large data sets what the best practices are and applying such knowledge to new cases to provide fast, yet accurate recommendations. In this study, we propose a data-driven framework that optimizes replenishment at warehouse level. First, a reference set of items is used for the machine learning models employed to learn how inventory performance is affected by demand patterns and key inventory policy parameters. Then, the extracted knowledge is exploited to optimize the inventory policy parameters of the target items. Effectively, the reference and target items may not be the same, thus enabling transfer learning and coping with cases where data availability is limited. Moreover, the proposed framework is model-independent in nature in terms of the methods used for forecasting demand and predicting costs. We use part of the M5 competition data set and indicative benchmarks to evaluate the performance of the proposed framework. Our results suggest that the proposed approach can significantly improve service level, while reducing total inventory cost.

## Sales forecasting with machine learning methods using external information for the retail Supply Chain

Presenter: Patrícia Ramos
Co-authors: Patrícia Ramos;José Manuel Oliveira;Mariana Cardoso Teixeira

The increasing complexity of Supply Chains reinforces the need for retailers to rely on sales forecasts to plan their activity and make decisions. The greater the accuracy of these forecasts, the greater the efficiency of Supply Chain management. The forecasting method adopted by the retailers must consider factors such as implementation and computational requirements, although highly accurate forecasts are mandatory to allow reducing losses, avoid inefficient stock management and mitigate customer dissatisfaction. Retailers need to forecast sales at the various levels of their internal organization in order to make appropriate decisions, whether at the store, distribution center or chain level but also at the product category, brand and SKU level. This concept of sales forecasting at different levels of aggregation relies on hierarchical forecasting. Forecasts can be generated individually for each level of the hierarchy, but this solution does not reflect the hierarchical structure of the retailer and does not guarantee that the sum of the forecasts at the disaggregated level is equal to the forecast at the corresponding aggregate level. However, this consistency between forecasts at different levels is necessary so that retailers can make appropriate decisions along the entire Supply Chain. Alternatively, hierarchical forecasting methods can be used to ensure consistency between forecasts at different levels of the hierarchy and additionally to improve the accuracy of those forecasts if the chosen forecasting method is appropriate. As a result of the growing competition in the retail sector, the adoption of promotional activities has become increasingly frequent, which implies the use of forecasting models that take into account the effect that these activities have on sales. In fact, the integration of variables, such as promotional information in the sales forecasting gains relevance since it increases accuracy. In this study we integrate promotional activity into hierarchical forecasting. Machine learning methods are used to generate base forecasts for the different levels of the hierarchy, namely Random Forest, eXtreme Gradient boosting, Support Vector Machines and Long Short-Term Memory which in recent studies have shown better performance compared to traditional methods. Working with a large dataset from a leading Portuguese retailer, we develop and test these modeling approaches to forecast the demand.

## Macroeconomic Extrapolation, Machine Learning, and Equity Risk Premium Forecast

Presenter: Yufeng Han
Co-authors: Yufeng Han;Yueliang Lu

We present a macroeconomic trend extrapolation approach that utilizes all economic fundamentals of different time periods simultaneously in the aggregate market. A simple pooling combination statistically and economically outperforms the historical average that assumes a constant equity risk premium, as well as Rapach, Strauss, and Zhou's (2010) mean combination that ignores the historical information in the macroeconomics. We further find that extrapolating via neural network produces out-of-sample R2 statistic as high as 4% and generates substantial economic value. Extrapolating economic fundamentals with a grid of moving averages more closely tracks important macroeconomic fluctuations and more effectively regulates the forecast variability, thereby generating superior and robust forecasting gains consistently over time.

## Forecasting the volatility of U.S. oil and gas firms: beyond linearity and volatility factors

Presenter: Erwin Hansen
Co-authors: Erwin Hansen;Juan Díaz;Gabriel Cabrera

This paper forecasts the daily volatility of U.S. oil and gas firms using an extended version of the model introduced by Lyócsa and Todorova (2021). We show that forecast accuracy can be significantly

improved by relaxing the linearity assumption and using several non-linear machine learning techniques. Furthermore, we demonstrate that besides Lyócsa and Todorova's (2021) volatility factors, economic variables are valuable predictors of firm volatility. Technical indicators also improve forecasting accuracy, albeit to a lesser extent. Finally, we also compare predictions using an economically motivated Value-at-Risk loss function as well as the more standard statistical loss functions.

## Hurst Exponent Dynamics of SandP 500 Returns: Implications for Market Efficiency, Long Memory, Multifractality and Financial Crises Predictability by Application of a Generalized Nonlinear Dynamics Analysis Framework

Presenter: Markus Vogl
Co-authors: Markus Vogl;

In this study, I apply a novel, generalized nonlinear dynamics analysis framework for non-stationary and nonlinear time-series. I conduct a rolling window approach to wavelet-filtered (denoised) S&P500 returns (2000–2020) to obtain time varying Hurst exponents. I analyse the dynamics of the Hurst exponents by applying statistical tests (e.g. for stationarity, Gaussianity and self-similarity), a recurrence quantification analysis (RQA), a wavelet multi-resolution analysis (MRA) and a multifractal detrended fluctuation analysis (MFDFA). Moreover, I discuss implications of Hurst exponent dynamics such as the complete invalidity of the efficient market hypothesis (EMH), long memory, multifractal properties, an explicative rationale for momentum crashes and potentials for crises predictability. Furthermore, forecasting attempts based on the determined dynamical traits are conducted with stochastic processes (e.g. multifractal Brownian motions) and Neural Networks (e.g. deep learning multiperceptron models [DL-MLPs] or wavelet Neural Networks [WNNs]). Besides, I display empirical findings taken out of the academic literature and critically elaborate on the impact of my findings and future prospects.

## Monitoring daily unemployment at risk

Presenter: Ignacio Garron
Co-authors: Ignacio Garrón;Helena Chuliá;Jorge Uribe

Using a high-frequency framework, we show that the Auroba-Diebold-Scotti (ADS) daily business conditions index significantly increases the accuracy of U.S. unemployment nowcasts in real-time. This is of particular relevance in times of recession, such as the Global Financial Crisis and the Covid-19 pandemic, when the unemployment rate is prone to rise steeply. Based on our results, the ADS index presents itself as a better predictor than the financial indicators widely used by the literature and central banks, including both interest and credit spreads and the VXO.

## Forecasting with subset regularization

Presenter: Bonsoo Koo
Co-authors: Bonsoo Koo;Didier Nibbering

This paper proposes a subset regularization method that partitions the predictor set into strong, weak, and noise predictors. We show that the asymptotic mean squared forecast error is minimized when strong signals are included in the forecasting model without regularization, and predictors identified as noise are excluded from the model. The weak signals are subject to a bias-variance trade-off, which is handled by regularization. An empirical application to forecasting U.S. inflation and industrial production shows that the common approach of preselecting unregularized predictors in macroeconomic forecasting models, is outperformed by subset regularization.

## Mapping the denominator: Predicting small area population distributions, demographics and dynamics for health applications

Presenter: Andrew Tatem
Co-authors: Andrew Tatem;

Accurate, timely and detailed data on human populations and their characteristics are needed to support disease surveillance, monitor progress towards health goals, and assess the impacts of disease outbreaks and policy decisions. Unfortunately, recent, reliable and complete data on population distributions, demographics and dynamics can be lacking in many settings around the World. Population and housing censuses are typically conducted only once a decade, registries can be incomplete and inaccurate and household surveys provide a broad snapshot. The advent of detailed and timely data from satellites and cellphones however, as well as improvements in computing power and algorithms to extract information from them, is providing opportunities to complement traditional demographic data sources. Prof Tatem will discuss how geospatial data integration and statistical modelling approaches can produce high resolution and timely data on populations in resource poor settings, and how these are being used by governments, scientists and international agencies to plan intervention campaigns, respond to disease outbreaks and improve health metrics.

## The benefit of clustering for cross-learning models on M5 data

Presenter: Claudia Ehrig
Co-authors: Claudia Ehrig;

Cross-learning can enhance prediction accuracy and save computational resources in comparison with series-by-series learning. Its benefit is however depending on the similarity of the time series used. Hence, we propose to cluster the data beforehand and then cross-learn one model per cluster. We benchmark different clustering methods with respect to the forecasting accuracy, runtime, and CO2 emissions of a simple MLP trained on those clusters. We use data from the M5 competition.

## Machine Learning forecasts – confirmation bias or value add?

Presenter: Anne Flore Elard
Co-authors: Anne-Flore Elard;

“If dataset A can be ingested on top of the data already used by the Machine Learning models, the forecast results would be 10% more accurate right?. Then if I add dataset B, it will add 5% additional accuracy so with the whole data we should get 95% accuracy.” “Why would I need to spend time on basic statistical forecast when I can go straight to Machine Learning models?” Here are two sample quotes I hear almost every day rooted in the concept that Machine Learning models applied to forecasting are superior to statistical forecasting models in that they are intrinsically more accurate. In this perhaps iconoclastic discussion, I will explore the value-add of Machine Learning for industry-based forecasts. First, let’s look at some research and review the accuracy obtained from different methods in the case of highly random data or specific patterns. Second, as a practitioner, I will spend time discussing the cost component of Machine Learning implementations, both quantitatively and qualitatively with interpretability. Last but not least, I will highlight some lessons learnt over the years on how to effectively implement Machine Learning methods to industry-focused forecasts”

## Forecasting Italian GDP Growth with Epidemiological Data

Presenter: Valentina Aprigliano
Co-authors: Valentina Aprigliano;Alessandro Borin;Francesco Paolo Conteduca;Simone Emiliozzi;Marco Flaccadoro;Sabina Marchetti

The Covid-19 epidemic affected the ability of traditional forecasting models to produce reliable scenarios for the evolution of economic activity. We combine macroeconomic variables with epidemiological indicators to account for the Covid-19 shock and predict the short-term evolution of Italian GDP growth. In particular, we use a mixed-frequency dynamic factor model together with a sophisticated susceptible-infectious-recovered epidemic model featuring endogenous policy responses. First, we simulate different scenarios of economic growth depending on the course of the pandemic in Italy. Second, we evaluate the forecast performance of the model for the period August 2020-March 2021. Taking epidemiological indicators into consideration is important for obtaining reliable projections.

## A two-step short-term demand forecasting algorithm based on generalized additive models

Presenter: Zaniar Ahmadi
Co-authors: Zaniar Ahmadi;Armin Pourkhanali;Gordon Leslie

The pattern of electricity prices in Australia has been changed significantly from amongst the cheapest in the International Energy Agency (IEA) countries to amongst the most expensive, from 2004 to 2018. One of the important primary reasons for these changes has identified peak demand-driven network investment as the dominant driver of electricity price increases. Critical peak demand events occurring less than 0.5% of the year are the chief driver of network investment. Indeed, this rare event happens to demand significantly higher than its average level for a short time. Extreme weather events that drive very high aggregate space heating or cooling demand can lead to this matter. Consequently, electricity prices are among Australia's most critical policy issues and represent a crucial component of the energy and climate change policy discussions. In this paper, we introduce the two-step algorithm based on using the Generalised Additive Model (GAM) approach. The proposed algorithm has the potential to match or even outperform more complex non-linear forecasting models. Two essential keys to the success of this model are the effective use of weather station data (temperature and humidity) which has direct interaction between seasonal patterns and electricity consumption. Secondly, using zone-substation data as an independent variable for forecasting electricity demand leads to significant improvement. Using zone-substation data to improve forecasting performance is novel in the literature and an important innovation in electricity demand forecasting. The proposed method has been used to forecast the half-hourly electricity for power systems demand in one major region of the Australian National Electricity Market (NEM). The performance of the proposed model outperforms the benchmarks by about 25% in terms of the mean absolute percentage error (MAPE) criterion. The gains are almost exclusively driven by the methodology, confirming its potential to match or even outperform more complex nonlinear forecasting models. Despite considerable spatial heterogeneity in temperature, weather and load patterns, including these data provide little additional forecast improvement in forecasting aggregate load relative to inputs from only a single location.

## Incorporating downstream, task-specific information in forecasting models

Presenter: Hussain Kazmi
Co-authors: Hussain Kazmi;Maria Paskevich

Demand forecasting is a critical step in most Supply Chain management problems. Inaccurate demand forecasts can lead to large scale disruption in downstream operations. This has been evident over the past two years as changes in demand profiles, caused by pandemic-induced lockdowns, have caused shortages of certain objects and excess of others. The energy domain perhaps best exemplifies this, ranging from negative electricity prices across Europe in spring 2020 off the back of depressed demand and elevated solar generation to unprecedented high prices in winter 2021 due primarily to high demand and supply side constraints. However, the energy sector is by no means an outlier in its demonstration of the vagaries of demand forecasting. Forecasting of demand is thus an important part of decision-making pipelines in business contexts, but so are decisions that are made based on these forecasts. This decision-making part should therefore be included in the actual evaluation of the performance of forecasts. In this talk, we highlight the importance of using downstream value as an evaluation metric for forecasting algorithms.

Using cases from energy and Retail Demand Forecasting, we show how this enables us to combine the risk-aware character of interval forecasts with the simplicity of point forecasts. The proposed framework allows for the identification of the optimal forecaster as measured not just by forecast accuracy but also downstream value or utility. Together with computation costs, this allows practitioners to quantify the real-world value of input data in a more relevant business metric (e.g. in dollar-terms) as opposed to traditional MSE or MAPE. Consequently, it helps reduce both the direct and indirect costs of training forecast models. More concretely, our results show that the relationship between forecast accuracy and downstream performance is not monotonic. We demonstrate this using both simulated and actual forecasts, and showcase the considerable improvement value-oriented forecasting can bring in practical settings. Unlike interval forecasts, such value-oriented forecasts do not require an adaptation of the downstream decision-making pipeline. This paper is based on work carried out for the IIF Annual Award 2020 in the business applications category.

## Assessing the Accuracy of Directional Forecasts

Presenter: Constantin Burgi
Co-authors: Constantin Burgi;

This paper introduces a new and unified framework for testing the accuracy of directional forecasts that allow for importance weightings. The flexibility of the new framework allows testing for directional accuracy for a large variety of benchmarks and loss functions including serial dependence and the profitability of trading strategies. The framework is based the weighted chi-squared tests of independence for contingency tables and has a similar power to existing specialized tests based on simulations. The framework is applied to the categorical EUR/USD forecasts in the ifo-Institute's World Economic Survey and to the point forecasts for quarterly GDP in the Philadelphia Fed's Survey of Professional Forecasters. We are able to pinpoint that the predictive accuracy identified for some exchange rate forecasts is likely due to serial dependence and that the GDP forecasts have predictive value up to two quarters ahead.

## "Three is not a crowd": A new test for Forecast Evaluation

Presenter: Nicolas Hardy
Co-authors: Nicolas Hardy;Pablo Pincheira

In this paper, we propose a new asymptotically normal test for the out-of-sample evaluation of two competing forecasts. The main idea of our strategy is to evaluate the relative performance of both forecasts through a third common benchmark. Our test builds on the following intuition: When comparing two competing forecasts, it is sometimes the case that one forecast overwhelmingly outperforms the benchmark and that the other does not. By transitivity, we might expect that a direct comparison between the two competing forecasts would yield a result favorable to the forecast outperforming the benchmark. However, traditional comparisons of MSPE (e.g., Diebold and Mariano (1995) and West (1996)) may fail to detect this superiority. Through a simple decomposition, we notice that the MSFs may play an important role. Our test tackles this issue by allowing a desirable notion of "transitivity." Additionally, building on West (1996) asymptotic theory, we show that our test can easily accommodate parameter uncertainty. Our Monte Carlo simulations suggest that our test displays adequate size, even in small samples. Finally, we illustrate the benefits of our test with an empirical illustration in the context of commodity currencies.

## Evaluating Incident Forecasts in the Presence of Sparsity

Presenter: Red Davies
Co-authors: Red Davies;

This paper examines the problem of evaluating "incident" forecasts for a large online retailer (Wayfair). An "incident" is defined as any order that does not go as planned (e.g., the product arrived damaged, the package is lost in transit, etc). The forecasting goal was to predict incident probabilities at the product

(SKU) level. There is a unique problem of sparsity, however, with incident forecasts. Not only do many products have only a few orders each month, but only a fraction of these orders actually have an incident. For example, if a product had five orders in a month and one of those orders had an incident then the observed incident rate would be 0.20. This value, however, might not be a very accurate predictor of how likely an incident is to occur for a future order.

To examine this problem, the paper conducted a simulation exercise. We simulated 10,000 products where each product had an unobserved ("latent") stochastic incident rate that varied between 0.09 and 0.13. Each product also had an associated Poisson process that was used to simulate orders for that product – with 2,000 of the products having a high-volume order process and the other 8,000 having a low-volume order process. The "observed" incident rates were defined as the product of the simulated number of orders and the latent incident rate rounded to the nearest integer. We simulated a 24 month training set and a one month evaluation set. Three benchmark forecasting models were used: an ETS model, an ARIMA model, and a simple historical average. The forecasts were evaluated using mean-squared error (MSE) and the mean absolute scaled error (MASE) metric introduced in Hyndman & Koehler (2006).

When compared against the observed incident rates, unsurprisingly, the forecasts for the high-volume products were more accurate. However, when compared against the unobserved latent rates, the forecasts of the low-volume products had MSE values much closer to those of the high-volume products and MASE values that were actually lower. It appears that our forecasts might be more accurate than the observed data would lead us to believe, and that the MASE error metric is taking into account some of the difficulty of forecasting in a situation of high sparsity.

## Forecasting tennis game outcomes: The predictive value of betting odds

Presenter: Steffen Q. Mueller
Co-authors: Steffen Q. Mueller;

While it is widely acknowledged that betting odds provide valuable information for forecasting game outcomes across various sports, the vast majority of studies analyzing the predictive value of betting odds use odds provided by only one bookmaker or average the odds of different bookmakers. However, not only have betting markets been shown to be biased in various ways, but there is evidence that bookmakers exploit bettors' biases when setting their odds to maximize profits. Moreover, different techniques exist to convert betting odds into winning probabilities, but previous sports forecasting research typically considers only one probability conversion method. In contrast, this study considers odds provided by various bookmakers and evaluates different probability conversion methods for forecasting individual game outcomes using men's professional tennis data from 2001 to 2017. Using different statistical and machine learning methods, my preliminary results indicate significant variation in forecasting accuracy with respect to both bookmakers and probability conversion methods; however, the most accurate forecast results from predictions incorporating all available odds and odds-based winning probabilities.

## Classifying Pantry Food Offerings by Client Preferences

Presenter: Benjamin Morrow
Co-authors: Benjamin Morrow;

Client choice food pantries offer an alternative approach to traditional pantries, featuring something comparable to the grocery store shopping experience where pantry guests are free to select from donations the foods they prefer. This research develops a systematic approach for classifying pantry food offerings according to client preferences, identifying key demographic predictors of those choices, and offering guidance on an optimal food stocking strategy.

## Predicting and Optimizing the Fair Allocation of Donations in Humanitarian Supply Chains

Presenter: Lauren Davis
Co-authors: Lauren Davis;

Non-profit hunger relief organizations (NPHROs) around the world depend mainly on the benevolence of donors to alleviate hunger. This implies that the quantity and frequency of donations they receive may vary over time. Moreover, their inability to determine additional food purchase requirements due to inaccurate donation estimation may pose a challenge as they strive to allocate food equitably throughout their service areas. This paper presents a simulation model to determine the quantity of food donations received per month in a multi-warehouse allocation network. The mean of the simulated donation estimates for each future month is used in an optimization model to estimate fair allocation of donations.

## Probabilistic reconciliation of count time series

Presenter: Giorgio Corani
Co-authors: Giorgio Corani;NIcolo Rubattu;Dario Azzimonti

We propose an algorithm for the probabilistic reconciliation of count time series. First, we form a joint distribution for the entire hierarchy considering only the base forecasts of the bottom time series; this is the probabilistic bottom-up reconciliation.

We then update such joint using the forecast for the upper time series (i.e., the non-bottom time series). We implement the updating via sampling in PyMC3, treating the base forecast of each upper time series as an uncertain observation of a sum of bottom time series. In the jargon of Bayesian networks, this is called updating with soft evidence. We perform multiple updates, one for each upper time series of the hierarchy. We eventually obtain a reconciled joint distribution for the entire hierarchy.

We test our method in the reconciliation of temporal hierarchies, comparing against probabilistic reconciliation based on the Gaussian and on the truncated Gaussian distribution. Our methods largely improves point forecast and prediction intervals compared to traditional methods on intermittent time series. Interestingly we obtain a reconciled distribution defined over counts whose prediction intervals are often asymmetric. On the other hand, when dealing with smooth time series, its reconciled forecasts are practically equivalent to those of minT.

Further developments could regard ad-hoc algorithms for faster sampling of the reconciled distribution.

## Improving hierarchical time series forecasting via resistant reconciliation

Presenter: Fernando Cyrino
Co-authors: Fernando Cyrino;Mauricio Lila;Erick Meira

Hierarchical Time Series (HTS) stand for a set of time series that can be aggregated at different levels, according to a well-defined hierarchical structure. Hierarchical forecasting methods take advantage of the hierarchical structure of the data through base forecast reconciliation, generating results that are usually unbiased and more accurate than those provided by benchmark methods. When combining base forecasts through reconciliation strategies, however, some forecasts may behave like outliers, causing distortions to the reconciliation process. This work introduces the concept of hierarchical forecast reconciliation using resistant regression techniques. By considering resistant statistics in the forecast reconciliation process, we address a significant gap in the modelling and forecasting of hierarchical time series: the substantial reduction of contamination in the involved series due to outlying forecasts. To demonstrate the potential and validity of the proposed strategy, we set forth an application using hierarchical data on monthly electric energy consumption in Brazil, generating forecasts up to four months ahead, a forecast horizon that is of paramount importance to several stakeholders in the energy supply industry. We compare the performance of our strategy with that from traditional benchmarks and with state-of-the-art methods in HTS. Overall,

resistant reconciliation show promising forecasting results under multiple settings and through the lens of different evaluation metrics. Consequently, the new approach is shown to be suitable to support decision making in the energy sector. Furthermore, the methodology developed is flexible, in the sense that it can be readily applied to other sets of hierarchical time series. Findings and policy implications are further discussed.

## Probabilistic Forecasting with modified N-BEATS networks

Presenter: Jente Van Belle
Co-authors: Jente Van Belle;Ruben Crevits;Wouter Verbeke

In this paper, we present a modification to the state-of-the-art N-BEATS deep learning architecture for the univariate time series point forecasting problem for generating probabilistic forecasts. Next, we propose an extension to this probabilistic N-BEATS architecture to allow optimizing probabilistic forecasts from both a traditional forecast accuracy perspective as well as a forecast stability perspective, where the latter is defined in terms of a change in the forecast distribution for a specific time period caused by updating the probabilistic forecast for this time period when new observations become available, i.e., as time passes. For the M4 monthly data set, we show that this leads to more stable forecast distributions without leading to a significant deterioration in their accuracy. Finally, we present a second extension to the probabilistic N-BEATS model for use in an inventory management context in that also probabilistic forecasts of temporal aggregates, i.e., cumulative demands, are produced by the network. Results are reported for the M4 monthly data set and indicate that large improvements in accuracy can be obtained over basic but well-established methods to produce probabilistic cumulative demand forecasts.

## Data-driven forecasting for a multi-level parcel distribution network

Presenter: Gabor Tamas
Co-authors: Gábor Tamás;Elisabeth Zehendner;Christian Paier;Emel Arikan;Gerald Reiner

Our research proposes an approach for data-driven parcel quantity forecasting in and between network nodes of a multi-level parcel distribution network. Actual parcel quantities for all nodes and edges depend on the total parcel quantity and the applied processing rules. Individual forecasts for each network node cannot capture these interdependencies nor known changes to the processing rules. To predict realistic parcel volumes at all levels, we use a three-stage approach combining forecasting and simulation. First, parcel quantities for origin-destination pairs are forecasted based on historical data. Several forecasting approaches are considered and compared quantitatively. Second, up-to-date routing rules are inferred from the system. These routing rules define the path a parcel follows through the network from its origin to its destination. Finally, parcel flows are simulated by combining forecasted parcel quantities, routing rules, and node capacities. The results are forecasts for all nodes and edges in the network and performance metrics for the entire system. The proposed three-stage approach helps to achieve better network performance. Routing rules can be changed to evaluate alternative network settings and managerial decisions. Parcel quantities and utilization rates for each node and overall network performance are obtained for each setup and can be compared. The feasibility and applicability of our approach are presented using real-world parcel sorting data from the Austrian Post.

## Information sharing via hierarchical modelling for collaboration in a retail Supply Chain

Presenter: Yves R. Sagaert
Co-authors: Yves R. Sagaert;Filotas Theodosiou

Collaboration in the Supply Chain between a retailer and a supplier has been extensively researched. Much of the relevant literature has focused on the theoretical impact of this collaboration, with several studies arguing for its potential benefits, ranging from counteracting the bullwhip effect to reducing the

likelihood of a stock-out. Despite this, most empirical work has focused on improving the accuracy of supplier-level forecasting. Using regression-based approaches, retailer sales information is included as exogenous variables to improve the supplier demand forecasts. However, this has limited benefits for the retailer. In this work, we aim to reduce total Supply Chain costs, both at the retailer and supplier level, by tracking decisions throughout the Supply Chain. These decisions are usually made on different frequencies. Our case study consists of a Supply Chain with three echelons: shop level, distribution centre (DC) and supplier level. The decisions about replenishing inventory are made at different times for each tier, causing the decision makers to model demand at different frequencies. Driven by the advances in the hierarchical forecasting literature, we propose ordering the demand time series into temporal and cross-sectional aggregations and structure them into a partial cross-temporal hierarchy. The decisions are not taken at the same moment across Supply Chain echelons but usually in a reactive manner that is prone to overstocking. The hierarchical approach allows to represent the propagation of decision-making, enhancing the potential for collaboration. This creates the potential for an AI-driven integrated Supply Chain. We examine how this hierarchical approach enables information sharing across decision levels and data frequencies and its impact on the total Supply Chain. We present the results based on data from a logistics network with 275 stores, 4 distribution centres and one supplier.

## Forecasting inflation with twitter

Presenter: Martin Llada
Co-authors: Martin Llada;Daniel Aromí

We use Twitter content to generate an indicator of attention allocated to inflation. The analysis corresponds to Argentina for the period 2012-2019. The attention index provides valuable information regarding future levels of inflation. A one standard deviation increment in the index is followed by an increment of approximately 0.4% in expected inflation in the consecutive month. Out-of-sample exercises confirm that social media content allows for gains in forecast accuracy. Beyond point forecasts, the index provides valuable information regarding inflation uncertainty. The proposed indicator compares favorably with other indicators such as media content, media tweets, google search intensity and consumer surveys.

## Political Polarization, Inflation's Perception and Inflation Expectations – Evidence form Poland

Presenter: Kamil Jonski
Co-authors: Kamil Jonski;

Since 2015 double electoral victory of right-wing Law&Justice (parliamentary and presidential) the country's political landscape is deeply polarized. That includes not only traditional conservative vs. progressives policy debate, but also unprecedented rule of law crisis and – crucially for this presentation – the core premises of the economic policy. As Law&Justice flagship policy, universal monetary child benefit had been introduced, their opponents – associated with free-market economists – rang alarm bells on budget balance and unfeasibility of the project. None of them happened, increasing Law&Justice credibility in the economic policy area. That perception seemed reinforced with massive economic stimulus during the covid pandemic – largely funded with tricks bordering with creative budgetary accounting. As post-pandemic inflation surge occurred, Law&Justice-nominated governor of de iure independent Polish Central Bank (former L&J politician and personal colleague of its almighty chairmen) first had been catch off guard (with asset purchases, 0 interest rates and forward guidance of no need for tightening). Second, with inflation approaching double digits, he (and the Law&Justice nominated Monetary Policy Council) embarked unexpected tightening to keep pace with other central banks of the region. Data on inflation perceptions (over last year), inflation expectations (over next year) collected by the Public Opinion Research Centre (CBOS) on the representative sample of adult poles (4th-14th Oct 2021 - N=1161, 3rd – 13th Jan 2022 – N= 1135, mixed CAPI, CATI, CAWI) had been examined to uncover whether political preferences (Law&Justice supporter, Law&Justice opponent, not engaged) and opinions on politically polarizing issues affects inflation expectations formation (controlling for demographic factors and economic situation of the

household). Moreover, using the Jan 2022 poll, respondents views on the mechanism behind inflation surge (domestic economic policies, external factors) as well as expectations on the effectiveness of government anti-inflation policies had been examined. Results will contribute to the understanding of the inflation expectations formation process in heavily politically polarized environments.

## A Multi-decomposed Wavelet Neural Network for Long Term Forecasting of CPI Inflation for BRIC Countries under economic and geo-political uncertainties

Presenter: Shovon Sengupta
Co-authors: Shovon Sengupta;Tanujit Chakraborty;Sunny Kumar Singh

Inflation is regarded as one of the key policy variables for the overall monetary policy of an economy. Thus, a good understanding of the factors that help forecast inflation accurately is fundamental to policymakers at the central banks. During the last few decades, emerging market economies like Brazil, Russia, India, and China (BRIC) have witnessed tremendous changes in their macroeconomic environment like changes in monetary policy (inflation targeting), globalization, and other factors like pandemic and financial crises. These changes pose a significant challenge for forecasters interested in predicting key macroeconomic variables like consumer price index (CPI) inflation. Given these dynamic changes in the macroeconomic environment, factors like economic policy uncertainty and geopolitical risk might help forecast inflation accurately.This study proposes a multi-decomposed wavelet-based autoregressive neural network (MD-WARNN) model that can produce reliable forecasts for CPI inflation prediction for BRIC countries. The proposal is flexible and can easily incorporate exogenous variables, namely economic policy uncertainty and geopolitical risk, in its scalable framework. The proposal uses 'ChristianoFitzgerald' approximation to the ideal band-pass filter and 'Hodrick-Prescott' filter to extract the cyclical and trend components and use them as explanatory variables during the data preprocessing stage. Further, wavelet decomposed time series data are transported into a neural network model to generate final inflation forecasts. The proposal's competitiveness compared to the state-of-the-art statistical, machine learning, and deep learning models (ARIMAX, SARIMAX,ARFIMAX, XGBoost, DeepAR, ARNN, TFT, N-BEATS etc.) is shown using computation experiments based on standard metrics. Finally, the theoretical robustness of the proposed model is studied by showing the asymptotic properties of the associated Markov chain.Overall, the study reveals that economic policy uncertainty and geopolitical risk indices, to some significant extent, capture the hidden uncertainty about inflation movement and turn out to be critical factors for generating more accurate forecast numbers for inflation for the BRIC countries. This approach can potentially serve as an alternative forecasting strategy for the central banks.

## Making Energy Forecasting Resilient to Missing Features: a Robust Optimization Approach

Presenter: Akylas Stratigakos
Co-authors: Akylas Stratigakos;Andrea Michiorri;Georges Kariniotakis

Short-term forecasting is key to the safe, reliable, and economic operation of modern power systems. As the majority of modern forecasting tools are purely data-driven, their performance relies heavily on the quality and availability of data. In this work, we examine forecasting when a subset of features used during model training becomes unavailable (deleted or missing) in an operational setting, a subject that has been largely overlooked by previous works. Several reasons could lead to feature deletion, including malicious data-integrity attacks, network latency, and equipment malfunctions, among others. We leverage tools from robust optimization and machine learning and formulate a linear regression model that is optimally resilient to the deletion of features at test time. The robust counterpart of the proposed model is a linear program whose size grows polynomially with the number of training observations and the number of features; we further provide a decomposition algorithm based on the alternating direction method of multipliers to deal with large problem instances that are typically found in energy forecasting applications. We further extend to the case of Probabilistic Forecasting by robustifying the standard linear quantile regression model. To validate empirically the proposed approach, we examine several prevalent forecasting practices in power

systems, namely electricity prices, load, wind production, and solar production forecasting. We compare against regularized and randomization-based models and benchmark their performance for the case of feature deletion. The results show that the proposed solution successfully mitigates the adverse effects of missing features, leading to the lowest overall performance degradation. Further, it successfully hedges against the most adverse scenario of deleting the most important feature from the test set. The results persist both for point and probabilistic forecasts and across the different series. Overall, this work highlights the benefits of leveraging robust optimization and provides a new perspective on how to deal with feature uncertainty in energy forecasting applications.

## Privacy-preserving solar forecasting using federated learning

Presenter: Attila Balint
Co-authors: Attila Balint;Hussain Kazmi;Johan Driesen

In light of climate change concerns, renewable energy sources, such as wind and solar, have rapidly proliferated over the past decades, and today form a significant part of the electricity mix in several countries. These are inherently variable and intermittent in nature, and introduce an element of stochasticity in grid operation. Forecasting them is therefore of vital importance, not just for grid operators but society as a whole. Several algorithms, ranging from classical techniques such as exponential smoothing to machine learning ones such as boosted trees and Neural Networks, have been proposed to predict electricity generation using solar panels. With the continuous increase in the number of sites where solar generation needs to be forecast, global forecasting methods, which enable cross-learning from multiple data sources, have gained in popularity. These algorithms can address the recurring cold start problem while also outperforming purely physics-based models. However, a common assumption in such global forecasting models is the availability of data from all sites at a central server. Especially when different solar systems are spread spatially and located in different sites, this assumption does not necessarily hold and aggregating the data can lead to undesirable privacy leakage. In this talk, using data from 50 different solar sites spread across the Netherlands, we show how federated learning can be used to tackle this issue without significantly degrading forecast accuracy. Formulating the forecasting problem in this way allows us to combine the best of both worlds: achieve the performance gains of global models due to cross-learning, while retaining the data privacy associated with local models. We benchmark our results against simpler algorithms as well as forecasters with access to all data, and present an overview of the key benefits and limitations of such algorithms.

## An Exploratory Study of Solar Irradiance Reanalysis data available in Brazil for Scenarios Simulation

Presenter: Margarete Afonso de Sousa
Co-authors: Margarete Afonso de Sousa;Fernando Cyrino Oliveira;Soraida Aguilar;Reinaldo Souza

Renewable energy sources are growing worldwide mainly due to efforts to obtain net zero. In 2020, the global renewable power capacity rose 7%, and Solar PV led this growth. In Brazil, the Installed Capacity of Solar Photovoltaic Electricity Generation quintupled in the last years, going from 935 MW in 2017 to 4,630 MW in 2021. The intermittence of solar energy increases challenges in forecasting Solar PV generation for decision-makers to plan flexible and adequate distribution systems. Brazil is a country that has a substantial potential for Solar PV, which makes it essential to obtain solar radiation data to simulate output generation scenarios and subsidize grid expansion. Despite the rise in investment in this energy source, the availability of measured public data is limited to some weather stations that do not cover all the Brazilian territory. Nowadays, reanalysis datasets to obtain climate data has become popular, and its use could minimize this lack of information. Nevertheless, how distant is the reanalysis data from the measured ones? This article aims to make an exploratory study of different reanalysis databases available for the Horizontal Global Irradiance variable, on an hourly basis, focused in Brazil and compared with public ground measured data. As a secondary objective, this study analyzes each reanalysis base and its characteristics by comparing them, creating a reference on the subject for the area of solar energy in Brazil.

## The M6 competition in progress

Presenter: Spyros Makridakis; Evangelos Spiliotis
Co-authors: Spyros Makridakis;Evangelos Spiliotis;Fotios Petropoulos;Norman Swanson;Ross Hollyman;Maria Michailidis

The M6 competition (https://mofc.unic.ac.cy/the-m6-competition/), the sixth installment of the M competition series organized by Spyros Makridakis, aims to empirically identify the most appropriate ways of forecasting financial (stock and ETF) prices and utilizing such forecasts to make profitable investments. The findings of the competition will shed new light on the efficient market hypothesis and investigate the connection between the accuracy of forecasts and estimation of risk with the associated returns on investment. The M6 "duathlon" competition was designed to attract participation from financial experts, data scientists, and economists, among other interested parties, providing significant monetary prizes for the best forecasting performance and for the highest, risk-adjusted investment returns, while also being live and consisting of 12 monthly submission points to allow for objective results. In this session, Spyros Makridakis and Evangelos Spiliotis will describe the purpose, objectives, and research hypotheses of the M6 financial competition, present the schedule, duration, input data, submission format, and evaluation measures of the competition, and provide some initial findings based on the results available till that point. Normal Swanson will then moderate a discussion in order for the attendees to express their views, exchange constructive ideas, and identify critical points of analysis.

## Predicting the M6 competiton with technical indicator Neural Network Ensembles

Presenter: Kamran Rismanchi
Co-authors: Kamran Rismanchi;Sven F. Crone

Forecasting financial markets has captured the imagination of researchers and practitioners alike, given the monetary reward despite evidence of the efficient market hypothesis. As a consequence, many techniques have been derived to predict stock market prices from technical and fundamental analysis, traditional time series approaches to artificial intelligence and machine learning techniques, daring back to the first papers on IBM stock prices by White et al (1988). The ongoing M6 forecasting competition aims at systematically evaluating these approaches. As stock market returns are known to be often noisy and volatile, sophisticated algorithms are employed to separate the signal from noise. Our approach attempts to do the opposite, by predicting stock prices by applying simple yet robust to train, shallow Multilayer Perceptron- Ensembles with selected external technical indicators as features to exploits local time trend and mean reversion, most notably Bollinger Band indicators.

## Bridging the gap between forecast and business value

Presenter: Johann Robette
Co-authors: Johann Robette;

Like any other business activity, forecasting must monitor, demonstrate and defend its performance. What is its actual added value to the business?

The norm in forecasting is to evaluate the performance of a forecast through an accuracy metric. Therefore, all attempts to measure business added value are based on some form of accuracy metric. In fact, our entire practice is based on the assumption that "the greater the accuracy, the better the decision".

What if this widely accepted assumption is wrong? What if optimizing our forecasts against an accuracy metric is counterproductive? In this thought-provoking presentation, we'll discuss the surprising relationship between forecast accuracy and value!

Based on the findings of empirical research using the M5 competition data set, we will answer the following questions: - Why is assessing the real added value of a forecast so important? - Does improving the accuracy

of a forecast necessarily lead to better decisions? - How can the quality of a forecast be assessed from a business perspective? - Where is it worthwhile to improve a forecast and where is it a waste of time and energy? - What exciting new perspectives do cost-oriented metrics and digital twins open up?

## Trends in the Field of Forecasting (1982-2022)

Presenter: Claudio Antonini
Co-authors: Claudio Antonini;

An analysis of the n-grams used in the abstracts and articles of the International Journal of Forecasting (IJF) and the abstracts presented at the International Symposium on Forecasting (ISF) indicate definitive trends, and possibilities of research and improvement in the field of forecasting. At a more general level, and invoking some of the assumptions used by Whorf and Wittgenstein in linguistics, both the IJF and ISF are simultaneously reservoirs and examples of what forecasting represents, providing a 'world view' of forecasting. In the same way that citation analysis is used to reveal the impact that an article had over time, dynamic word (or n-gram) analysis can also indicate what sets of words represent what is considered 'forecasting,' in this case since the 1980s. However, the n-gram analysis promises more depth than merely counting citations, as the individual words and the n-grams have inherent meanings that may change with time, representing new facets in the field. At the same time, various editions of Box and Jenkins' traditional book, 'Time Series Analysis: Forecasting and Control,' were analyzed. It is evident that Box and Jenkins use forecasting as a tool, and if they look at accuracy metrics is to design control laws to correct "potential deviations … from a desired target." Thus, their objective is not to use forecasting blindly—in a narrow sense—as a test bench to make techniques compete in terms of their accuracies on known datasets, but to create a framework (control law) that can self-correct in an unknown future. This approach should guide policies to be implemented by decision-makers, as some of the IJF early articles and ISF abstracts also demonstrate. The analysis carried out so far indicates the n-grams that, since the 1980s, have increased, decreased, or being used in a definite period. They point to increased usage of progressively larger scale numerical techniques, but to a decrease of theory, research, and guidance on firm's practices (general decision-making, or how to implement courses of action as a result of a forecast). A tool in R/Shiny was designed to demonstrate these trends and allow discovering more patterns.

## BYOL (Bring Your Own Language): The Case for the Tower of Babel

Presenter: Michele Trovero
Co-authors: Michele Trovero;Spiros Potamitis;Joe Katz;Sahbi Chaieb

The current volatile economic conditions have forced many organizations worldwide to operate on razor-thin margins. As a result, they are focused on developing forecasting processes that provide the highest level of accuracy to optimize their business processes. Modern forecasting ecosystems should be able to provide a framework to incorporate and scale the most advanced forecasting algorithms in a governed and consistent way. At the same time, data scientists need to have the option of using algorithms, from both the proprietary and open-source worlds, or to be able to program and quickly productionalize and scale new algorithms in the language of their choice. This has been a constant challenge because businesses have had a difficult time identifying an easy method of seamless integration between proprietary systems, which offer a high level of consistency and governance, and the open-source world, which continuously evolves and includes a wide variety of advanced algorithms that data scientists want to use. We illustrate the main points with practical examples of scaling of popular open-source algorithms in R and Python using SAS® Visual Forecasting software, an open and extensible framework that integrates and parallelizes algorithms, making them shareable and reusable.

## Bias in the reporting of probabilistic expectations: evidence and implications

Presenter: David Comerford
Co-authors: David Comerford;

Survey measures of likelihood beliefs are widely used. For instance, inflation beliefs elicited as probabilities are published by the New York Fed. Also, several recent influential papers have implicated survival beliefs elicited as probabilities as explaining undersaving for retirement and a bias against annuitization. Given their influence, it is important to assess whether a face value interpretation of these reported probabilities is valid. Leveraging features of the SHARE survey of 14 European countries and Israel, I demonstrate national differences in the belief implied by a response of "100%". This result implies that a change in reported beliefs of 5 percentage points conveys a different magnitude of change in beliefs at the top of the distribution than elsewhere in the distribution i.e. there is curvature in the function that translates respondents' beliefs to reported probabilities. For instance, the expectations reported in the Health and Retirement Study are biased towards 50% relative to respondents' beliefs. An implication of this reporting bias is specious pessimism: people understate their probabilistic beliefs to a greater degree when they happen to be asked about higher probability events than when they happen to be asked about lower probability events. This reporting bias fully explains why older respondents appear more optimistic regarding their survival than younger respondents and why men appear more optimistic than women. I show that age and gender differences in survival optimism are illusory in US, English and Irish data. A further insight concerns predicting behaviours and forecasting outcomes. Modelling reporting bias allows us better identify beliefs. Accounting for reporting bias in reported survival beliefs should deliver more accurate forecasts of mortality and should better predict a survival belief-contingent behaviour i.e. buying an annuity. I specify a simple model that includes a quadratic term to control for curvature in the reporting function. I apply it to survival probabilities reported in Wave 10 of the HRS. The quadratic term is a significant predictor of respondent's survival and of their claiming an annuity 6 years later, in HRS Wave 13. I conclude that knowledge of reporting bias is valuable for interpreting survey expectations elicited as probabilities.

## Extremal dependence modelling of global horizontal irradiance with temperature and humidity: An application using South African data

Presenter: Thakhani Ravele
Co-authors: Thakhani Ravele;Caston Sigauke;Lordwell Jhamba

This paper discusses extremal dependence modelling of global horizontal irradiance (GHI) with temperature and relative humidity (RH) at one radiometric station using South African data from 16 November 2015 to 16 November 2021. The station of interest is the University of Venda radiometric station. Multivariate adaptive regression spline (MARS) models are used to model the impact of weather variables such as temperature and RH on the GHI power production. The results from the MARS models revealed that the relationship between GHI and temperature is directly proportional, whereas the relationship between GHI and RH is inversely proportional. Empirical results showed that the marginal increases of GHI converge to 0.12W/m2 when temperature converges to 44.26°C and the marginal increases of GHI converge to 0.1W/m2 when temperature converges to 103.26%. Conditioning on GHI, the study found that temperature and RH variables have a negative extremal dependence on large values of GHI. The modelling approach from this study may help the system operators reveal information for operating the power plant during the heatwave period.

## The coverage probability of forecast intervals in the presence of unpredictable and predictable spikes

Presenter: Samaneh Sheybanivaziri
Co-authors: Samaneh Sheybanivaziri;Jonas Andersson

Probabilistic Forecasting for electricity prices has become more critical with the increased power production from wind, solar, and, other renewable resources. The capacity from those on a future time point is often quite uncertain. As an example, wind turbines cannot produce very low or very high wind speeds. This is one of the reasons, mentioned by Nowotarski and Weron (2018), that uncertainty in supply, demand,

and prices increases. In this paper, we compare methods to compute forecast intervals for electricity price forecasts. In our experiment, the electricity prices are assumed to be generated by a mixture model with two components, one for regular prices and, one for spikes; a specification that we argue captures the most essential features of distributional and, temporal properties of electricity prices. Employing a Monte Carlo simulation study we, systematically, compare different methods. Knowing the underlying stochastic process of the data makes this a complement to studies on observed electricity price data, such as the empirical study in the review article by Nowotarski and Weron (2018). We also argue for the realism of our assumed data generating process by studying a day-ahead electricity price series.

## Angular Combining of Forecasts of Probability Distributions

Presenter: James Taylor
Co-authors: James Taylor;

When multiple forecasts are available for a probability distribution, forecast combining enables a pragmatic synthesis of the available information to extract the wisdom of the crowd. A linear opinion pool has been widely used, whereby the combining is applied to the probability predictions of the distributional forecasts. However, it has been argued that this will tend to deliver overdispersed distributional forecasts, prompting the combination to be applied, instead, to the quantile predictions of the distributional forecasts. Results from different applications are mixed, leaving it as an empirical question whether to combine probabilities or quantiles. In this paper, we present an alternative approach. Looking at the distributional forecasts, combining the probability forecasts can be viewed as vertical combining, with quantile forecast combining seen as horizontal combining. Our alternative approach is to allow combining to take place on an angle between the extreme cases of vertical and horizontal combining. We term this angular combining. The angle is a parameter that can be optimized using a proper scoring rule. For angular averaging, we show how varying the angle affects the density and distribution functions. As with vertical and horizontal averaging, angular averaging results in a distribution with mean equal to the average of the means of the distributions that are being combined. We provide empirical support for this new form of combining using weekly distributional forecasts of COVID-19 mortality at the national and state level in the U.S.

## Detecting Crises, Jumps, and Changes in Regime with Saturation Techniques

Presenter: Neil Ericsson
Co-authors: Neil Ericsson;

Crises, jumps, and changes in regime can be persistent, time-dependent, and difficult to detect, yet have substantive implications for policy analysis. Saturation techniques provide a coherent framework for detecting and quantifying such structural breaks. Saturation techniques are inherently high-dimensional, requiring automated model selection with non-standard inference and utilizing machine learning algorithms to handle more potential regressors than observations—a characteristic common to big data. This paper characterizes several roles for saturation techniques in empirical economic analysis. They can demonstrate the robustness of a model to a wide range of feasible alternatives. They can also yield statistical and economic improvements to a model when structural breaks are present and thereby offer insights into the practical justification of empirical evidence. Additionally, saturation techniques provide a framework for creating near-realtime early-warning and rapid-detection devices, such as of financial market anomalies. Empirical applications include U.S. equity prices, U.K. exchange rates, government forecasts, climate change, and economic consequences of hurricanes.

## What Does it Take to Control Global Temperatures? Prospective and Counterfactual Carbon Abatement Policies in a Cointegrated Vector Autoregressive Model

Presenter: Guillaume Chevillon
Co-authors: Guillaume Chevillon;Takamitsu Kurita

This paper performs an empirical historical analysis of a Stochastic Integrated Model of Climate and the Economy cast into an integrated-cointegrated Vector Autoregressive Model and estimated over years A.D. 1004-2001. Extending previous results on Control Theory in nonstationary systems, we ask whether and quantify to what extent carbon abatement policies can effectively reduce global temperatures, using atmospheric carbon concentrations as an intermediate policy target. Using an extended dataset predating the industrial revolution, we also show how our econometric model for policy analysis can be used for counterfactual analyses. Our policy feasibility test shows that carbon abatement can have a significant long run impact and policies can render temperatures stationary around a chosen long run mean. Provisional results show that the counterfactual cost of carbon abatement for a retrospective policy aiming to keep global temperatures close to their 1900 historical level is about 40% of the observed 2003 level of world GDP, a cost equivalent to reverting to levels of output historically observed about 30 years ago.

## Discriminating direct from induced equilibrium mean shifts

Presenter: Jennifer Castle
Co-authors: Jennifer Castle;Jurgen Doornik;David Hendry

Equilibrium mean shifts can result directly from changes in intercepts with constant dynamics, or be induced by shifts in dynamics (or other parameters). The impacts of in-sample induced shifts modify previous taxonomies of forecast errors. Step-indicator saturation (SIS) helps detect any resulting location shifts, but even when all variables in the data generation process (DGP) and all indicators matching DGP shifts are selected, mis-forecasting can occur. To discriminate direct from induced shifts, we add to the model multiplicative indicators formed by interacting all selected step indicators with the lagged regressand. When equilibrium mean shifts are induced by changes in dynamics, forecasts can be markedly improved.

## The Effects of Privacy Protection on Forecast Accuracy

Presenter: Cameron Bale; Matthew Schneider
Co-authors: Cameron Bale;Matthew Schneider;Jinwook Lee

Forecasts generated using protected time series change significantly from those using the original time series. While prior experiments have demonstrated severe degradations in forecast accuracy from a VAR model applied to differentially private time series, little is known about how privacy protection affects other forecasting models. We measure the effects of several data protection methods (top and bottom coding, additive noise, differential privacy, and cluster-based swapping) on both simple and complex forecasting models. We find that data protection degrades forecast accuracy the majority of the time regardless of forecast horizon. Surprisingly, when the time series are protected with differential privacy or additive noise, we find that exponential smoothing models have better accuracy than LGBM models for all forecast horizons. We investigate the reasons behind these results and offer guidance for practitioners in selecting a forecast model for privacy protected time series data.

## Choosing Aggregation Levels for Forecasting and Fairness

Presenter: Jade Zhang
Co-authors: Jade Zhang;Lanqing Du;Jinwook Lee;Matthew Schneider

Fairness metrics capture the behavior of subgroups across multiple time periods and are widely used by organizations today. One issue in calculating fairness metrics is that decision makers do not analyze every possible subgroup and often pick the most favorable aggregation level. To explore this issue, we use University of California Davis enrollment data with race, gender, and residency subgroups over several years. We find that fairness metrics and their forecasts change significantly depending on how the time series data are aggregated. We propose a new sequential aggregation method based on pseudo-Boolean functions that minimizes the changes in fairness metrics. We forecast fairness metrics for subgroups so that

decision makers can react to potentially unfavorable fairness levels in the future. Our methodology is also generalizable to forecasters interested in choosing aggregation levels to improve hierarchical forecasts.

## Forecasting Criminal Justice Outcomes While Reducing Negative Consequences

Presenter: Nick Powell
Co-authors: Nick Powell;

Tasked with blending rehabilitative and accountability functions to improve public safety, probation and parole agencies must rely heavily on evidence-based practices. Through relationships with the academic community, the Georgia Department of Community Supervision (DCS) strives to bridge the gap between theoretical knowledge and real-world application. To this end, using statistical analyses to predict future events is a valuable means of making informed decisions. This workshop discusses some challenges (i.e., privacy concerns, data limitations, and explanations to the public) and implications (i.e., risk assessment, intervention targets, and officer discretion) when building research partnerships and translating research findings into practice.

## On the inventory performance of hierarchical forecasting approaches

Presenter: M. Zied Babai
Co-authors: M. Zied Babai;Bahman Rostami-Tabar

Demand forecasts are the basis of most decisions in Supply Chain management. The granularity of these decisions, either at the time level or the product level, lead to different forecast requirements. The most accurate forecasts are not always obtained from data at the 'natural' level of aggregation. In some cases, forecast accuracy may be improved by aggregating data or forecasts at lower levels, or disaggregating data or forecasts at higher levels, or by combining forecasts at multiple levels of aggregation. To do so, methods dealing with forecasting hierarchies (over time and other dimensions), and combinations of forecasts across hierarchical levels, have been recommended. Several hierarchical forecasting approaches have been developed in the literature over the last decade. However, despite the recent developments, most of the research in literature has looked at the forecast accuracy of the proposed approaches without any interest at their utility, and in the particular context of Supply Chains, their inventory performance has been ill-researched. This paper empirically analyses the inventory performance of existing hierarchical forecasting approaches. A reorder point and order-up-to-level policy is considered for the inventory control at the different levels of the hierarchy and the inventory performance is measured through the cost and service efficiency. The empirical investigation is based on data related to the Supply Chain of vaccines distributed in Kenya with 12 different vaccines and four levels of hierarchies. The results show that a forecasting approach with a high accuracy does not necessarily lead to a high inventory performance.

## Forecast of New Products in a Portuguese Brewery

Presenter: Ricardo Galante
Co-authors: Ricardo Galante;Teresa Alpuim

To be relevant to competitors, organizations from different industries regularly produce new products. And this is more present when the company comes from the CPG (Consumer Packaged Goods) sector, as a brewery that produces products to be consumed daily by consumers. For a brewery to launch a new brand or a new sub-brand on the market it is first necessary to forecast the expected demand, so that the raw material can be purchased in advance and the labor needed to develop these products can be hired. The big challenge lies in the fact that the product forecast has no history. In this context, what is done is to use data mining to identify products that are "similar" to that new product and to use the information from these similar products as a "starting point" for the new product. That is, the forecast is made for similar products and the values of this forecast are used as a "good guess" for the forecast of the new product. So, we have the whole process divided into two steps. The first one is related to the data mining process and

the second one is related to the forecasting process. Our proposal in this work was to compare methods both data mining and forecasting. For the data mining process, we made the segments using k-prototype and deep cluster (deep learning). For the forecasting process we used more traditional models, such as Time Series Regression models, and models based on Neural Networks. Then we made a comparison between the approaches used, i.e.: data mining using k-prototype and forecasting using conventional methods (Time Series Regression) versus data mining using deep cluster and forecasting using Neural Networks. As a result, we produced a benchmark between more traditional methods and methods using deep learning. We select a sub-brand of the brewery we are studying, ignore its history and "treat" it as a "new" sub-brand. Thus, this sub-brand is used as a reference for all this work.

## Hierarchical forecasting for inventory planning

Presenter: Oliver Schaer
Co-authors: Oliver Schaer;Nikolaos Kourentzes;Doug Thomas;Vidya Mani

Hierarchical forecasting has been proposed to achieve coherent forecasts across different planning levels of organizations, for example, at SKU/store level and at SKU/distribution center, where coherency suggests that the more disaggregate forecasts will always sum up to the aggregate ones. Similarly, temporal hierarchies have been proposed to connect different planning horizons. The literature suggests that hierarchical forecasting have the beneficial side effect of improved forecast accuracy, with temporal hierarchies exhibiting higher gains. Arguably, coherent forecasts lead to aligned decisions that should result in better outcomes, but little work evaluates the benefits of hierarchical forecasting in decisions. We focus on inventory management and investigate how the use of hierarchies benefits inventory performance. On a dataset from a large retailer, we empirically investigate this connection between inventory and the use of cross-sectional and temporal hierarchies. We draw conclusions on the connection between accuracy and decision benefits from hierarchical forecasting, as well as investigate any benefits stemming from the qualitative benefits of hierarchical forecasting, such as enforcing coherency between forecasts.

## Inventory Control for Periodic Intermittent Demand

Presenter: Sarah Van der Auweraer
Co-authors: Sarah Van der Auweraer;Joachim Arts;Thomas van Pelt

Intermittent demand is difficult to forecast, as many periods have no demand at all. Forecasting methods for such demand usually create separate estimates for the time between demand occurrences and the size of a demand occurrence. These methods implicitly assume that the time between demand occurrences is memoryless. Data from practice, however, indicates that the times between demand events is often not memoryless but –contrary to implicit model assumptions—displays periodicity. Consequently, the time since the last demand is an important predictor for future demand. We propose a demand model that accommodates such periodicity and show that the optimal inventory policy is a state-dependent base-stock policy, where the order-up-to-levels depend on the time since the last demand. We benchmark the performance of our approach against heuristic policies both in a numerical experiment and on five real data sets in terms of average inventory costs.

## Scores for Multivariate Distributions and Level Sets

Presenter: Xiaochun Meng
Co-authors: Xiaochun Meng;James Taylor;Souhaib Ben Taib;Siran Li

Forecasts of multivariate probability distributions are required for a variety of applications. Scoring rules enable the evaluation of forecast accuracy, and comparison between forecasting methods. We propose a theoretical framework for scoring rules for multivariate distributions, which encompasses the existing quadratic score and multivariate continuous ranked probability score. We demonstrate how this framework can be used to generate new scoring rules. In some multivariate contexts, it is a forecast of a level set that

is needed, such as a density level set for anomaly detection or the level set of the cumulative distribution as a measure of risk. This motivates consideration of scoring functions for such level sets. For univariate distributions, it is well-established that the continuous ranked probability score can be expressed as the integral over a quantile score. We show that, in a similar way, scoring rules for multivariate distributions can be decomposed to obtain scoring functions for level sets. Using this, we present scoring functions for different types of level set, including density level sets and level sets for cumulative distributions. To compute the scores, we propose a simple numerical algorithm. We illustrate our proposals in a simulation study and a financial application to portfolio construction.

## Bagged Value-at-Risk Forecast Combination

Presenter: Ekaterina Kazak
Co-authors: Ekaterina Kazak;Roxana Halbleib;Winfried Pohlmeier

Recent developments in financial econometrics literature on joint scoring functions for Value-at-Risk and Expected Shortfall allowed for consistent implementation of statistical tests based on the Model Confidence Set (MCS). MCS is shown to be a great tool for model comparison, both in-sample and out-of-sample. Another branch of literature focused on the superior performance of convex forecast combinations, which often outperform stand-alone forecasting models. This paper combines both results and proposes a novel approach to a forecast combination of Value-at-Risk and Expected Shortfall based on the MCS. We exploit the statistical properties of bootstrap aggregation (bagging) and combine competing models based on the bootstrapped probability of the model being in the Confidence Set. The resulting forecast combination allows for a flexible and smooth switch between the underlying models and outperforms the corresponding stand-alone forecasts.

## Penalized quantile regression for time series data with heavy tails

Presenter: Sander Barendse
Co-authors: Sander Barendse;

We study penalized quantile regression in high-dimensional sparse models for data with time series structure and heavy tails. This setting is important in the study of systemic risk, which is commonly quantified in terms of the (conditional) quantiles of the P&L of an individual financial firm and the financial sector as a whole. Such data is characterized by dependence across time and heavy tails. We obtain nonasymptotic estimation and prediction bounds by leveraging a Fuk-Nagaev inequality for a class of -mixing processes. We use our estimator to estimate the systemic risk measure $\Delta$CoVaR, and condition on the FREDMD factors and transformations. We show that good performance of the estimator requires a rich set of transformations of the FREDMD factors.

## Dynamic Co-Quantile Regression

Presenter: Timo Dimitriadis
Co-authors: Timo Dimitriadis;Yannick Hoga

The popular systemic risk measure CoVaR (Co-Value-at-Risk) is widely used in financial risk management, banking and insurance regulation. Formally, it is defined as an (extreme) conditional quantile of the asset return $Y_t$, given that the return of another asset $X_t$ is beyond its (extreme) conditional quantile. It measures how extreme observations of $X_t$ influence extreme observations of $Y_t$ and hence, it measures the spillover risk from one asset to another. In this article, we propose a dynamic ''Co-Quantile Regression" which jointly models the VaR and CoVaR semiparametrically and propose a two-step M-estimator drawing on recently proposed lexicographical loss functions for the pair VaR and CoVaR. Among others, this allows for the estimation of joint dynamic forecasting models for the CoVaR. We establish the asymptotic properties of the estimator suggest asymptotically valid inference. Our method can further by extended to the Marginal Expected Shortfall (ES) and the Co-Expected Shortfall (CoES) in a straight-forward fashion.

We apply our co-quantile regression to correct the statistical inference in the existing literature on CoVaR, and to generate CoVaR forecasts for real financial data, which show to be superior to existing methods.

## Confidence bands for the PIT histogram

Presenter: Felix Kiessner
Co-authors: Felix Kiessner;Malte Knüppel;Matei Demetrescu

The histogram of probability integral transforms (PIT) is a diagnostic tool to assess the consistency between forecasts of the probability distribution for a target variable and the corresponding realizations of that variable. Under consistency, we can expect the same number of observations in each bin of the PIT histogram. If, for instance, the probability distribution is too dispersed, we can expect more observations in the inner bins than in the outer bins. However, these histograms are commonly presented without any indication of the uncertainty surrounding the height of the histogram bars, i.e. without confidence bands. Therefore, in practice, it is often difficult to judge whether deviations from equal height are due to misspecification of the probability distribution forecast or due to sampling uncertainty. Sometimes, sampling uncertainty is presented in the form of individual confidence intervals for each bar. Yet, this approach does not take the multiple testing problem into account, which results from the simultaneous consideration of many bars. We consider various methods to construct simultaneous confidence bands for the PIT histogram, allowing for serial correlation of the PITs. These methods include Bonferroni and sup-t bands based either on long-run covariance estimation or on bootstrapping. We evaluate the performance of these methods in Monte Carlo simulations. We also compare their size and power properties to existing methods. An empirical application shows the usefulness of the new methods proposed.

## Reordering variables in VARs with stochastic volatility: implications for forecasting and structural analysis

Presenter: Florens Odendahl
Co-authors: Florens Odendahl;Gergely Ganics

Bayesian Vector Autoregressions (BVARs) with stochastic volatility are almost exclusively estimated using a lower triangular decomposition of the covariance matrix of the reduced-form innovations. It is well-known that in this case the ordering of the variables matters for the posterior distribution of the parameters. To circumvent this problem we propose to use an ordering-invariant autoregressive inverse Whishart (ARIW) process to model stochastic volatility. Using U.S. data in standard BVAR specifications, we document the relevance of the ordering problem for point and density forecasts, and show that the ARIW specification provides a competitive alternative. Additionally, the impulse response functions after a monetary policy shock also tend to differ considerably depending on the order of the variables when using the standard model specification, while this is not the case when using the ARIW specification.

## Robust Inference in Structural VAR Models Identified by Non-Gaussianity

Presenter: Lukas Hoesch
Co-authors: Lukas Hoesch;Geert Mesters;Adam Lee

Existing methods that exploit non-Gaussian distributions to identify structural impulse responses and conduct inference in SVAR models are not robust to small deviations from Gaussianity. This leads to coverage distortions for the impulse responses. We propose a robust and efficient semi-parametric approach to conduct hypothesis tests and compute confidence bands in the SVAR model. The method exploits non-Gaussianity when it is present, but yields correct coverage regardless of the distance to the Gaussian distribution. We evaluate the method in a simulation study and revisit several empirical studies to highlight the limitations of using non-Gaussianity for identification.

## Constructing the Term Structure of Uncertainty from the Ragged Edge of SPF Forecasts

Presenter: Gergely Ganics
Co-authors: Gergely Ganics;Todd Clark;Elmar Mertens

We construct term structures of expectations and uncertainty that are consistent with point and density predictions observed from the Survey of Professional Forecasters (SPF), or similar forecast sources. We derive a state space model that exactly matches any set of fixed-horizon or fixed-event forecasts that can be observed in the data. Model-implied expectations can be set to equal observed point forecasts not only in the model's prior but also in the model's posterior. In addition, we can match observed SPF density predictions (histograms) by application of entropic tilting. Applied to data from the US SPF, we construct fixed-horizon fan charts for up to four years. We document considerable variation in forecast uncertainty. In response to the onset of the COVID-19 pandemic, tilting with annual SPF histograms considerably reduces model-based estimates of forecast uncertainty.

## Generating synthetic load profiles of buildings from metadata

Presenter: Evgenii Genov
Co-authors: Evgenii Genov;

While deployment of smart meters has been carried out rapidly across Europe, the data availability remains an important issue in the energy sector. Synthetic Load Profiles are a more realistic and variable extension to the established Standard Load Profiles, which are developed for general types of customers in the electricity market. The profiles make a more accurate representation of electric demand over time for individual or a particular group of users. Moreover, Synthetic Load Profiles are suitable for forecasting and simulation of electric demand of a consumer with specified characteristics. Such functionality is critical for design and energy management on a microgrid level. Current approaches for generation of synthetic load profiles adopt either bottom-up models simulating the operation of select devices or applying a prior clustering to identify a representative profile. This paper proposes a novel data-driven method to generate unique stochastic profiles, with quantified uncertainty estimates, based on the given metadata, such as the information on users' demographics, lifestyle, and home sizes. The main benefit of the proposed strategy is no requirement in a prior clustering or knowledge of the yearly consumption of a household. The method is based on a global regression model, using a supplementary local model for predicting the yearly consumption value, which is used for scaling. The level of uncertainty is quantified using a conformal prediction framework. The obtained synthetic profiles are validated using out-of-sample time-series. We demonstrate the effectiveness of the method by comparing it with existing baselines on the CER Smart Metering Project open-access dataset.

## Iterative Load Shifting Disaggregation Algorithm

Presenter: Colin Quinn
Co-authors: Colin Quinn;Richard Povinelli;Ronald Brown

It is a common problem in natural gas demand time series forecasting that the required granularity level of the forecast does not match the existing granularity level of the historical gas consumption time series data. For example, only historical cycle-billing data is available and a daily forecast is needed. A forecaster can obtain an estimated, higher frequency series through temporal disaggregation of low frequency measurements to forecast at the desired frequency. Current direct-disaggregation methods are focused on the transformation of a single low frequency time series with uniformly spaced intervals and are unable to operate over multi-source time series with nonuniform and overlapping intervals. In this paper, we introduce an Iterative Load Shifting (ILS) disaggregation algorithm used to disaggregate multiple data sources structured at nonuniform, low levels of aggregation into a single daily representation while maintaining the direct-disaggregation constraints traditional methods rely on. Load shifting is carried out in an iterative two-step process; where first a prediction phase uses multi-parameter linear regression to

generate high frequency time series based on the relationship between the observable low frequency series and high frequency independent correlated variables. Then second, load profiles are used in a piecewise linear update phase to "shift" or redistributing the low frequency load observations amongst the underlying high frequency periods within each low frequency measured interval. This two-step process is repeated, with each subsequent prediction phase modeling the updated estimates produced in the preceding update phase. The Iterative Load Shifting disaggregation algorithm produces an 8.6% MAPE evaluated over a three-year period and the most accurate disaggregation results as shown in the case studies included in this work.

## Electricity Demand Forecasting: Modelling the COVID-19 Lockdown in Spain

Presenter: Eduardo Caro
Co-authors: Eduardo Caro;Jesús Juan

Many governments around the world had to enforce a strict lockdown during the coronavirus COVID-19 pandemic, closing or limiting all non-essential businesses and ordering citizens to stay at home. As a result of this decision, the electric power demand pattern varied significantly during the affected months. Considering that the electric consumption forecasting models are estimated or trained using historical data, their numerical accuracy decreased during months of confinement. In this work, the electric power consumption in Spain during the lockdown has been analyzed, and the forecasting software (currently used by the Spanish Transmission System Operator) has been modified to: (i) properly model the lockdown policies during the year 2020, and (ii) to anticipate the change of behavior of load consumption during working and non-working days. In order to evaluate the methodology performance, a real data set from the Spanish market has been used and the developed models has been tested employing the RMSPE (Root Mean Squared Percentage Error) as the accuracy metric.

## Credit Image: A method to convert data for credit scoring to images for convolutional Neural Networks

Presenter: Junhao Liang
Co-authors: Junhao Liang;Xingjie Wei;Barbara Summers

Improving the forecasting capability of credit scoring models has long been a concern of financial institutions. Various prediction models have been tested to improve predicting loan defaults on heterogeneous data (e.g. tabular data). However, it is still challenging for credit scoring models to utilise powerful deep learning networks that perform extremely well on homogeneous data (e.g., images). To meet these challenges, we propose a novel credit data presentation method called 'credit Image' that converts tabular credit data into images for credit scoring to take advantage of the powerful 2D convolutional neural networks (CNN), which are widely used in visual pattern recognition. Firstly, a novel binning algorithm, the best-KS algorithm, is applied to the credit data to identify the optimal bin allocation that can maximise the distance between normal and default accounts. Then a Weight of Evidence (WOE) transformation is conducted on the discretised data to transform credit data into homogeneous data. After that, a linear segmented colour map is generated to map the WOE values into colours. Finally, a blank image is formed and filled with colours, with the area of each feature on the image set by the ratio of each feature's information value to the sum of all features' information values. The 'Credit Image' is then applied to a range of credit scoring benchmark datasets and the transformed data is used as an input for several 2D CNN models. Our result demonstrates that basic 2D CNN models can achieve comparable performance results to state-of-the-art credit scoring models.

## A verifiable estimation and parametric inference of nonlinear equations using neural networks

Presenter: Luigi Longo
Co-authors: Luigi Longo;Mohammad Soltanieh-ha

Methodologies such as situational importance and Shapley Additive explanations (SHAP) are increasingly used to explain machine learning (black-box) models' predictions. Based on parametric inference, we develop a methodological framework that uses Shapley-features relation to generalize the data generating process learned by long short-term memory (LSTM) neural networks. We provide a Monte-Carlo analysis proving that it is possible to retrieve the exact data generating process of the data from the LSTM training when the process is additive. We make an empirical application where LSTM is trained on US macroeconomic data from 1970Q1 to 2007Q4 to make inference on the data generating process. We find significant polynomial relations between the variables that hold for a test period from 2008Q1 to 2019Q4.

## Forchestra: Towards a Scalable and Flexible Time Series Prediction Framework for Demand Forecasting

Presenter: Young-Jin Park
Co-authors: Young-Jin Park;

Demand forecasting is a crucial component of Supply Chain management for revenue optimization and inventory planning. Traditional time series forecasting methods, however, have resulted in small models with limited expressive power because they have difficulty in scaling their model size up while maintaining high accuracy. In this paper, we propose Forecasting orchestra (Forchestra), a scalable and flexible framework capable of accurately predicting future demand for a diverse range of items. Forchestra consists of two parts: 1) base predictors and 2) a neural conductor. For a given time series, each base predictor outputs its respective forecast based on historical observations. On top of the base predictors, the neural conductor adaptively assigns the importance weight for each predictor by looking at the representation vector provided by a representation module. Finally, Forchestra aggregates the predictions by the weights and constructs a final prediction. In contrast to previous ensemble approaches, the neural conductor and all base predictors of Forchestra are trained in an end-to-end manner; this allows each base predictor to modify its reaction to different inputs, while supporting other predictors and constructing a final prediction jointly. We empirically show that the model size is scalable to up to 0.8 billion parameters ($\approx$ 400-layer LSTM). The proposed method is evaluated on our proprietary E-Commerce (100K) and M5 (30K) dataset, and it outperforms existing forecasting models with a significant margin. In addition, we observe that our framework generalizes well to unseen data points when evaluated in a zero-shot fashion on downstream datasets. Last but not least, we present extensive qualitative and quantitative studies to analyze how the proposed model outperforms baseline models and differs from conventional ensemble approaches.

## Multilevel Lifetime Value Forecast for VR Users at Meta

Presenter: Slawek Smyl
Co-authors: Slawek Smyl;Huigang Chen;Pablo Solis;Yilun Chen

User lifetime value forecast provides an important input for customer management and product enhancement in non subscription based platforms. We developed a set of multilevel, both user-level and cohort-level, lifetime value models to address the analytical and product needs for VR users at Meta. The key modeling challenge is the sparsity and irregularity of the user spend pattern. We use a special model architecture in the Recurrent Neural Network Time Series family (RNN-TS) to achieve high performance at both user level and aggregate level forecast. This includes innovative RNN cells and layers of small (state size) cells. In addition, we introduce a new Causal RNN-TS model to estimate the heterogeneous treatment effect of product intervention. The model also has built-in capability that generates feature importance for model interpretability.The models attain improvements in accuracy evaluated with sMAPE and absolute difference as compared with several baseline models, such as tree-based models and parametric statistical models.

## Forecasting the tourism demand cycle: A Markov-VAR approach

Presenter: Andrea Saayman
Co-authors: Andrea Saayman;Ilse Botha

Forecasting the tourism demand cycle has received increasing attention over the past two decades, with research efforts focused on predicting turning points in the cycle and using leading indicators in forecasting cyclical behaviour. The research falls within the second category and aims to investigate to what extent the business cycles of origin countries predict cyclicality in tourism demand for South Africa. South Africa is a developing country and tourism destination, with its main international source markets, the United States, the United Kingdom, Germany, France and the Netherlands. The methodology includes a Markov-VAR approach, which differs from the approaches followed by other researchers. Forecasts from the Markov-VAR model will be compared to time-series approaches. The findings of this study will shed some light on the predictability of tourism demand cycles, which could assist tourism managers and policy makers to manage and mitigate cyclicality in tourism demand.

## Investigation of substitution effects and adoption patterns across Swiss hotel distribution channels - A multigeneration perspective

Presenter: Miriam Scaglione
Co-authors: Miriam Scaglione;Roland Schegg

The evolution of distribution channels in the hospitality industry has varied over time depending on the technology used. In IFS 2014, this research team presented an analysis of the evolution of the market shares of different clusters or generations of distribution channels using multi-generation diffusion methods for an annual series from 2002 to 2014. The data comes from an annual member survey of the Swiss hotel association HotellerieSuisse, which has been monitoring the development of the market shares of 15 individual distribution channels since 2002. The aim of this survey is to divide the distribution channels into three generations, starting with the pre-world-wide-web era (travel agencies, tour operators, etc.); the middle generation includes internet-based direct booking channels (email, reservation forms on hotel websites, etc.) and the latest generation includes online intermediaries such as online travel agencies (OTA) and social media. The objectives of this research are twofold. Firstly, to extend the analysis for a further 7 years to 2021 and assess the accuracy of previous forecasts. Second, to shed light on the impact of Covid-19 on the dynamics of competing sales funnels. From a practitioner's perspective, the study provides insight into the future evolution of the sales mix. This evolution is critical for hotel marketing strategies, as the Internet is the most powerful marketing tool in the hospitality industry today.

## Time Series Forecasting in the Field of Property Management in Tourism Regions

Presenter: Corsin Capol
Co-authors: Corsin Capol;Stefano Balestra

With the rise of alternative accommodations such as rental holiday homes, the need for a service to manage such accommodations is growing. The planning of human resources is still a major challenge. This is especially true for the management of multiple holiday homes where distances between properties vary and demand is flexible. The goal is therefore to develop a software that predicts the number of people departing for each day of the next four weeks. Within this context, the forecast supports the management in mid-term human ressource planning. The authors offer a research prototype, predicting a flexible demand and translating this demand into a staff requirement. The time series exhibits several interesting characteristics. It consists of strictly positive integer values, of which many are zeros, it shows multiple seasonalities with different cycles and numerous anomalies due to Covid-19. Multiple state-of-the-art models and methods were tuned and tested until Facebook Prophet was chosen as forecasting technique. The addition of context specific regressors to the model is essential for generating valuable forecasts. The prototype software allows

the training of the model once new data becomes available. Furthermore, its integration into the staff assignment planning and time recording software of the tourism property management company ensures its ease of use.

## Can Fashion Forecast Tourism?

Presenter: Emmanuel Silva
Co-authors: Emmanuel Sirimal Silva;Hossein Hassani

The United Kingdom (UK) is a world-renowned fashion hub where the economic importance of the tourism sector was recording continuous growth prior to the pandemic. Interestingly, tourism shopping is widely experienced yet seldom discussed from a tourism demand forecasting context. Driven by the potential relevance of tourism shopping and in hope of motivating increased collaboration between the tourism and creative industries, we analyse whether fashion retail sales can be a leading indicator for tourism demand in the UK. Adopting the Multivariate Singular Spectrum Analysis leading indicator algorithm, we forecast UK tourism demand and compare the results with six benchmark univariate forecasts. We find statistically significant evidence for the existence of cross-sector relations between the UK's fashion and tourism industries.

## Density forecast comparison in small samples

Presenter: Laura Coroneo
Co-authors: Laura Coroneo;Fabrizio Iacone;Fabio Profumo

We apply fixed-b and fixed-m asymptotics to tests of equal predictive accuracy and of encompassing for density forecasts. To accommodate forecasts reported as probabilities for intervals, we use two loss functions: the Quadratic Probability Score and the Ranked Probability Score. With these loss functions, both tests can be performed in the popular Diebold and Mariano (1995) (DM) framework, which is well-known to suffer from small sample size distortions, see Diebold and Mariano (1995) and Clark (1999).

To overcome the small sample bias of the DM test, we use an alternative approach based on fixed-smoothing asymptotics. In particular, we consider fixed-b asymptotics by Kiefer and Vogelsang (2005) and fixed-m asymptotics by Hualde and Iacone (2017). This approach proved capable of eliminating size distortion in the DM test for comparing point forecasts, see Coroneo and Iacone (2020) and Coroneo, Iacone, Paccagnini and Monteiro (2022). In an original Monte Carlo exercise, we document that standard asymptotics deliver unreliable tests in small samples, and we verify that fixed-b and fixed-m asymptotics can be used with success to perform tests of equal predictive accuracy and of encompassing for density forecasts.

We apply our methodology to test the accuracy of the ECB SPF density forecasts for three key macroeconomic variables (real GDP growth, inflation and the unemployment rate). We are interested in establishing whether ECB SPF density forecasts can beat and/or encompass simple benchmarks, such as a uniform distribution, a Gaussian random walk distribution, and a naive forecast taken from the previous round of ECB SPF forecasts.

Results indicate that ECB SPF density forecasts for unemployment and real GDP growth outperformed and sometimes encompassed the benchmarks, especially at one-year ahead. On the contrary, survey forecasts for inflation do not easily outperform nor encompass the benchmarks. For all the variables, however, we find evidence of an improvement in predictive ability since 2010, suggesting a change in the forecasting practice after the financial crisis. We also find that the professional forecasters always encompass the naive benchmark, indicating that they update their information set when making their predictions.

## Forecasting Colombian inflation in real time

Presenter: Hector Zarate
Co-authors: Hector Zarate;Miguel Angel Manrique

This paper analyzes the forecasting performance of some machine learning techniques and statistical models with high dimensional administrative data to assemble accurate forecasts of the Colombian consumer price inflation at different horizons. We focus on a real-time forecasting exercise up to one year ahead using a large number of macroeconomic predictors available at the time when the forecasts were made. The statistical methods rely on dimensionality reduction techniques and penalized models to overcome overfitting and improve inflation forecasts' accuracy. The main findings indicate that the proposed modeling can identify subsets of explanatory variables or latent factors providing alternative methods to macroeconomic forecasting.

## Evaluating Probabilistic Classifiers: The Triptych

Presenter: Alexander Jordan
Co-authors: Alexander Jordan;Timo Dimitriadis;Tilmann Gneiting

Predicting the occurrence probability of binary events is presumably the most common forecasting task throughout the sciences. We propose a "triptych" of evaluation displays for binary probabilistic classifiers consisting of the receiver operating characteristic (ROC) curve, the CORP reliability diagram, and the Murphy diagram. Individually, these three displays focus on different and complementary aspects of the forecast's performance. The ROC curve assesses discrimination, the reliability diagram evaluates calibration, and the Murphy diagram combines both properties and visualizes overall predictive ability. This intuition is supported by a theoretical result connecting these plots: For autocalibrated forecasts, the ROC curve and the Murphy diagram display congruent information. Combined, these three displays give a full picture of the forecast's predictive ability. We illustrate the proposed triptych in several applications.

## A methodology for identifying new U.S. non-stop flying routes using algorithms from Machine Learning

Presenter: Maria Rosa Nieto
Co-authors: Maria Rosa Nieto;Rafael Bernardo Carmona

This paper introduces and applies the concept of unmet demand, or the demand that has not been satisfied, for airline passenger transportation industry. The unmet demand may exist because of the economic effects of supply and demand. The estimation of the unmet demand is important for strategic decisions. In literature, there are evidences of studies and analysis that tackle the problem of calculating the unmet demand, mainly from the point of view of human, economic and financial needs. Recently, it has been studied in medicine and Supply Chain but not in transportation. All the studies, propose approaches for forecasting the unmet demand. This paper proposes a methodology to estimate the unmet demand per origin destination (OD pair) in five steps. Step 1, segment airports in airport clusters using a machine learning classification method; step 2, segment OD pairs in subclusters (ORI-DES relation) OD pairs are classified into subclusters according to the relation between the ORI cluster and the DES cluster; step 3, forecast the total market size of the OD pairs subclusters (ORI-DES relation) with the ARIMA-GARCH-Bootstrap time series method developed by Nieto & Carmona-Benitez (2018); step 4, forecast the pax demand of the OD pair with the ARIMA-GARCH-Bootstrap time series method; and step 5, calculate the OD pair unmet demand with the mathematical model proposed in this paper. Data from The United States domestic air passenger transport market (1st quarter of 1993 to the 4th quarter of 2012) is used for estimation and models, and data from the 1st quarter of 2013 to the 4th quarter of 2015 is used for the model validations.

## Forecasting Appeals in Polish Corporate Income Tax Cases Involving Bilateral Tax Agreement

Presenter: Dominik Gajewski
Co-authors: Dominik Gajewski;Kamil Jonski

Full-text database of the universe of Polish Administrative Courts Tax-related verdicts had been queried to select the body of cases adjudicated over 2005-2019 period, where one of bilateral tax agreement (on avoiding double taxation) had been invoked (Poland have effective agreements with 86 jurisdictions). Resulting database of 653 first instance verdicts contains information on the case processing (date of case lodging and the first instance verdict, result, information on the appeal – and if, who brought it and – except some end-of-the-sample cases - the results of appeal) as well as full text of the first instance verdict. Since Polish tax disputes are nearly always litigated (settlements are extremely rare), collected database offers quite good insight into the decision-making of the courts and parties. Probit regression had been employed in order to identify key variables allowing for forecasting (i) the fact, that the appeal will be lodged and (ii) its effectiveness (in terms of whether Supreme Administrative Court will upheld the verdict). Sample had been randomly split for estimation sample (50%) and validation sample (25%), applied to select appropriate model specification while avoiding over-fitting. Finally, out of sample performance had been evaluated using the test sample (25%).

## Forecasting Realized Equity Volatility From Text Sentiment Revealed by Company Filings

Presenter: Massimo Guidolin
Co-authors: Massimo Guidolin;Manuela Pedio;Kevin Tikvic

We analyse the effect of sentiment on realized equity return volatility in the week following the filing of a 10-K document by the company. We combine the terms in the word lists developed in the seminal paper by Loughran and McDonald (2011), using a market-based weighting scheme to summarize word frequencies into one sentiment measure. We find that negative, positive, assertive, and litigious tones in the 10-Ks filings have a significant impact on post-filing realized volatility. Our results also show that a market-based weighting scheme produces more reliable results compared to traditional, corpus-based approaches.

## Shrinkage Estimator for Exponential Smoothing Models

Presenter: Kandrika Pritularga
Co-authors: Kandrika Pritularga;Ivan Svetunkov;Nikolaos Kourentzes Kourentzes

Exponential smoothing is widely used in the practice and has shown its efficacy and reliability in many business applications. Yet, there are cases, for example when the estimation sample is limited, that the estimated smoothing parameters can be erroneous, often unnecessarily large. This can lead to over-reactive forecasts, and high forecast errors. Motivated by these challenges, we investigate the use of shrinkage estimators for exponential smoothing. This can help with parameter estimation and mitigating parameter uncertainty. Building on the shrinkage literature, we explore $\ell_1$ and $\ell_2$ shrinkage for different time series and exponential smoothing model specifications. From the simulation and the empirical study, we find that using shrinkage in exponential smoothing results in forecast accuracy improvements and better prediction intervals. In addition, using bias-variance decomposition we show the interdependence between smoothing parameters and initial values and the importance of the initial value estimation on point forecasts and prediction intervals.

## Global Output Factor and Predictability of Commodity Prices

Presenter: Arabinda Basistha
Co-authors: Arabinda Basistha;Richard Startz

Estimation of a global output factor, often used as a measure of global demand, is a current empirical challenge facing many global macroeconomic applications. Recent research by Baumeister and Guerin (2021) documents five alternative global measures: world industrial production, global steel production based on Ravazzolo and Vespignani (2020), a shipping cost index based on Kilian (2009), Kilian and Zhou

(2018), Kilian (2019), a real commodity price factor, and a global economic factor based on the common factor of multiple global indicators including the world industrial production. The common factor based global economic factor was first proposed by Baumeister et al. (2020). Comparison of the five indicators in Baumeister and Guerin (2021) show superior predictive performance by the common global economic factor in prediction of world output growth. One important use of measures of global output factors has been commodity returns predictability. Wang and West (2014) document limited predictive information present in the world industrial production for disaggregated commodity returns. In contrast, Hamilton (2019) present predictive evidence favoring the world industrial production index relative to the shipping cost based index. A strong alternative presented by Wang and West (2014), and supported by Delle Chiaie et al. (2021), is based on a common factor of disaggregated real commodity prices. This factor shows strong predictive performance for commodity returns. This study examines the predictive performance of all five alternative factors for nominal and real, aggregated and disaggregated, commodity prices upto one year ahead. The outcomes show relatively limited in-sample predictability with the real commodity factor providing superior performance. We then propose a mixed frequency state-space model that incorporates the broader GDP measure at quarterly frequency with the monthly industrial production data for G7 countries for measuring global output factor. The in-sample predictive results are comparable with the previous output based factors although show limited improvement. We adapt our mixed frequency model to incorporate further information from disaggregated real commodity price indexes. The in-sample outcomes show 10 to 20 percent reduction in MSPE ratios for several commodities over multiple horizons, relative to other five alternatives. However, we note that the distribution of our predictive gains show a fair amount of heterogeneity.

## Markov intermittent demand

Presenter: Giacomo Sbrana
Co-authors: Giacomo Sbrana;

Intermittent data is an uncharted issue for standard Time Series Models, since they can hardly control for it. We provide a framework for modeling intermittent time series, embedding a Markov chain approach. This assumption allows introducing dependency between switching observations as well as measuring the degree of persistence of shocks across time. Since the switching is directly observable in real data, this greatly simplifies the treatment. In addition, we provide closed-form results for estimating the parameters, forecasting future observations and constructing prediction intervals. It turns out that integrating the Markov assumption within a basic teime series framework, greatly improves the predictive performance. Indeed an empirical analysis, carried out over 30000 weekly retail sales data by Walmart, shows that the proposed approach is more accurate than standard benchmarks.

## The Australian retail sector through the COVID-19 pandemic

Presenter: George Athanasopoulos
Co-authors: George Athanasopoulos;Megan Ortega;Anastasios Panagiotelis;Farshid Vahid

National governments and local authorities around the world implemented diverse policies in order to manage and contain the effect of the Covid-19 pandemic, both on health systems but also economies. That was also the case for measures implemented around Australia. An initial countrywide lockdown, implemented by the Australian federal government, was followed by several state lockdowns and border closures; with the city of Melbourne, in the state of Victoria, being declared the most locked down city in the world. In this paper we explore and analyse the effect of the various lockdowns around Australia on the retail sector and its various components. Using a forecast reconciliation approach we generate coherent and robust counterfactual Covid-free forecasts. These form a base in order to estimate the effect of the lockdowns. The results show the clear effect of the multiple lockdown measures, especially in the states of Victoria and New South Wales. Some components of the retail sector show a nice recovery and even exceed expectations after the lockdowns, while others never recover, having a severe effect on livelihoods.

## Demand forecasting for pricing in fashion retail – an overview

Presenter: Johannes Stephan
Co-authors: Johannes Stephan;

About a decade ago Zalando, a fashion e-tailer, started to implement algorithmic tools that assist pricing managers to clear seasonal stock. Over the course of various incarnations of our pricing engines, a commonality remained: each pricing engine predicted sales or demand at a given price point which in turn is used to compute and optimize profit until a specified season end. Thus, with the need for algorithmic pricing, forecasting demand has become a central domain within Zalando's Applied Science organisation. In this talk, I will give an overview of our current demand forecast that we use with the latest version of our price recommendation tool. Furthermore, I will shed light on some key challenges that we face: (1) How to infer demand from sales using partially constrained stock and (2) how to ensure that predicted demand increases when we apply higher discounts. Finally, I highlight ongoing work on testing alternative sales forecasters in a simulated marked environment.

## Retail analytics – Integrated forecasting and inventory management for perishable products

Presenter: Anna-Lena Sachs
Co-authors: Anna-Lena Sachs;Stewart Liu;Philip Kaminsky;Stefan Minner

Recent studies have caught attention to the large amounts of food wasted in retailing. Food waste can be reduced by better matching supply and demand. We investigate a retail chain that owns several stores and keeps inventory at the stores and at a central warehouse. The inventory policy must include order quantities for the warehouse and stores, and an inventory allocation policy between stores when stocks are limited. We propose a data-driven approach that integrates demand forecasting and inventory optimization by determining the optimal policy parameters as a function of exogenous variables such as price, weekday and weather.

## An integrated-signal approach to selective hedging

Presenter: Ana-Maria Fuertes
Co-authors: Ana-Maria Fuertes;Joelle Miffre;Adrian Fernandez-Perez

Extant theory and empirical evidence indicate that economic agents engage in selective hedging, namely, trading in futures contracts with the dual objective of covering the price risk of their spot position and to earning a premium through pure speculation. This paper proposes a selective hedge ratio with a risk-minimizing or pure hedge component that is aimed purely at covering spot price risk and a speculative component that is aimed at extracting an excess return. As in prior studies, the speculative component is inversely related to the risk aversion of the hedger and to the uncertainty associated with the spot position of the commercial trader as in the theoretical models of Anderson and Danthine (1983) and Stulz (1984). The main novelty of the selective hedge proposed is that it allows economic agents additionally to speculate upon the information content of any number of predictors for futures returns. The selective ratio proposed builds on the fruitful notion of characteristic- or signal-integration in the portfolio management literature (e.g. Brandt et al., 2009; Fernandez-Perez et al. 2019) by proxying the expected change in the futures price by a linear combination of predictors – the integrated signal. We deploy both the selective hedge and the underlying pure hedge individually on 19 commodities covering four major categories – agriculture, energy, livestock and metal. Seeking to aid the hedging problem faced by commercial traders in real time, we sequentially construct out-of-sample hedging portfolios at sample week t based on past information. The integrated-signal blends various types predictors – commodity characteristics (futures term structure, hedging pressure, momentum and basis momentum), early indicators of economic activity, and purely directional price-trend indicators from technical analysis. The results indicate that while the selective hedge covers the risk of the spot position as effectively as the pure hedge, it provides a Sharpe ratio improvement at 58% on average across commodities. This added profitability is partly compensation to the

commodity producer for the drawdowns incurred when funding liquidity plunges and economic conditions worsen. The findings survive the consideration of transaction costs, sub-period analyses, and redesigns of the pure risk-minimizing and speculative components of the selective hedge.

## Modelling and adaptive forecasting of trends in temperature distributional characteristics under structural change

Presenter: Robinson Kruse-Becher
Co-authors: Robinson Kruse-Becher;

It is widely accepted that global warming is defined as an increasing trend in the global temperature mean. Gadea and Gonzalo (2020, JoE) find strong evidence in favor of positive trends, not only in global temperature means, but also in various quantiles. We follow their important contribution and analyze global warming by modelling and short-term now- and forecasting trends in distributional characteristics of global temperatures beyond the mean. First, we apply a dynamic stochastic coefficient process as proposed by Giraitis, Kapetanios and Yates (2014, JoE) to model temperature anomalies. Thereby, we decompose the climate series into a random persistent attractor and a dynamic part with a time-varying autoregressive coefficient. Smooth structural changes in trends of temperatures are clearly documented. Second, we apply adaptive now- and forecasting methods for climate time series of distributional characteristics. When forecasting under recent and ongoing structural changes, robust adaptive methods can be quite useful to enhance forecast accuracy, see Giraitis, Kapetanios and Price (2013, JoE) in the short-run. Therein, recent and past observations are weighted in various ways (e.g. exponential smoothing, rolling window and nonparametric). Most schemes are quite sensitive to the unknown underlying trend, potential structural breaks and persistence. Tuning parameters (e.g. bandwidth and weighting coefficients) are selected in a data-dependent way via cross-validation techniques. We consider a balanced annual raw station panel data set collected from more than 10,000 weather stations around the globe (1850-2020, provided by the CRU). This enables us to investigate how successful adaptive forecasting devices work around historical structural changes in global temperature distributions at several breakpoints, over the last centuries. Moreover, we study recent developments and consider the case of rapidly changing trends in lower temperature quantiles relative to upper quantiles and mean averages.

## Are consensus FX forecasts valuable for investors?

Presenter: Michał Rubaszek
Co-authors: Michał Rubaszek;Joscha Beckmann;Marek Kwas

We establish a new link between the cross-section of currency returns and survey-based forecasts. Using data from Consensus Economics, we show that surveys provide trading signals which are not entirely driven by benchmark trading strategies: momentum, carry and value. We evidence the sizable economic value of survey-based trading strategies, as they provide additional excess returns of up to two percentage points per year compared to benchmarks. This illustrates that professionals effectively explore available information and their expertise can be used to diversify exchange rate portfolios. Our findings are robust against various tests and different currency portfolios structures.

## Modelling Okun's Law – Does non-Gaussianity Matter?

Presenter: Tamás Kiss
Co-authors: Tamás Kiss;Hoang Nguyen;Pär Österholm

In this paper, we analyse Okun's law – a relation between the change in the unemployment rate and GDP growth – using data from Australia, the euro area, the United Kingdom and the United States. More specifically, we assess the relevance of non-Gaussianity when modelling the relation. This is done in a Bayesian VAR framework with stochastic volatility where we allow the different models' error distributions to have heavier-than-Gaussian tails and skewness. Our results indicate that accounting for heavy tails yields

improvements over a Gaussian specification in some cases, whereas skewness appears less fruitful. In terms of dynamic effects, a shock to GDP growth has robustly negative effects on the change in the unemployment rate in all four economies.

## Does forecast optimism affect outcomes? A re-examination of Beaudry-Willems

Presenter: Christopher Gilbert
Co-authors: Christopher Gilbert;

Beaudry and Willems (BW, American Economic Journal: Macroeconomics, January 2021) examine how forecasting errors affect macroeconomic performance. They look specifically at IMF forecasts of GDP growth rates for 189 member countries published in the semi-annual World Economic Outlook (WEO) review. They argue that over-optimism about growth prospects may result in excessive expenditure or over-investment with adverse consequences for debt levels so that short term stimulus is offset by negative longer term consequences. Re-examination of their data fails to find any support for this claim. First, BW misinterpret their own results. Using a large panel, they regress growth outcomes on a forecast accuracy variable defined as the average of the one, two and three year forecasts made three years previously. This variable includes the forecast error for the current year. Recognizing the endogeneity of this variable, they instrument using the past forecast accuracy of the relevant IMF country managers. I recast their estimated equations in terms of the extent to which the current WEO forecast, rather than the forecast error, influences the growth outcome. The two formulations are observationally equivalent but, making this switch, it becomes clear that the forecast errors drop out of the relationships. The presence of the current forecast error on the right hand side of the BW equations confuses interpretation. The implications are the opposite of BW's claims. Second, standard Granger non-causality analysis that shows that forecast residuals have no incremental predictive power for growth outcomes. This underlines the claim that the conclusions in BW derives from inclusion of the current forecasting error in their measure of over-optimism. BW argue that forecasters should be cautious and specifically that the IMF should curb the over-optimism of their country directors. While over-optimism is to be discouraged, there is no evidence that it has the impact BW claim.

## High-Resolution Peak Demand Estimation Using Generalized Additive Models and Deep Neural Networks

Presenter: Jonathan Berrisch
Co-authors: Jonathan Berrisch;Michał Narajewski;Florian Ziel

This paper presents a method for estimating high-resolution electricity peak demand given lower resolution data. The technique won a data competition organized by the British distribution network operator Western Power Distribution. The exercise was to estimate the minimum and maximum load values in a single substation in a one-minute resolution as precisely as possible. In contrast, the data was given in half-hourly and hourly resolutions. The winning method combines generalized additive models (GAM) and deep artificial Neural Networks (DNN) which are popular in load forecasting. We provide an extensive analysis of the prediction models, including the importance of input parameters with a focus on load, weather, and seasonal effects. In addition, we provide a rigorous evaluation study that goes beyond the competition frame to analyze the robustness. The results show that the proposed methods are superior, not only in the single competition month but also in the meaningful evaluation study.

## Peak electric load days forecast performance improvement by combining base models predictions using machine learning based ensemble methods

Presenter: Omar Aponte
Co-authors: Omar Aponte;Katie McConky

In the energy forecasting domain, as well as in other forecasting domains, ensemble forecasting ap-

proaches that combine the results of two or more base models often outperform the base models that make them up. These ensemble models tend to deliver better generalization performance by integrating a number of base models to generate a final forecast. This integration can compensate for the individual imperfections of the base models. Ensemble classification techniques have consistently been top performers at generating day-ahead peak electric load days forecasts in the published literature. However, the published works have only evaluated arithmetic based ensemble classification techniques such as majority vote and single vote for this particular application. These techniques are incapable of automatically determining a level of importance or weight to assign to each base model forecast based on the model's past performance. Machine learning techniques such as classification trees, classification random forest, adaptive boosting, and artificial Neural Networks on the other hand, are designed to automatically determine weights for each input based on how much each input contributes to achieve an accurate result. Our research evaluates the performance of each of these machine learning techniques as an alternate ensemble mechanism to forecast if an upcoming day will be a peak electric load day or not based on the forecast provided by diverse base models and additional inputs. The performance of these machine learning based ensemble approaches is then compared to that of their published arithmetic based counterparts, and to that of the base models that make them up. Results based on real electricity consumption data collected specifically for this research from three different types of consumers (industrial, educational, and residential) will be presented. The methodology is applicable to consumers with and without behind the meter electricity generation. This research provides electricity consumers with improved and more accurate ensemble peak electric load days forecasting techniques that allow the consumers to minimize the financial impact of demand charges. Furthermore, classification trees and classification random forest techniques will provide important insight to determine the level of importance of each base model in generating the most accurate forecast.

## A Hybrid Weighted LSTM-NBEATS-RBFNN Method for Day-ahead Peak Load Forecasting

Presenter: Abhishek Sharma
Co-authors: Abhishek Sharma;Sachin Jain

Day-ahead peak load forecasting plays a crucial role in the reliable operation of the power system. The volatile nature of load profiles, renewable integration, rapid climate change even further increases the importance of accurate peak load forecasting. Traditional and machine learning approaches are widely used to predict load. However, due to several assumptions and preprocessing needs, they fail to provide accurate and consistent predictions of peak load. This paper presents a hybrid method comprising of Long Short Term Memory, Neural Basis Expansion Analysis for Time Series, Radial Basis Function Neural Network LSTM-NBEATS-RBFNN. LSTM is used to transform the inputs into a latent/hidden space that has more meaningful temporal dynamics. These hidden spaces are fed to NBEATS which acts as multivariate regression. RBFNN is used to fuse the outputs of the NBEATS. A weighting mechanism is proposed based on Dynamic Time Warping (DTW), using which similarity between the actual and forecasted value is found. The weights are proportional to the similarity i.e., more emphasis is given to the forecast having more similarity with actual output during model training process. Finally, the weighted average of all outputs is taken to provide a single forecast. The performance of the proposed method has been validated and compared on the actual load data of three geographical zones Madhya Pradesh State (MP), India. The results confirm that the proposed methodology achieves higher and more consistent forecasting accuracy with low variations in forecast error.

## How concepts like Feature Evolution, Automatic Machine Learning and ML Ops are rapidly advancing AI applications for electrical grid operators, utilities and energy suppliers.

Presenter: Jon Farland
Co-authors: Jonathan Farland;

Over the last decade, there has been extensive global investment in "Smart Meter'' technology used

to meter both electricity and natural gas consumption at service points across entire distribution networks. This has led to an explosion of high-frequency time series data that can conceivably help manage demands, optimize behind-the-meter generation like solar, and even improve electrical load forecasting visibility and accuracy across the grid. Unfortunately, many of the regulators, rate payers and stakeholders of this investment are still asking the question: "how exactly can we use this data?". In this situation, being able to rapidly prototype, experiment and test analytical concepts requires the tools and the mindset to learn what works and what doesn't. Concepts such as Feature Evolution and Automatic Machine Learning allow data scientists in the energy industry to experiment extensively with their data at scale and to quickly learn what algorithms and applications are the most meaningful. Well-built tools that facilitate Machine Learning Operations (ML Ops) also enable data scientists to quickly put models into production, track their performance, and alert them when predictive performance has degraded, or when the data generating process has significantly 'drifted'. Lastly, being able to build light-weight Apps on top of these AI engines allow forecasters to quickly demonstrate the performance and value of their models in near real-time. This talk will introduce several advanced analytics use cases being deployed across the energy industry, as well as provide a live demonstration of how the H2O.ai Cloud can bring state-of-the-art predictive performance to these use cases in hours instead of days or weeks.

## A Machine Learning based Approach to Forecast Hierarchical Time Series using Non-linear Mappings

Presenter: Kasun Bandara
Co-authors: Shanika Wickramasuriya;Kasun Bandara;Hansika Hewamalage;Maneesha Perera

Nowadays, many operational forecasting tasks are associated with vast collections of time series spanning across several dimensions, such as product categories, geographical areas etc., that form a hierarchical structure. In these operational environments, forecasts at multiple hierarchically aggregated levels and coherent forecasts across the hierarchy (i.e., forecasts sum up across the structure as the data) are required for better and aligned decision-making and strategic planning. However, time series at different levels of aggregation possess distinct features, therefore leading to incoherent forecasts across the hierarchy when modelled individually. Traditional hierarchical forecasting methods first produce individual base forecasts for each time series in the hierarchy and then follow a reconciliation step to ensure the coherency of the forecasts. Most state-of-the-art reconciliation methods reconcile the forecasts using linear mappings to project the base forecasts onto a coherent subspace. They also use assumptions such as the forecasts are unbiased and residuals are jointly covariance stationary. In this study, we propose a non-linear forecast reconciliation approach using machine learning methods, obviating the assumptions of traditional linear reconciliation approaches. We introduce a novel loss function incorporating non-linear mappings to obtain a set of coherent forecasts from the individual base forecasts. To obtain the weights of the non-linear mapping between the base forecasts, we train a feed-forward neural network with the proposed loss function using fitted values of the base models as inputs to the network. We evaluate the proposed methodology against state-of-the-art linear reconciliation approaches on multiple benchmark datasets, and the results are very competitive.

## Machine Learning based Financial Forecasting and Planning

Presenter: Hamin Oh
Co-authors: Hamin Oh;Anbang Xu;Jerry Shan;Faisal Farooq;Phil Gaudreau;Neelima Rathi

In this paper, we will walk through a use case for our financial forecasting and planning, where we forecast bookings for 6+ months ahead, and use the model insights to allocate budget for headcount expenses, commissions expenses and marketing expenses. We develop a business-model-based forecasting approach, in which we mimic the business practice and process with forecasting intermediate metrics first and then arriving at the final ones. For example, instead of forecasting renewal bookings directly, we forecast Renew Increment Growth (RIG) and Renew Target Amount (RTA) separately, and then use their product as the forecasted renewal bookings. By breaking down with more interpretable pieces, we achieve

better accuracy on average, provide predictions at lower granularity, and obtain more explainability and insights. Utilizing business metrics (e.g. baseline, RTA closed) as indicators, we on average improved the MAPEs by 7% points and decreased the volatility by 2% points. We estimate a 10% savings on every dollar correctly forecasted in regards to headcount investments, sales compensation structure, and opportunity cost.

## Reconciliation of low frequency quantile forecasts

Presenter: Alaleh Razmjoo
Co-authors: Alaleh Razmjoo;Slawek Smyl

When the underlying data is of high frequency (e.g. 1 minute) but the task is to forecast at much lower frequency (e.g. daily), it is unworkable to forecast at the base (high) frequency. In such situations data is often preprocessed into lower frequency (e.g. 995th percentile at daily grain) and the forecast is made in this low-frequency domain. However a reconciliation problem arises when attempting hierarchical forecasting (e.g. network traffic of a region vs. traffic of data centers within the same region) and existing solutions do not cater for this situation. We propose a hierarchical reconciliation method that extends definition of the summation matrix used in standard (average) approaches, into one that reflects relationship between components and their aggregate, when they are quantiles, so when they do not aggregate with sum.

## The effect of Big Data and Artificial Intelligence on Forecasting in Defence and Military Applications

Presenter: André Hoogstrate
Co-authors: André Hoogstrate;

In this research an analysis of the effects Big Data and Artificial Intelligence (AI) have and will have on the practice of forecasting for defence in military applications is given. Big Data brings an abundance of heterogeneous data from many sources to the table and AI and its subfield Machine Learning present the forecaster with a whole host of new models, algorithms, and forecasting methods. It is foreseen that forecasting practices will be greatly influenced by these two developing trends. The impact on applications at the strategic, operational as well as tactical level is considered.

## Honest calibration assessment for binary outcome predictions

Presenter: Marius Puke
Co-authors: Marius Puke;Timo Dimitriadis;Lutz Dümbgen;Alexander Henzi;Johanna Ziegel

Probability predictions from binary regressions or machine learning methods ought to be calibrated: If an event is predicted to occur with probability $x$, it should materialize with approximately that frequency, which means that the so-called calibration curve $p(x)$ should equal the bisector for all $x$ in the unit interval. We propose honest calibration assessment based on novel confidence bands for the calibration curve, which are valid only subject to the natural assumption of isotonicity. Besides testing the classical goodness-of-fit null hypothesis of perfect calibration, our bands facilitate inverted goodness-of-fit tests whose rejection allows for the sought-after conclusion of a sufficiently well specified model. We show that our bands have a finite sample coverage guarantee, are narrower than existing approaches, and adapt to the local smoothness and variance of the calibration curve $p$. In an application to model predictions of an infant having a low birth weight, the bounds give informative insights on model calibration.

## Testing for equal predictive accuracy with strong dependence

Presenter: Fabrizio Iacone
Co-authors: Fabrizio Iacone;Laura Coroneo

We revisit the properties of the Diebold and Mariano (1995) test in presence of autocorrelation in the loss differential. This situation can arise not only when a forecast is sub-optimal under MSE loss, but also when it is optimal under an alternative loss, or it is evaluated on a short sample, or when a forecast with weakly dependent forecast errors is compared to a naive benchmark.

We characterise strong dependence as local to unity as in Phillips (1987) and Phillips and Magdalinos (2007b). This definition is well suited to derive reliable guidance when the sample is not very large, as it is the case in many applications. With this definition, the strength of the dependence is determined also by the sample size: a stationary AR(1) process with root close to unity may be treated as weakly dependent in a very large sample, but standard asymptotics may be poor guidance for cases with smaller samples, and local to unity asymptotics may be more informative.

We show that the power of the Diebold and Mariano (1995) test decreases as the dependence increases, making it more difficult to obtain statistically significant evidence of superior predictive ability against less accurate benchmarks. We also find that after a certain threshold the test has no power and the correct null hypothesis is spuriously rejected. Taken together, these results caution to seriously consider the dependence properties of the selected forecast and of the loss differential before the application of the Diebold and Mariano (1995) test.

To illustrate the problems associated with the DM test when there is dependence in the loss differential, we consider the case in which an AR(1) forecast for inflation in the Euro Area is compared to two naive benchmarks: a constant 2% prediction (that represents the inflation target in the Euro Area) and a moving average prediction. These benchmark predictions have highly dependent forecast errors. As a consequence, the loss differential is dependent, and the DM test fails to reject the null of equal predictive accuracy, even if the benchmarks are less accurate than the AR(1) forecast for short forecasting horizons.

## Machine Learning Methods are Better – But When?

Presenter: Keith Ord
Co-authors: Keith Ord;

The M5 Competition suggested that machine learning (ML) methods led to significant improvements in forecasting performance. But under what circumstances is this conclusion justified and how should more traditional statistical procedures be modified to benefit from the finding of M5? We focus primarily upon interval estimation and first examine how the data structure and the choice of criteria in M5 affected the conclusions. M5 attracted a large number of (mostly) ML entries and this element also had an impact upon the results. Finally, we consider the lessons for statistical (and ML) model-building going forward.

## Using big data analytics to forecast governance performance

Presenter: Georgios Karamatzanis
Co-authors: Georgios Karamatzanis;Konstantinos Nikololopoulos;Anna Tilba

Forecasting the firm's performance is critical for all stakeholders. Big data analytics can process data in volatile, complicated circumstances and help study organisations in near-real-time to forecast their response and predict an event's likelihood. They can assist organisations' governing bodies discover patterns for developing emerging possibilities, risks and how to adapt appropriately. This study aims to create a model in which interventions at the governance level can lead to better performance for the company. We aspire that by improving corporate governance mechanisms through interventions, forecasting games, new technology, new data analysis path/data set, we can forecast corporate governance performance and improve the firm's performance. This will allow the respective value of the firm to be improved by adopting a new processing style of data which will allow us to improve forecasting.

## Monthly active user forecasting and its positive externalities at Frontify

Presenter: Michal Lapinski
Co-authors: Michal Lapinski;

Frontify empowers people to connect and build brands together, by making branding incredibly simple, accessible, and fun. One of our key measures for the successful adoption and pricing of our product is the number of monthly active users (MAU). We are currently developing MAU forecasting models to support Customer Success, Sales, Finance and the management team. The project is in a very early phase and generating forecasts for about 80% of the monthly active users of Frontify on customer level. We apply state of the art bivariate and multivariate forecasting models for each individual customer and also panel data based forecasts. Some of our main challenges are identifying key drivers for MAU due to little present business knowledge and short time series in a big data context. Nevertheless, our present results are very promising and already supporting our Customer Success Managers in a prototype phase. In general, I want to present on the one hand our current state and the implementation plan for our multiple MAU forecasting projects. However, I also want to empathize and illustrate which other processes and thoughts the forecasting project has triggered.

## Spatiotemporal Modeling with General and Geographical Covariates: Insights on Crime in Philadelphia

Presenter: Nicolò Bertani
Co-authors: Nicolò Bertani;Ville Satopää;Shane Jensen

We explore whether the urban geography (i.e. the individual components of the city, such as parks, hospitals, or churches) is an important predictor of urban crime. To do this, we extend Gaussian Process modeling to high-dimensional count data and devise a general procedure to relate any spatiotemporal phenomenon with important characteristics of its environment. The procedure encodes raw spatial objects (e.g the point location of a restaurant or the perimeter of a park) into geographical covariates. Geographical and general covariates (e.g. socioeconomic indicators such as income or population density) are then used as competing predictors to estimate how, how far, and how strongly they predict crime. As a case study, we focus on the city of Philadelphia. Our results show that, when controlling for the urban geography, socioeconomic indicators are inconsequential predictors. Moreover, our results suggest possible interventions to improve the roughest parts of Philadelphia. Finally, we showcase how our model can be used for evaluating urban redevelopment plans: we analyze the ongoing redevelopment in Sharswood-Blumberg.

## A novel approach to the identification and estimation of outliers and structural changes in unobserved component models

Presenter: Matteo Pelagatti
Co-authors: Matteo Pelagatti;Antonio Candelieri;Simone Frigerio

In this work we propose a novel approach for the inference of possibly non-Gaussian unobserved component models (UCM), where outliers and structural changes are identified and estimated during the estimation process. The basic idea consists in adding at every time point a positive parameter to the variances of the disturbances that let the unobserved components evolve and then using LASSO regularization to shrink these additional parameters towards zero. The resulting model is a UCM where most variances are constant unless extreme outliers or structural changes let the additional parameter inflate to better fit the data. Our approach entails the solution of a high dimensional optimization problem and the determination of few tuning parameters governing the regularization. We alleviate the first problem by providing analytical derivatives of the regularized likelihood function and attack the second problem using only in-sample information (avoiding cross-validation) and by exploring the tuning parameters grid in an efficient way through Bayesian optimization. Our method is compared to multi-step approaches such as those implemented in SAS and STAMP and its advantages and disadvantages are discussed.

## Connecting the dots: how to make ETS work with ARIMA

Presenter: Ivan Svetunkov
Co-authors: Ivan Svetunkov;

Exponential smoothing has made a long trip from an ad-hoc forecasting method, considered by statisticians as a special case of ARIMA, to an approach with an underlying proper statistical model (called ETS) with unique properties. Over the years, ETS has been considered as an alternative to ARIMA, a different family of models, which behaves differently but has some connections with it. Still, some researchers noticed that the ETS model can be improved in some cases by the introduction of AR elements. However, nobody has done a proper analysis to identify what a combination of ETS and ARIMA could mean and how to make them work together. At most, analysts would fit ARIMA to the residuals of ETS, reflecting the idea that the two models are different. We propose using the Single Source of Error framework to unite the two models, making them work in harmony, not competing with each other. In this presentation, we will discuss what the united model implies how it can be formulated, estimated and used in forecasting. This model is already implemented in adam() function from the smooth package for R and discussed in the online textbook: https://openforecast.org/adam/

## Commonality in time series: PIC (parameters, initial states and components) and its application to vector exponential smoothing

Presenter: Huijing Chen
Co-authors: Huijing Chen;Ivan Svetunkov;John Boylan

In the context of short-term demand forecasting, businesses are often required to forecast the demand of seasonal products, based on very few complete seasonal cycles of data. This makes the forecasting task difficult, but one possible solution is to make use of the readily available cross-sectional information from groups of homogeneous data. For example, a product family may contain essentially the same product with many variations such as size or colour. It might be reasonable to assume that seasonality is common within the family of such products and can thus be estimated using cross-sectional information. Motivated by this practical problem, we develop a new taxonomy of parameters, initial states and components (PIC) to explore homogeneous features of time series. We show that the taxonomy can be applied to a multivariate model or a set of univariate models. In this presentation, the taxonomy is applied to vector ETS (VETS) models, leading to a VETS-PIC framework. We demonstrate that the PIC taxonomy affords flexibility and makes the VETS models easier to adopt and estimate, especially for seasonal time series in the retail context. Experiments with artificially generated data provide an example of how this taxonomy works. This is also evaluated empirically on M5 competition data using a variety of point forecast and prediction interval measures.

## Using Internet-of-Things Point-of-Consumption Data for Smart Replenishment

Presenter: Sandria Weißhuhn
Co-authors: Sandria Weißhuhn;Yale T. Herer;Kai Hoberg

Internet-of-Things-enabled systems that monitor inventories and sales are the latest technological advancement in demand forecasting and replenishment. Unlike traditional systems that record purchases via cash registers or RFID technology at the point-of-sale, these novel systems can track product usage via smart, connected devices at the point-of-consumption, i.e., directly at the end user. This usage data promises to be a valuable basis for automated ordering and replenishment services. We study such a system in the context of commercial coffee machines through collaboration with the B2B business unit of a large manufacturer/retailer in the coffee industry. Our data set contains information on more than 75 million drinks recorded since late 2017 by nearly 6,500 IoT-enabled coffee machines for customers such as office kitchens, restaurants, and gas stations. Challenges in this context include service providers lacking preparedness to leverage usage data in a scalable way, customers characterized by little error tolerance and motivation to cooperate, and machines only allowing for imperfect inventory visibility. We approach these

challenges by developing models for demand forecasting, inventory control, and correction of inventory record inaccuracy, only the combination of which ensures a smoothly running system. Overall, we seek to develop tools for managing smart replenishment systems in practice. Our findings extend to consumer contexts and suggest important implications for both manufacturers and retailers.

## Predictably unpredictable: How judgmental and machine learning forecasts complement each other

Presenter: Devadrita Nair
Co-authors: Devadrita Nair;Arnd Huchzermeier

Demand forecasting for seasonal products becomes especially challenging in the case of fast innovations where the product portfolio is upgraded every season. In addition to the problem of forecasting demand without any historical data, companies also have to deal with frequent stockouts, which bias past sales and provide an unreliable anchor for making new forecasts. We show how one can leverage information on comparable products from the past to improve forecasting accuracy over a purely judgmental prediction. We propose a three-step demand prediction framework using machine learning (i) to estimate the demand in the presence of stockouts for the previous year, (ii) to forecast the end-of-season demand for the current year, and (iii) to forecast the demand for the next year using expert forecasts, price and other relevant features from the previous steps. We use data from Canyon Bicycles, a manufacturer and online seller of premium bicycles to test our framework and find a forecast error reduction of 29%, measured by the weighted mean absolute percentage error (WMAPE), over a purely judgmental forecast. However, the forecast error increases by 27% if we omit the expert forecast from the model. Our study shows that expert predictions remain important for forecasting demand for rapidly innovating seasonal products. Our paper's novel approach of combining the experts' knowledge of the market and the machine learning algorithms' ability to leverage historical information works best in this setting.

## Promotional Forecasting using Machine Learning

Presenter: Michał Kurcewicz
Co-authors: Michał Kurcewicz;Piotr Zambrzycki

Accurate promotional demand estimation is essential for retailers. Failure to accurately predict promotional demand can result in lost sales, stock-outs and, in the case of perishable items, in spoilage. Recent development in machine learning techniques allow retailers to use much richer data sets than traditional time series techniques. These data not only include detailed sales and promotion information but also granular product and store attributes, price changes, in-store display information and data from external sources. Using machine learning algorithms retailers can make use of those rich data and generate more accurate demand forecast even for complex promotional scenarios. This paper discusses a promotional demand forecasting project delivered for a large Polish retailer. We present the supported business process, discuss data sources as well as data cleansing and data augmentation steps. Next, we discuss explanatory features supporting the forecasting process and the machine learning algorithms used to generate forecasts. Finally, we discuss forecast postprocessing steps and compare the forecast accuracy with the existing forecasting system finding significant improvements for both perishable and non-perishable products.

## The equity markets of the BRICS and the world: raw material suppliers vs manufacturing economies

Presenter: Angi Roesch
Co-authors: Angi Roesch;Harald Schmidbauer;Yongda He;Erhan Uluceviz

In a network theoretic setup for equity markets, the importance of any market as a risk spreader can be monitored on a daily basis by a concept called risk propagation value, which can be interpreted as a measure of centrality under current shock spillover gradients. These propagation values provide a means of

monitoring the concentration of the power to spread risk and create volatility in the network, and of gauging the network's systemic risk. Within the BRICS group of equity markets, Brazil, Russia and South Africa are mainly raw material suppliers, while India and China are predominantly manufacturing economies. For example, we can show that the shock propagation power of raw material suppliers tends to be at a higher level when energy prices tend to go down, that is, when their business is not going well.

### Forecasting the Swiss Market Index using market entropy

Presenter: Harald Schmidbauer
Co-authors: Harald Schmidbauer;Erhan Uluceviz;Angi Roesch

The Swiss Blue Chip equity market is known to be very efficient. So, is it futile to devise efforts to forecast tomorrow's return on the Swiss Market Index (SMI)? Probably so, if the forecast aims at summary measures such as the expectation. However, we can use information from the network of constituent stocks of SMI, and we can use quantile regression to forecast tomorrow's return distribution in detail. This framework is similar to an approach called the mixture of distributions hypothesis (MDH), but our information structure is different. While the MDH essentially utilizes volume information to explain the nonnormality of returns, our approach consists of extracting market entropy in a network theoretic setup to forecast daily return distributions. We show that our approach can beat early efforts to forecast SMI returns in certain ways which are practically useful.

### Spillovers among Energy Commodities and the Russian Stock Market

Presenter: Marco Lorusso
Co-authors: Marco Lorusso;Michele Costola

We examine the connectedness in the energy commodities sector and the Russian stock market over the period 2005-2020 using the variance decomposition approach. Our analysis identifies the booms and busts in the correspondence of political and war episodes that are related to spillover effects in the Russian economy, as well as the energy commodities markets. Our findings show that the Russian Oil & Gas and Metals & Mining sectors are net shock contributors of crude oil and have the highest spillovers to other Russian sectors. Furthermore, we disentangle the sources of spillovers that originated from the financial and energy commodity markets and find that a positive change in the energy commodity volatility spillover is associated with an increase in Russian geopolitical uncertainty. Finally, we show that the spread of COVID-19 increases the stock market volatility spillover, whereas it lowers the energy commodity volatility spillover.

### Asymmetric New Keynesian Phillips Curve for Mexico, 2005Q1-2021Q4

Presenter: Eduardo Loría
Co-authors: Eduardo Loría;Raúl Tirado

The recovery from the corona crisis has raised inflation around the world above central bank targets. In Mexico, this has caused a restrictive monetary policy. We estimate three Augmented Asymmetric Accelerationist Phillips Curves with Labor Precariousness (2005Q1-2021Q4) with NLARDL (Shin et. al., 2014). With an in-sample forecast, we choose the model that minimizes the RMSE, the Theil inequality coefficient, and the MAE. We found a cointegration relation between the variables and that there are significant short and long-run asymmetric effects of the unemployment gap and the rate of critical labor conditions over inflation. The predominant short-run effect comes from increasing unemployment, but in the long term, the opposite occurs. Additionally, we establish that there is no econometric evidence that negative unemployment gaps raise inflation.

## Getting the ROC into Sync

Presenter: Kajal Lahiri
Co-authors: Kajal Lahiri;Liu Yang;Adrian Pagan

Judging the conformity of binary events in macroeconomics and finance has often been done with indices that measure synchronization. In recent years, the use of Receiver Operating Characteristic (ROC) curve has become popular for this task. This paper shows that the ROC and synchronization approaches are closely related, and each can be represented as a weighted average of correlation coefficients between a set of binary indicators and the target event over different thresholds. An advantage of such a representation is that inferences on the degree of conformity can be made robust to serial dependence in the underlying series in the standard framework of a linear regression model. Such serial correlation is common in macroeconomic and financial data. We illustrate our theory using yield spread as a predictor of recessions.

## Band-Pass Filtering with High-Dimensional Time Series

Presenter: Alessandro Giovannelli
Co-authors: Alessandro Giovannelli;Marco Lippi;Tommaso Proietti

We deal with the construction of a synthetic indicator of economic growth obtained by projecting a measure of aggregate economic activity, such as gross domestic product (GDP), onto high-frequency smooth principal components representative of the medium-to-long-run component of growth in a high-dimensional time series. The smooth principal components result by applying a suitable cross-sectional filter. The result is a monthly nowcast of the medium-to-long-run component of GDP growth. After discussing the theoretical properties of the indicator, we deal with the assessment of its reliability and predictive validity with reference to US data, and consider the estimation of the indicator in the presence of structural change induced by the COVID19 pandemic.

## Characterizing the temporal dependence structure between wind speed variables and wind energy generation via Copula Theory

Presenter: Tuany Esthefany Barcellos
Co-authors: Tuany Esthefany Barcellos;Reinaldo Castro;Fernando Luiz Cyrino;Marco Aurélio Sanfins

The increased interest in the development of renewable energy sources has grown exponentially over time. The use of wind energy eliminates unwanted waste that is harmful to health and the environment from other energy sources such as coal and nuclear power plants. This work aims to analyze the dependence relationship between the wind speed variables and the generation of wind energy, in order to understand the stochastic nature of both phenomena. As a methodological tool, the theory of copula was used, which consists of formulating multivariate distributions so that several dependency structures can be represented through the copula function. For the study, it is necessary to identify adequate multivariate distributions. The construction of the joint distribution by copulas helps in the modeling for simulation of scenarios related to wind energy generation based on the information obtained about variable wind speed, that is, the work is based on the analysis and modeling of the dependence between the speed data and generation on a monthly basis for an hourly database of a wind farm in the state of Bahia in the period 2017.

## An AI/ML/Statistical Framework for Improving Numerical Weather Prediction (NWP) and Renewable Energy Production Forecasting

Presenter: Thomas Sherman
Co-authors: Thomas Sherman;

Weather forecasting plays an important role in a variety of weather dependent industries. One such industry is renewable energy, where wind and solar energy production is directly impacted by local environmental and weather influences. Utilities, farm/plant operators, and energy traders all leverage

weather-derived time series data to forecast energy production and the optimal times to purchase/sell energy in spot markets. CRCL Solutions has developed a suite of time series correction algorithms, which ingests weather forecast data, correct it based on historical analysis, and translate the corrected weather data into a renewable energy production forecasts for major wind farms in the United States. This talk will present CRCL's methodologies broadly and how time series correction algorithms are used to improve weather and energy forecasting. Weather forecasts are derived from numerical weather predictions (NWP), which are physics-based models which solve the governing atmospheric equations through space-time. NWP models output data on a wealth of atmospheric variables and are optimized over continental scales. They require immense computational resources and are typically run by large federal organizations, e.g. NOAA. While the models do a remarkable job predicting weather, there are systematic biases because the model is optimized over so many variables and such a large spatial scale. For wind energy production, wind speed is the most important weather variable. CRCL has developed a suite of AI algorithms that analyzes historical wind speed NWP forecasts and identifies where it is likely wrong. This has enabled CRCL to improve upon NOAA's NWP predictions via a post-processing methodology. By leveraging public data, CRCL can improve forecasting error (mean-squared) by over 20%; these results come from a case studying considering a year of hourly data at 259 wind farm sites in the Texas' ERCOT energy market.

## A wagering mechanism for prediction markets

Presenter: Pierre Pinson
Co-authors: Pierre Pinson;Aitazaz Raja;Jalal Kazempour

We design a forecast elicitation market, a mechanism that allows expert forecasters to sell predictions for an uncertain event and buyers to obtain an aggregate of these predictions, which would be of higher quality than their own forecast. The ever-increasing collection of data makes it naturally become distributed in terms of ownership, also with diverse expertise of agents to extract high-quality predictive information from it. Thus, platforms for the pooling of forecasts based on distributed data and heterogeneous predictive skill give the opportunity to gain a collective value, for a prediction task. The key challenges in designing such platforms involve the proposal of payoff functions with theoretical guarantees, to compensate forecasters, and an efficient approach to aggregation of reported forecasts. With the aim to address these challenges, we propose a one-shot (history-free) marketplace based on a wagering mechanism where, buyers intending to improve their forecasts post a prediction task and sellers respond with their forecast reports and wagers. Then, the platform delivers an aggregated forecast to the buyer (pre-event) and allocates a payoff to the sellers (post-event) for their contribution. First, we show that, in our setup, the utilized forecast aggregation method, i.e., quantile aggregation produces higher quality forecasts compared to the widely used linear pooling methods. Then, we propose a payoff allocation that elicits truthful individual predictions of the sellers and rewards them jointly based on the relative quality of their forecasts and on the improvement in the buyer's utility. We show that the proposed payoff mechanism satisfies several desirable economic properties, including those specific to electronic platforms. To illustrate these properties, we provide several numerical examples. Finally, we present a real-world case study of wind energy forecasting, using data from the Global Energy Forecasting Competition 2014, to highlight the workings of the proposed market mechanism.

## Vector Quantized Autoregressive Probabilistic Time Series Forecasting

Presenter: Kashif Rasul
Co-authors: Kashif Rasul;

In practice, deep learning based Time Series Models come in many forms, but at a high level learn some continuous representation of the past and use it to output point or probabilistic forecasts. In this talk, we introduce a novel autoregressive architecture that instead learns a discrete set of representations that are used to predict the future distribution. We will highlight this model's performance as well as the shortcomings of this approach and hope that researchers can further investigate this inductive bias for forecasting.

## When MIDAS Meets LASSO: Forecasting Tail Risk Using Effective Macroeconomic Variables

Presenter: Xiaohan Xue
Co-authors: Xiaohan Xue;Yi Luo

A new framework for the joint estimation and forecasting of Value at Risk (VaR) and Expected Shortfall (ES) is proposed, which incorporates low-frequency macroeconomic and financial indicators into the quantile-based GJR-MIDAS model. Using an innovative machine-learning approach that maximizes the penalized Asymmetric Laplace (AL) likelihood function with an Adaptive-Lasso penalty, the most informative variables are selected in a ''big data" setting. A dynamic selection process enables the visualizing of the variable-selection evolution. In the empirical analysis, four variables (namely, realized volatility, implied volatility, term spread and housing starts) are consistently selected for most of the rolling windows. More information may be required to predict extreme VaR and ES. The out-of-sample backtesting results show that our method passes most backtests with relatively higher $p$-values and achieves the minimum loss in the joint forecasting of VaR and ES.

## Improving the forecasting accuracy of global models / cross-learning in large datasets by finding clusters of similar time series

Presenter: Pablo Montero-Manso
Co-authors: Pablo Montero-Manso;Angel Lopez-Oriona;Jose Vilar

Cross learning/global models is a recent paradigm in forecasting that pools together the data of a group of time series and fits a single model to them. Many successful applications of cross-learning rely on the time series in the group being 'similar'. Similarity is often judged by high-level criteria such as time series being of the 'same nature' (e.g. all come electricity consumption). Under the cross-learning paradigm we can think of two potential avenues for improving forecasting accuracy: 1) Exploring the space of models (e.g. new deep learning architectures, new data transformations) 2) Exploring methods that try to find better groups time series. This presentation focuses on the second avenue. We introduce an 'end-to-end' notion of time series similarity for cross-learning: time series are similar if they forecast accurately when pooled together under a given model class. We will show a selection of algorithms that leverage that notion of similarity to find groups of time series within a larger dataset, each of these groups is then forecasted global model. These clustering algorithms can be combined with most forecasting models (linear, deep networks, trees). We will give empirical results on real an simulated dataset. The main result: we can almost always improve forecasting accuracy of a given model, even if marginally.

## Forecasting 'Accuracy' vs 'Explainability' – Experience from Demand Forecasting in Supply Chain

Presenter: Nilmadhab Mandal
Co-authors: Nilmadhab Mandal;

Thanks to mainstreaming of Machine Learning (ML) in the last decade, demand forecasting methodology has evolved beyond the traditional statistical/econometric models. As a result, the potential for 'accuracy' improvement has received an encouraging boost. In the pursuit of better accuracy, an advanced ML model today has several complex features and algorithms as standard methodology. However, ML models have inadvertently introduced an element of opaqueness and have turned into more 'black-box'. Business decission makers, while benefiting from the delta accuracy, are now struggling with the 'explainability' of the forecast resulting into lower adoption. This paper will highlight the main methodological reasons for the explainability challenge, identify key concerns for operationalization of forecast and throw light on methods of addressing the challenges. The insight will be based on real-life experiences from multiple business implementations across consumer product companies.

## Superforecasting Revisited: The Human Forest Effect Dominates Differences in Predictive Skill and Expertise

Presenter: Regina Joseph
Co-authors: Regina Joseph;Pavel Atanasov

Faith in expertise as a primary criterion for predictive accuracy has been challenged by the evidence of superforecasters: generalists who exhibit forecasting accuracy that consistently outperforms that of domain experts in such broad areas as geopolitics. But does such predictive dominance extend to more technical domains like the life sciences? Are the cognitive traits of elite forecasters the key determinant in above-average accuracy or can elicitation modalities have a significant impact? Vetted elite forecasters are a limited resource; can better forecasting systems augment the predictive performance of those who are neither elite forecasters nor domain experts? We report the results of applying the Human Forest method of reference class forecasting to two seasons of a forecasting tournament devoted to predicting clinical trial success for new infectious disease vaccines and treatments. Human Forest acts as an "information scaffold," allowing forecasters to arrange comparison classes from ingested exogenous datasets, from which the system then generates a base rate to aid the user's prediction. We compare the individual accuracy of forecasters randomly assigned to either the Human Forest elicitation platform or a control poll condition. We also compare accuracy among three cohorts: life science professionals, volunteers with no life sciences background, and superforecasters. Using standardized Brier scores across 60 questions and two seasons, we show that individuals using the Human Forest method significantly outperformed those in the control condition. The descriptive difference was Cohen's d = 0.51, and remained above d > 0.35 across specifications. Neither superforecasters nor life science professionals outperformed in accuracy versus participants without backgrounds in forecasting or life sciences. Forecasters scoring high on numeracy measures, and those demonstrating more general science knowledge, tended to produce more accurate forecasts. Our results show that information design and elicitation user interfaces play a key role in the Supply Chain to predictive accuracy.

## 20 Years of FVA: A Critical Retrospective

Presenter: Michael Gilliland
Co-authors: Michael Gilliland;

In the 20 years since Forecast Value Added entered the lexicon, FVA has gained traction among practitioners as a way to identify wasted efforts in the forecasting process (i.e., process steps and participants that are failing to "add value" by improving forecast accuracy). Many organizations now use FVA to complement traditional forecast error metrics like MAPE. And many FVA users and researchers have advanced the basic concept with creative new ways to evaluate forecasting performance and drive process improvement (e.g., Morlidge & Goodwin's "Unavoidability Ratio" and de Kok's "Stochastic Value Added"). This presentation examines the initial motivation for an FVA-guided approach, highlights extensions of the FVA mindset, and addresses misconceptions about FVA's purpose and implementation. In addition, we will explore criticisms of the approach and problematic scenarios, while providing guidelines for the proper interpretation and use of FVA results.

## Alternative Measures to Evaluate Density Forecasting

Presenter: Roberto Morales-Arsenal
Co-authors: Roberto Morales-Arsenal;José Javier Núñez-Valázquez

A density forecasts shows all the uncertainty associated with the point forecast. To evaluate the quality of this estimation the standard technique computes the so-called probability integral transform and, in order to make comparisons between alternative density forecasts it is common to use scoring rules. In this work the standard statistical techniques are extended using decision theory techniques such as: 1) stochastic dominance, 2) expected Monetary value, 3) bayesian methodology and 4) presenting the evaluation process

of a density forecast as a problem of multiobjetive programming. The results indicate a certain coherence between the statistical techniques and the techniques coming from decision theory.

## Forecasts and Order Decisions: Reactions to Demand Variability

Presenter: Mustafa Sinan Gonul
Co-authors: Mustafa Sinan Gonul;Dilek Onkal;Ayse Kocabiyikoglu;Itir Gogus

In a typical supply-chain management setting, making order decisions inherently entails forecasting the uncertain demand for the relevant products. Through this translation of demand forecasting into final order decisions, one of the persistent findings in recent years is the pull-to-centre effect. This effect can be summarized as the tendency of the decision makers to set their order decisions between the mean demand and the normative order quantity. In the current study, we attempt to explore how decision makers will react to demand uncertainty, particularly to changes in variability of the demand and investigate the corresponding pull-to-centre effect. We also try to identify a potential cognitive bias, overprecision, that may prevail in this forecasting and decision process. Findings are discussed and directions for future research are suggested.

## Predictability of Temporally Aggregated Real Series

Presenter: Stephen Snudden
Co-authors: Stephen Snudden;Reinhard Ellwanger

Real macroeconomic variables are often aggregated from higher-frequency data. We show that this seemingly innocent feature has far-reaching consequences for the predictability of such series. First, the series are predictable by construction. Second, conventional forecast comparisons are uninformative about the effectiveness of the random walk forecast, making it necessary to re-evaluate the merit of existing forecasting approaches. Third, forecasting models should be estimated with end-of-period observations, even when the goal is to forecast the aggregated series. We highlight the relevance of these insights for forecasts of real prices of commodities and assets. Link to paper: https://www.dropbox.com/s/5y02xg5vxq69asu/BS_TestingForecasts.pdf

## Evaluating Forecasters in Real-Time by Finding Individuals who are as Wise as the Crowd

Presenter: Mark Himmelstein
Co-authors: Mark Himmelstein;David Budescu;Emily Ho

Research evaluating the talent of human forecasters has achieved a consensus: forecasting is a unique skill best measured by knowing how accurate individual forecasters have been in the past, on average. However, unlike the measurement of other traits, evaluating forecasting skill requires substantial time investment. Before their average accuracy can be estimated, forecasters must make a sufficiently large number of predictions about events that subsequently resolve. Our work builds upon research on peer similarity methods, such as proper proxy scoring rules and surrogate scoring rules, to show how peer similarity can substituted as a scoring criterion to identify talented forecasters from among a crowd in real-time, before any events resolve. Studies have even shown that asking forecasters to predict what their peers would forecast can yield gains in accuracy, while also serving as an incentive compatible mechanism for scoring rules based on peer similarity. We define a new peer similarity-based scoring rule: the Expected Brier Score (EBS). We test the utility of EBS as an evaluation metric in a unique forecasting experiment, in which forecasters made repeated predictions about a set of events at fixed time intervals. Because each forecaster forecasted all events at the same points in time, many of the confounds common to forecasting tournaments or observational data were eliminated. This allowed us to conduct straightforward analyses, which account for changes in information and accuracy as time progressed, to demonstrate the efficacy of our methods. EBS, which was available immediately after forecasters made their forecasts, was not only successful at

selecting the best forecasters, but was even more reliable than actual mean Brier scores, which only became available after the events resolved. Our results indicate that selecting even a single forecaster based on their past EBS can yield subsequent forecasts that approximate the actual accuracy of much larger crowd aggregates. We also found that EBS along with a measure of probabilistic coherence dominate several other known predictors of forecasting talent.

## Factors affecting consumers' inflation expectations: Purchase frequency and method of elicitation

Presenter: Xiaoxiao Niu
Co-authors: Xiaoxiao Niu;Nigel Harvey

National surveys have been used to measure consumers' inflation expectations because they help to inform the monetary policy decisions of central banks. However, survey responses have shown overestimation and considerable disagreement between respondents. Relatively little is known about how inflation expectations are generated. This study examined three factors that previous research suggests could be important in this respect: people's financial literacy, their purchasing experience, and the method used for eliciting responses. We found that only the last two of these influenced inflation judgments: 1) Consistent with previous reports of a frequency bias, higher price change estimates occurred in product categories in which the frequency of consumption was higher, 2) Estimates derived indirectly by eliciting prices in product categories in successive years were much higher than estimates of inflation produced directly by intuition. Our results reinforce some previous findings and have implications for improving the accuracy of inflation expectations derived from surveys.

## A recommender system for forecast adjustments

Presenter: Anna Sroginis
Co-authors: Anna Sroginis;Nikolaos Kourentzes;Robert Fildes

Expert judgment is the primary tool to incorporate additional knowledge when a forecasting model is unable to account for such information. Empirical studies show that using human judgment can be beneficial in adjusting forecasts, even though it is prone to many cognitive biases and inefficiencies. In practice there can be several types of adjustments: (i) adjusting individual forecasts for a single item; (ii) batch-adjusting, i.e., correcting several time-series or categories at the same time; (iii) model tuning, such as indicating a location of corrections rather than the size and letting a statistical model estimate them, e.g., by introducing indicator variables. We observe that all types of human judgment perform inconsistently and, at times, are very damaging. This may be due to adjusting based on misinterpreted contextual cues or using the wrong type (and size) of adjustment. Past literature has investigated restricting or guiding forecasters through the refinement of the forecasting process or the system interface, with mixed results. Other contributions have explored the effectiveness of post-processing adjustments via models, although their practical applicability can be questioned. Nonetheless, these only consider adjustments of type (i). We develop a model to assess the forecast value added for different types of adjustments. The aim is to provide real-time guidance to users while they are producing their forecasts. In this work, we investigate the feasibility of such models, using a UK retailer as a case study.

## Copula-based combination of point prediction systems to calibrated probabilistic forecasts

Presenter: Jonas Rieger
Co-authors: Jonas Rieger;Oliver Grothe

We formulate a copula-based algorithm for combining multiple point forecasts to a calibrated density forecast, the copula-combined density (CCD) method. The combination proceeds by a copula times series model of the point forecasts' errors and, thus, explicitly accounts for marginal serial and forecast errors' joint

dependence structure. Consequently, the model is designed to perform particularly well for complicated, non-linear dependencies and non-Gaussian margins. Non-linear dependencies and non-Gaussian margins frequently arise in applications, as they can stem from various reasons, e.g., from latent, shared influence factors of models or sensors, bounded marginal supports, or heavy tails. We show theoretically that the algorithm can produce calibrated forecasts for any forecast errors. In the case of multivariate Gaussian forecast errors, the presented algorithm embeds the results of the point forecast combination by Bates and Granger (1969). We compare the algorithm to other forecast combination methods in an extensive simulation study and provide a worked-through example on temperature forecast data for the Zugspitze, the highest mountain in Germany. Especially for non-Gaussian dependence structure or margins, the CCD method outperforms its competitors in the simulations, and, thus, the above advantages of the CCD method are substantiated.

## A mixed-frequency combination approach to forecast covariance matrices of asset returns

Presenter: Malvina Marchese
Co-authors: Malvina Marchese;Richard Payne;MIchael Tamvakis

To improve on the forecasting accuracy of conditional covariance matrices of asset returns, we propose a novel forecast combination approach based on mixed information from high and low frequency data. The combination strategy relies on an economic loss function based on portfolio selection. Specifically, it identifies the mixing weights using portfolio diversification optimality criteria. Our approach does not require a proxy for the latent conditional covariance matrix and facilitates the economic interpretation of the combination strategy for decisioning maker. Two empirical applications involving energy futures and S&P100ETs show that the proposed combination strategy leads to minimum variance portfolios with lower risk on an out-of-sample basis with respect to a number of alternative specifications based on pure statistical loss functions. The results suggest that low-frequency data improve volatility forecasting even when high frequency data is available at medium and long horizons.

## On Volatility Impact of Russian Invasion War on the Crude Oil Market

Presenter: Oyebimpe Adeniji
Co-authors: Oyebimpe Adeniji;ISAAC AJAO;BODE AJETUNMOBI

An empirical study was carried out on volatility impact on crude oil prices before, now and its forecast during Russia 2022 Invasion War of Ukraine. The Impact of Russian Invasion war on crude products is observed and reported in this paper. Daily crude oil data was used in this study. The time plots indicate non-stationary and high volatility persistence in the series, the descriptive statistics also revealed evidence of stylized facts of volatility features. Traditional GARCH models and extensions of GARCH models were used to capture the volatility of the series. it is observed that the traditional GARCH models with normal innovations have low forecast performance but are still suitable when the error term is tampered. This study is therefore comparing tampered innovations of GARCH(p,q) models , extension of GARCH models such as Generalized Autoregressive Score (GAS) for improved forecast estimates. The two error innovations considered were the Generalised Length Biased Scaled-t (GLBST) and Generalised Beta Skewed-t (GBST) distributions, obtained by remodifying Fisher Concept of Weighted Distribution and McDonald Generalised Beta Function, respectively, in the Student-t distribution.. The proposed innovations were imposed on GARCH(p,q) models. Data from West Texas Intermediate Crude Oil daily data were used to illustrate the models. The proposed models were compared with the jumps GARCH model. The performance of the proposed models over the existing ones was investigated using the Log-likelihood function, Root Mean Square Error (RMSE), Adjusted Mean Absolute Percentage Error (AMAPE) and Akaike Information Criterion (AIC). The ongoing war has a great and mixed Impact on the crude oil market and its end products, also the jumps GARCH models captured the information more properly than the traditional models. The suitable models were selected using Selection Criteria, forecast values at 20 days ahead were also predicted. Based on empirical data, the report indicates positive growth in crude oil prices but the

impact spikes negative pressure on Nigeria crude products and more hardship on the economy.

## Volatility Forecast and Risk Measure Calculation for Modified Asymmetric GARCH Model with Markov Switching

Presenter: Nurhayati Nurhayati
Co-authors: Nurhayati Nurhayati;Utriweni Mukhaiyar;Khreshna Syuhada

Volatility forecast is very important in finance and insurance, particularly in quantitative risk management. In this paper, we aim at modifying the well-known GARCH model such that it is able to accommodate the asymmetric volatility property and the changing of volatility level or state. The proposed model is called Modified Markov Switching (MMS)-GARCH. First, we derive theoretical as well as empirical properties of such a model including kurtosis, autocorrelation function, asymmetric effect, and volatility clustering. We then do a volatility forecast and provide a calculation of future observations through risk measures of Value-at-Risk (VaR) and Expected Shortfall (ES).

## Adaptive commodity price models for Inflation Forecasting

Presenter: Nicos Pavlidis
Co-authors: Nicos Pavlidis;Alisa Yusupova;Efthymios Pavlidis

The wide fluctuations of many commodity prices, especially food and energy prices, over the last two decades have attracted the attention of academics and policy makers and have renewed their interest in assessing their impact on overall price stability. It has been well established in the literature that commodity prices could play a crucial role in inflation dynamics through first-round effects on consumers and producers, and second-round effects related to expectations that may impact negatively on consumption and investment. In this paper, we examine the predictive content of commodity prices to forecast inflation by allowing for both model uncertainty and time variation in the data generating process. A growing empirical literature provides strong evidence in favour of structural instability in numerous application areas, including macroeconomics and finance. If unaccounted for structural instability can have detrimental consequences for inference and forecasting. To investigate the predictive content of commodity prices in the presence of structural instability we propose in this work an approach to address that relies on dynamic linear models with discounting. Our approach can accommodate both gradual and abrupt change. We also consider state-of-the-art point and density model combination methods to account for time-variation in model specification, and (or) the optimal linear model combination.

## Persistence in Economic Networks

Presenter: Jozef Barunik
Co-authors: Jozef Barunik;Michael Ellington

This paper studies heterogeneous network structures driven by different degrees of persistence in economic connections. Using frequency domain techniques, we introduce measures that characterize network connections stemming from transitory and persistent components of shocks. Our approach permits testing for statistical differences in such connections that evolve over time. We estimate uncertainty networks from the main US sectors and argue that they track transitory and persistent sources of systemic risk. Hence they may serve as a useful tool for macro-prudential supervisors and investors alike.

## Forecasting Economic Supply Chain Risk Capital

Presenter: Kamil Mizgier
Co-authors: Kamil Mizgier;

Supply Chain networks are complex and interconnected, with Supply Chain visibility often limited

to the first-tier vendors and a reactive approach to risk management. To address this challenge, first, a conditional mean network analysis (Diebold and Yilmaz, 2014) could reveal interesting information about the connectedness of complex Supply Chains. Second, risk factors at both macro level (e.g., catastrophic, sociopolitical, cultural, war, energy and terrorism) and micro level (i.e., failures specific to a supplier and/or its supply network) need to be identified. Third, a risk index associated with nodes and arcs in the Supply Chain network should be developed. A set of risk index measures was proposed by Mizgier et al. (2013) and further expanded by Mizgier et al. (2015) and Mizgier (2017). The authors proposed to use Value at Risk to measure the risk of disruptions in a complex Supply Chain network. It can then be translated to the concept of Economic Supply Chain Risk Capital (ESCRC) which measures the amount of capital that a firm needs to cover Supply Chain risk losses materializing at some future date with a certain probability. In other words, ESCRC converts Supply Chain risk to an amount of capital that is needed to support it. Finally, by using more sophisticated forecasting techniques based on big data, firms can identify Supply Chain bottlenecks and manage risk in a proactive manner.

## Assessing GDP forecasts from autoregressive models: the impact of model complexity and training dataset

Presenter: M. Paula Bonel
Co-authors: María Paula Bonel;

We evaluate different strategies to forecast GDP through autoregressive models. The analysis covers 46 emerging and advanced economies. Two key dimensions are explored. We consider alternative training datasets: single country vs. panel of countries. Simultaneously, we evaluate models that differ in terms of complexity. The results indicate that expanding the training dataset from single country to a panel of countries results in significant improvements in performance. In addition, we verify a strong complementarity between dataset size and model complexity. If individual country data is used for training, more complex models do not improve performance. In contrast, when a panel of countries is used for training, higher model complexity results in more accurate forecasts. The performance of these model forecasts is similar to professional forecasts. The results are verified for different forecast horizons and for emerging and advanced economies. Our results can prove valuable for forecasting exercises in other contexts and in the case of alternative forecast models.

## Benchmarking Historical Consistent Neural Networks

Presenter: Nico Beck
Co-authors: Nico Beck;Julia Schemm;Benedikt Sonnleitner;Claudia Ehrig

Historical Neural Networks are a Neural Network class, specifically designed for macroeconomic forecasting. Hans Georg Zimmermann presented those at several workshops at the ISF. Although they have achieved remarkable results in industrial practice (e.g., copper price forecasting), a rigorous benchmark study is currently missing. To address this research gap, we evaluate under which conditions (i.e., forecast horizon, data properties) the model is superior to state of the art methods. We use real world and synthetic data sets.

## An Approach to Forecasting Good Growth: Georgia in the Crosshairs

Presenter: Edward Raupp
Co-authors: Edward Raupp;

This study aims to develop a new composite index of indicators as an alternative or supplement to the most commonly used measure of a nation's economy, Gross Domestic Product (GDP), and its correlate GDP per capita. It is said, paraphrasing Winston Churchill, that GDP is the worst measure of the economy in the world—except for all the others. This study shows that the growth of GDP, while it tracks the sum of all expenditures and incomes, may be fatally flawed as an indicator of good growth. The composite index

(G) of good growth developed in this study with over 70 components may offer a more useful measure of the entire economy, including its sustainability and its balance between efficiency and equity. The unit of study is the post-Soviet nation of Georgia, a small country with a population of 3.7 million. Georgia is threatened by a virulently aggressive expansionist neighbor, the Russian Federation, which invaded the country in 2008. Nearly 70 percent of Georgian educators surveyed believe that another attack is somewhat likely or highly likely in the next five to ten years. Forecasting any Georgian macroeconomic variable, therefore, must be done under conditions of a high degree of uncertainty, so the study uses a mix of methods and attempts to use an innovative construction to forecast good growth in the coming decade. Using the historical values of G, the study proposes an approach to forecasting the prospects for good growth in Georgia from 2023 to 2033.

## Calibration window selection based on change-point detection for forecasting electricity prices

Presenter: Weronika Nitka
Co-authors: Weronika Nitka;Julia Nasiadka;Rafał Weron

We employ a recently proposed change-point detection algorithm, the Narrowest-Over-Threshold (NOT) method, to select subperiods of past observations that are similar to the currently recorded values. Then, contrarily to the traditional time series approach in which the most recent observations are taken as the calibration sample, we estimate autoregressive models only for data in these subperiods. We illustrate our approach using a challenging dataset - day-ahead electricity prices in the German EPEX SPOT market - and observe a significant improvement in forecasting accuracy compared to commonly used approaches, including the Autoregressive Hybrid Nearest Neighbors (ARHNN) method and autoregressive models estimated on a continuous calibration sample.

## Smoothing Quantile Regression Averaging for probabilistic electricity price forecasting

Presenter: Bartosz Uniejewski
Co-authors: Bartosz Uniejewski;Rafał Weron

Probabilistic Electricity price forecasting} (EPF) for a couple of years is gaining popularity and energy analysts have become aware of its importance in energy systems planning and operations. A variety of approaches have been considered, including bootstrapping, Bayesian statistics, deep learning, or Quantile Regression Averaging (QRA). The latter one is one of the most successful methods to obtain probabilistic forecasts. However, recent studies have revealed the method's deficiency. Firstly, the quality of forecasts significantly decreases when the set of regressors is larger than just a few. Secondly, the interval prediction obtained with QRA is too narrow, which makes the forecast of electricity price, especially for early hours, less reliable. Uniejewski and Weron (2021) depicted that the nominal coverage is never reached or exceeded while using QRA. Here, we introduce Smoothing Quantile Regression Averaging (SQRA), a pioneering approach that has managed to deal with the problem of too narrow interval prediction. We evaluate the proposed solution using datasets from four energy markets and compare it with two popular classes of forecasting models. The results indicate that the SQRA outperforms the standard QRA in both reliability and sharpness. Finally, we proposed a speculative trading strategy based on probabilistic forecasts that allow the market participants to obtain profits by trading solely on the day-ahead market.

## Electricity Price Forecasting: Main trends and models as of 2022

Presenter: Rafał Weron
Co-authors: Rafał Weron;

Until the early 2010s, the Electricity Price Forecasting (EPF) literature was dominated by relatively parsimonious linear regression and neural network models. As more data and computational power became

available, the models became more complex to the extent that expert knowledge was no longer enough to handle them. This paved the way for machine learning (ML) in EPF. Here, we provide an overview of the main trends and models as of 2022, with a particular focus on ML.

## Multi-horizon wind power forecasting using multi-modal spatiotemporal Neural Networks

Presenter: Eric Stefan Miele
Co-authors: Eric Stefan Miele;Nicole Ludwig;Alessandro Corsini

Wind power is one of the leading renewable energy sectors and is considered instrumental in the ongoing decarbonization process. Accurate forecasts are essential for a reliable large-scale wind power integration, allowing efficient operation and maintenance (O&M), planning of unit commitment, and scheduling by system operators. However, due to non-stationarity, randomness, and intermittency, forecasting wind power is challenging. This work investigates a multi-modal approach for wind power forecasting by considering turbine-level time series collected from SCADA systems, reanalysis maps, and high-resolution Numerical Weather Prediction (NWP) maps. A neural architecture based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) is proposed to process and combine different data sources containing spatiotemporal patterns. This architecture allows combining the local information from the turbine's internal operating conditions with past (reanalysis) and future (NWP) meteorological data from the surrounding area. Specifically, this work focuses on multi-horizon turbine-level hourly forecasts for an entire wind farm with a lead time of 4 days. Reanalysis and NWP maps are regular square grids on a mesoscale level centered around the wind farm. The choice of including a larger area and not only the meteorological data associated with the specific farm location was motivated by the presence of patterns that evolve both in time and space. In fact, by considering the XGBoost regressor as a baseline, the forecasting error decreases by considering larger regions surrounding the farm. The proposed neural architecture produces better forecasts by including larger maps and outperforms the XGBoost regressor baseline, achieving a skill score of 25%. Finally, an ablation study is carried out to understand each data source's impact on the neural network performance.

## Learning the Distribution of Economic Variables

Presenter: Luboš Hanus
Co-authors: Luboš Hanus;Jozef Baruník

We propose a distributional deep learning approach to Probabilistic Forecasting of economic time series. Being able to learn complex patterns from a large amount of data, deep learning methods are useful for a decision making that depends on uncertainty of large number of economic outcomes. Such predictions are also informative to decision makers facing asymmetric dependence of their loss on outcomes from possibly non-Gaussian and non-linear variables. We show the usefulness of the approach in the two distinct problems where machine learns the pattern from data. First, we construct big data-driven macroeconomic fan charts from a machine learning approximated model. Second, we illustrate gains in the prediction of stock return distributions which are heavy tailed and suffer from a low signal-to-noise ratio.

## Weight Initialization for Neural Network diversity - an empirical evaluation

Presenter: Sven F. Crone
Co-authors: Sven F. Crone;

Artificial Neural Networks require the random initialisation of parameter weights to break symmetrie and facilitate adjustment of weights through learning by gradient descent, and thus "learning". A variety of weight initialisation methods have been proposed in literature, from simple uniform random sampling across different (small) ranges of weight values (see, e.g. Sarle, 1990) to more intuitive uniform distributions (Hill et al, 2000), to sophisticated approaches by Nguyen-Widrow, and LeCunn. However, the majority of

initialization methodologies developed in the 1990s were aimed at the reduction of training time and/or the convergence of a single network topology towards a suitable minimum, but largely ignored issues of out-of-sample accuracy or diversity in ensembles. To date, no studies were conducted dedicated on how to set initial weights to (a) create improved out-of-sample predictions of single network topologies, or (b) create diverse Neural Networks destined for ensembles, which require diversity to achieve robust ensembles which increase accuracy and reduce variance of predictors. Our experiments on a large real-world dataset employing multiple step-ahead errors show that the setting multiple initializations wisely will impact accuracy, robustness of accuracy, and runtime as well, determining the operational efficacy of the models in use.

## How to be a Forecasting Superhero!

Presenter: Ruth Abbott
Co-authors: Ruth Abbott;

The role of a Business Forecaster can be challenging, so what tools do you need in your toolkit to ensure you can be a Forecasting Superhero? Let's explore together four crucial elements required for success, based on lessons learnt from the field over the past 2 years. This interactive presentation will cover the importance of:

1. Accurate forecasts using advanced models
2. A collaborative planning platform to share and discuss forecasts
3. Proactive insight from causal analysis to help your colleagues make proactive, informed decisions
4. The ability to quickly run and compare forecasting scenarios that are joined up to the rest of the organisation

## Fathoming Forecasting practices in the Pharmaceutical Sector

Presenter: Konstantinos Nikolopoulos
Co-authors: Konstantinos Nikolopoulos;Vasileios Bougioukos;Samantha Buxton;Konstantia Litsiou;Bethany Kersse;

We investigate organisational practices and processes, as well as the forecasting function itself, into (initially) UK pharmaceutical SMEs and a network of GPs, and at a second stage at SMEs and MNEs across the globe. We follow a mixed methods approach combining qualitative analysis of rich semi-structured interviews and quantitative analysis of surveys as well as meta-analysis of publicly available secondary data. We provide new theoretical development via a grounded theory approach.

## Time Series Forecasting on Coronavirus Data in Turkey using Combination of Machine Learning Algorithms and ATA Method

Presenter: Tugce Ekiz Yilmaz
Co-authors: Tugce Ekiz Yilmaz;Guckan YAPAR

The coronavirus has deeply affected the whole world as we know. The whole world has struggled to cope with this pandemic for two years. Many studies have been done on this subject in order to help the government and policymakers to cope with the pandemic. At this point, it is very important to obtain more accurate forecast. Although there are many methods that forecast is obtained, we have seen many times before that the combination technique gives the best result. So, at this study we have used combination methods and also only statistical and only machine learning methods for comparing. We have seen that ATA method as an alternative of simple exponential smoothing method has given more accurate results comparing its counterpart. Moreover, combination methods that compound ATA method and machine learning methods such as artificial neural network (ANN) have given a considerably good result. Thus, we can say that ATA method is innovative forecasting technique in the literature.

## Nonlinear smooth transition regression models in electricity price forecasting

Presenter: Andrzej Puć
Co-authors: Andrzej Puć;Katarzyna Maciejowska

Most statistical methods of electricity price forecasting rely on linear regressions. Here, we apply nonlinear smooth transition regression (STR) to predict the behaviour of electricity markets. We specify the STR model for intraday and day-ahead electricity prices in the German market. The resulting short-term point forecasts are compared with outcomes of an ARX benchmark, commonly used in EPF literature. Next, we apply three different variance stabilizing transformations: asinh, t-PIT and N-PIT, which reduce the impact of the outliers and help in describing nonlinear relationships. Finally, we address the issue of tuning the boundaries for STR model parameters by applying a state-of-the-art hyperparameter optimization framework. Such solutions are already commonly used by the machine learning community, but can also successfully automatize statistical models. The main conclusions are: (i) there is firm evidence that the electricity prices are generated by a nonlinear process, (ii) the STR model, which directly describes the nonlinearity, outperforms its ARX counterparts (iii) the estimation process of the STR model can be significantly improved by optimizing model hyperparameters.

## Impact of decomposition on time series bagging forecasting performance

Presenter: Anyu Liu
Co-authors: Anyu Liu;Xinyang Liu;Gang Li;Li Chen;

Time series bagging has been deemed an effective way to improve unstable modelling procedures and subsequent forecasting accuracy. However, the literature has paid little attention to decomposition in time series bagging. This study investigates the impacts of various decomposition methods on bagging forecasting performance. Eight popular decomposition approaches are incorporated into the time series bagging procedure to improve unstable modelling procedures, and the resulting bagging methods' forecasting performance is evaluated. Using the world's top 20 inbound destinations as an empirical case, this study generates one- to eight-step-ahead tourism forecasts and compares them against benchmarks, including non-bagged and seasonal naïve models. For short-term forecasts, bagging constructed via seasonal extraction in autoregressive integrated moving average time series decomposition outperforms other methods. An autocorrelation test shows that efficient decomposition reduces variance in bagging forecasts.

## New 2-parameter families of advanced forecasting functions: non-seasonal and seasonal models, and comparison to ARIMA and exponential smoothing models

Presenter: Nabil Kahouadji
Co-authors: Nabil Kahouadji;

We introduce twenty-four new two-parameter families of advanced time series forecasting functions, using three forecast estimate methods along with eight optimization criteria. Non-seasonal and seasonal Time Series Models are derived and applied to time series in education, sales, economics, industry and finance. We also compare the performance of our twenty-four functions/models to both exponential smoothing and ARIMA models using non-seasonal and seasonal time series. We show in particular that a decomposition of a seasonal time series into a trend, seasonal and random components is not needed for our models. Finally, we compare these twenty-four models using 5-year stock market data of 467 companies of the S&P 500 companies.

## Perspectives on Diversity in Forecast Aggregation and the Wisdom of Crowds

Presenter: Victor Richmond Jose
Co-authors: Victor Richmond Jose;David Budescu;Yael Grushka-Cockayne;Jack Soll

Across disciplines, scholars have observed that diversity is an essential ingredient to the wisdom of

crowds. Yet research paradigms vary in how they define and operationalize diversity. In this talk, we distinguish between two dimensions of diversity that have been the focus of different literatures — knowledge-based diversity (differences among people) and statistical diversity (differences among judgments). Although statistical diversity is generally recognized as the proximate cause of crowd wisdom, many social scientists have focused on the differences in knowledge that can potentially produce it. We consider both direct and indirect sources of knowledge-based diversity. Direct sources are those that most immediately cause judgment, such as the information people hold and how they combine it. In contrast, indirect knowledge-based diversity refers to social and dispositional factors that lead to people to encounter or use information differently in the first place. An example would be having different life experiences that lead people to interpret the meaning of a given piece of information differently. We explore how these different dimensions of diversity interrelate, with an emphasis on implications for the wisdom of crowds.

## Effects of multiple adjustments in Supply Chain forecasting on forecast accuracy

Presenter: Niles Perera
Co-authors: Niles Perera;Banusha Aruchunarasa;Dilek Onkal

Forecasting is considered the lifeblood of Supply Chains (SC). Accurate forecasts are significant to Supply Chain management and efficient organizational planning. Decisions affecting upstream operations are often triggered by forecasts generated through information systems in Supply Chains. Forecasting professionals frequently modify system-generated forecasts for various reasons, leading to multiple adjustments over the life cycle of a forecast. Surveys suggest that such nested or multiple adjustments to system-generated forecasts are a common phenomenon in Supply Chains and they impact forecast accuracy. Despite this, multiple adjustments to forecasts remain a sparsely addressed research gap in the literature, with very few studies investigating them. Thus, to examine the potential effects of multiple adjustments to forecasts and their repercussions for enhancing/deteriorating forecast accuracy in the SC, a behavioural experiment was employed. The experiment involved four treatments to gauge the behaviours with respect to multiple adjustments to forecasts. 194 undergraduate and MBA students were recruited as participants for the experiment. In the Control Group, participants had to deal with forecasts containing initial adjustments while other treatments entailed subsequent adjustments with different levels of information provision. It was found that multiple adjustments to forecasts can significantly improve forecast accuracy with the increase of relevant information as well as industry exposure. The findings provide insights on judgmental adjustments to already-adjusted forecasts and suggest that the provision of relevant information related to the previous adjustment allows forecasters to perform better. The results underscore the importance of industry exposure and understanding the practical contexts for a forecaster to improve his/her decision-making regarding judgmental adjustments. This study stresses that Supply Chain management-related degree programs ought to provide industry exposure to students to develop an appreciation of practical implications of forecasting on other Supply Chain issues. These results have implications for businesses on their efforts to increase information visibility among Supply Chain actors with the aim of improving forecast accuracy cascading into optimal results in the Supply Chain. Further, system developers are encouraged to integrate these findings when developing or improving the forecast support system.

## Newsvendor Problems: Naive Judgemental Adjustments Can Help

Presenter: Congzheng Liu
Co-authors: Congzheng Liu;Ivan Svetunkov;Adam Letchford;Ivan Svetunkov

Newsvendor problems are an important and much-studied family of stochastic inventory control problems. One strand of the literature on Newsvendor problems is concerned with the fact that practitioners often make judgemental adjustments to the theoretically "optimal" order quantities. Although judgemental adjustment is not always a bad thing, two specific kinds of adjustment are normally considered to be particularly naive: demand chasing and pull-to-centre. Using a simulation study, we show that even such naive adjustments can be useful in practice. We discuss the conditions under which they are beneficial. Finally, we conduct a real-life experiment to see the impact of adjustments on inventory and costs. We find

that by using judgemental adjustment, decision makers can achieve lower costs and a service level closer to the target one.

## The aggregation of expert judgment for forecasting uncertain quantities

Presenter: Fergus Bolger
Co-authors: Fergus Bolger;

Quantities must often be forecast when good quality data for either statistical or machine learning approaches is lacking, for example, the economic effects of a new technology. In such cases we may need to supplement – or substitute entirely – 'hard' data with expert judgment. In expert knowledge elicitation (EKE) the judgments of several experts are typically collected on the assumption that the aggregate of these judgments will modally be more valid than the judgment of any randomly selected expert. However, there is disagreement about how best to aggregate judgments, for example, with a weighted or unweighted average. In the former case, relative expertise must be established somehow. Advocates of the 'Classical' EKE method propose that this is done by means of a pre-test requiring judgments of 'seed' variables from the same domain of expertise as the target of interest. Crucially, the values of seeds are known to elicitors, but not (explicitly) by the experts.

There is a growing body of evidence that such 'performance weighting' can improve the validity of judgments relative to equal weighting, however, the positive effects are hit-or-miss. I argue that to benefit fully from this approach there is need to better understand how experts make their judgments of uncertain quantities across different task environments, to guide seed selection and weighting. To this end, I present a theory of expert judgment of uncertain quantities based on Brunswik's 'Lens Model' – judgment tasks and expertise are described as linear models of relationships between cues and the criterion to be estimated. Uncertainty in the environment is captured by the predictive ability of the task model, whereas substantive expertise is instantiated as the degree to which experts have acquired the ideal task model, and normative expertise by their ability to report the predictive power of their judgment models. Using this theory, I create expert agents of varying substantive and normative abilities, then place them in task environments with differing levels of uncertainty, and other structural properties (e.g., the degree of interdependence between cues). I then use such simulations to address questions regarding the aggregation of expert judgment.

## Analysis and Forecast of the Global Maize Price: A Comprehensive Integration of Machine Learning and Econometric Approaches

Presenter: Rotem Zelingher
Co-authors: Rotem Zelingher;

This study is the first to supply a comprehensive medium-term analysis and forecasting of the global price of maize while integrating the combined effects of over 100 possible variables into seven econometric and machine learning (ML) methods. First, it uses different cross-validation techniques to avoid any forced pre-research assumptions and capture these complex relations realistically. The learning process begins with comprehensive stationary and causality tests, which detect the nature of each possible variable and its suitability to serve as an explanatory factor for the change in the maize price. Second, we perform a retrospect analysis, which considers a large number of variables from three different groups: market fundamentals, financial and climatic. Third, using relative importance, we reduce the number of features to include only those most relevant for the performance of accurate maize price forecasting. The process is employed with three ML algorithms in a monthly application. The results point out significant differences in the ML models' forecasting ability and in the type of variables to consider when analysing the price of maize, depending on the month and in cases of severe shocks in certain factors. Our tool can be easily trained from publicly available data and is made available open-source. It is thus easily adaptable to analyse and forecast the price of other commodities with a minimal budget and may help promote global food security.

## On the use of mean square error and directional forecast accuracy for model selection: A Monte Carlo investigation

Presenter: Robert Kunst
Co-authors: Robert Kunst;Mauro Costantini

We propose a new procedure for model selection based on simultaneously targeting the mean square error and directional forecast accuracy criteria. The procedure combines the two types of accuracy measures using a weighting scheme for the selection of the forecasting models. Monte Carlo analysis under different scenarios serves as a tool that assesses the strength of the procedure. To this end, we consider various Time Series Models as generation mechanisms, in particular time-homogeneous univariate and vector autoregressions but also generating laws that involve thresholds and structural breaks. We evaluate two aspects of forecast model specification: forecast model selection chooses one out of two rival models that are both evaluated over a training sample, whereas forecast model combination determines weights on an average of the two rival models from the training performance. For the evaluation of the training samples, we study rolling and expanding forecast windows and a combination of the two. Although results are quite heterogeneous across designs, we generally find that finding reliable tools for improving directional accuracy is difficult and that a price must be paid by deteriorating performance such as measured by the mean square error.

## On the importance of model selection methods in time series forecasting

Presenter: Mahesh Joshi
Co-authors: Mahesh Joshi;

Forecasters have a plethora of time series forecasting methods at their disposal, from classical statistical techniques to machine-learning-based techniques to combinations of both. Automatically choosing the best model becomes an important task when you are dealing with routine forecasting tasks that involve thousands or millions of time series, such as in retail applications. Model selection can occur at different levels. For model families that use exogenous variables, you need to select an optimal subset of these variables to include in the models. For model families that have several tuning parameters, you must select an optimal set of parameters. To limit the combinatorial explosion, it might also help to first choose the candidate model families you want to explore by examining salient features of the time series. Novel techniques that use machine learning or optimization are being researched. This talk provides a survey of some promising model selection techniques and presents results.

## Forecast selection and representativeness

Presenter: Fotios Petropoulos
Co-authors: Fotios Petropoulos;Enno Siemsen

Effective approaches to forecast model selection are crucial to improve forecast accuracy and to facilitate the use of forecasts for decision-making processes. Information criteria or cross-validation are common approaches of forecast model selection. Both methods compare forecasts with the respective actual realizations. However, no existing selection method assesses out-of-sample forecasts before the actual values become available - a technique used in human judgment in this context. Research in judgmental model selection emphasizes that human judgment can be superior to statistical selection procedures in evaluating the quality of forecasting models. We therefore propose a new way of statistical model selection based on these insights from human judgment. Our approach relies on an asynchronous comparison of forecasts and actual values, allowing for an ex-ante evaluation of forecasts via representativeness. We tested this criterion on numerous time series. Results from our analyses provide evidence that forecast performance can be improved when models are selected based on their representativeness.

## When is the Next Order? Forecasting the Timing of Retail Orders Using Point-of-Sales Data and Channel Inventory Estimations

Presenter: Tim Schlaich
Co-authors: Tim Schlaich;Kai Hoberg

Slow-moving goods are common in many retail settings and occupy a vast part of food retail shelves. Since stores sell these products irregularly and in small quantities, the replenishing distribution center may only place batched orders with food manufacturers every few weeks. In addition to order quantity, manufacturers facing such intermittent demand are required to forecast the order timing. In this paper, we explore the value of Point-of-Sales data to improve a manufacturer's order timing forecast for slow-moving goods. We propose an inventory modeling approach that uses the last order, Point-of-Sales data from retail stores, and a production lead-time forecast to estimate the retailer's channel inventory. With this estimate, we can approximate the inventory reach and anticipate the arrival of a new order. To illustrate our methodology, we first conduct a discrete-event simulation and numerical experiment and compare our results to established intermittent demand forecasting methods. Next, we validate our approach with empirical data from a small German food manufacturer that serves a grocery retailer with a central distribution center and 50 retail stores. We find that our approach can improve the accuracy of order-timing predictions by up to 43 %. We overcome a shrinkage-induced bias by incorporating a wastage correction factor. Our approach describes a new way of utilizing PoS data in multi-layered distribution networks and can complement quantity-based estimates of established forecasting methods such as Croston. Particular application opportunities arise when historic order data is scarce (e.g., product launch) or unreliable to predict future demand (e.g., external shocks like COVID-19).

## Learning customer and context specific response to price and availability for dynamic pricing and forecasting in attended home delivery

Presenter: Ozden Gur Ali
Co-authors: Ozden Gur Ali;

As customers are increasingly using the internet channel, attended home delivery volume is increasing. Retailers use pricing and availability of delivery slots to improve retailer's operational efficiency goals, while maintaining customer satisfaction. Machine learning methods have been shown to be more accurate than multinomial choice models in forecasting demand for slots, as they can model arbitrary functional shapes and identify heterogeneity in preferences. When the model is used for influencing customer behavior, the learned price sensitivity can be biased due to confounding. We propose a hybrid method to forecast the load as a function of the price and availability of slot alternatives by learning customer and context dependent sensitivity to delivery fee (price) using observational data. We show that the sensitivity estimates are more meaningful (non-positive price sensitivity), and more useful (more heterogeneous), leading to more efficient operations.

## The evolving methodology of Tesco's mid-pandemic forecasting

Presenter: Alex Marsh
Co-authors: Alex Marsh;

As part of running operations and setting up longer term budgets, Tesco uses automated forecasting along with judgemental adjustments. During the pandemic, some of these automated methods became unsuitable and we will discuss how we have dealt with these new modelling challenges.

## Is Large-Scale Demand Forecasting with Machine Learning ready to be Democratized? - Evidence in the Retail Industry

Presenter: Arnoud Wellens
Co-authors: Arnoud Wellens;Maximiliano Udenio;Robert Boute

Despite the recent outperformance of machine learning (ML) based methods, simple statistical forecasting techniques remain the standard approach in retail. We propose and validate a simple-to-use decision-tree framework with the goal to democratize ML forecasting for large-scale retail applications. Based on a test set of 4,548 products of a leading Belgian retailer and a variety of external variables (e.g., promotions and national events), our method improves the forecast accuracy of commonly used statistical methods up to 25%. While more sophisticated versions of our framework only marginally improve the forecast accuracy, we show using Shapley values that external variables and processed input features (e.g., the rolling mean of the unit sales) are crucial for successful implementation. Our findings provide evidence that 'off-the-shelf' ML is finally ready for large-scale retail applications if retailers are willing to invest in computation power, big data and data science skills.

## Dynamic industry uncertainty networks and the business cycle

Presenter: Mattia Bevilacqua
Co-authors: Mattia Bevilacqua;Jozef Barunik;Robert Faff

Industry uncertainty network structures extracted from option prices contain valuable information for business cycles. Classifying U.S. industries according to their contribution to system-related uncertainty, we uncover an uncertainty hub role for the communications, financials, IT and real estate industries. Conversely, shocks to industrials, materials, and utilities do not create strong linkages. The ex-ante network of uncertainty is a useful predictor of business cycles, especially when it is constructed from uncertainty hubs. Policy actions should target uncertainty hubs since these are the main contributors to uncertainty shocks and show a tighter link with the real economy.

## Option-Implied Network Measures of Tail Contagion and Stock Return Predictability

Presenter: Manuela Pedio
Co-authors: Manuela Pedio;

The Great Financial Crisis of 2008 – 2009 has raised the attention of policy-makers and researchers about the interconnectedness among the volatility of the returns of financial assets as a potential source of risk that extends beyond the usual changes in correlations and includes transmission channels that operate through the higher-order co-moments of returns. In this paper, we investigate whether a newly developed, forward-looking measure of volatility spillover risk based on option implied volatilities shows any predictive power for stock returns. We also compare the predictive performance of this measure with that of the volatility spillover index proposed by Diebold and Yilmaz (2008, 2012), which is based on realized, backward-looking volatilities instead. While both measures show evidence of in-sample predictive power, only the option-implied measure is able to produce out-of-sample forecasts that outperform a simple historical mean benchmark.

## Exploiting High-Frequency Option Prices for Modelling and Forecasting Volatility and Equity returns

Presenter: Rodrigo Hizmeri
Co-authors: Rodrigo Hizmeri;Lykourgos Alexiou;Mattia Bevilacqua

We propose option realized measures of variance, skewness, and kurtosis to summarize the informa-

tion from high-frequency option data. These measures aggregate high-frequency option returns from mid-quote prices to estimate higher moments for a given day. Our analysis includes the S&P 500 as well as individual stocks for over 15 years. We study the information content of the option realized measures relative to the daily implied and objective measures for predicting the realized variance and equity returns. In addition, we also construct jump measures and risk-premia, which provide additional information content for predicting the aforementioned variables. Our results show that including high-frequency option information results in superior predictive in- and out-of-sample performance.

## Currency Network Risk

Presenter: Mykola Babiak
Co-authors: Mykola Babiak;Jozef Barunik

This paper identifies new currency risk stemming from a network of idiosyncratic option-based currency volatilities and shows how such network risk is priced in the cross-section of currency returns. A portfolio that buys net-receivers and sells net-transmitters of short-term linkages between currency volatilities generates a significant Sharpe ratio. The network strategy formed on causal connections is uncorrelated with popular benchmarks and generates a significant alpha, while network returns formed on aggregate connections, which are driven by a strong correlation component, are partially subsumed by standard factors. Long-term linkages are priced less, indicating a downward-sloping term structure of network risk.

## A fast and scalable ensemble of global models with long memory and data partitioning for the M5 forecasting competition

Presenter: Kasun Bandara
Co-authors: Kasun Bandara;Hansika Hewamalage;Puwasala Gamakumara;Rakshitha Godahewa

In this study, we present key insights on the model development strategies used in our cross-learning based retail demand forecast framework. The proposed framework outperforms state-of-the-art univariate models in the time series forecasting literature and has achieved overall 17th position in the accuracy track of the M5 competition, which is among the top 1% of solutions.

The proposed framework is an ensemble of pooled regression model and Light Gradient Boosting Machine (LightGBM) model, both trained globally across a collection of time series. These forecast models with the cross-learning capability, also known as Global Forecasting Models (GFM), have recently offered numerous possibilities for forecast practitioners, which have never been realised by the traditional univariate forecasting methods that forecast in isolation. In addition to their outstanding forecast accuracy, GFMs offer much better scalability to the increasing volumes of time series.

Following the recent theoretical recommendations, the base global models of our ensemble framework are carefully crafted to gain complexity through long model memory and a time series grouping strategy. This grouping methodology builds separate global models for each identified group of time series from the retail product assortment hierarchy. On top of the model complexity, the proposed grouping strategy also attempts to address the issue of data heterogeneity in GFMs. Furthermore, to bring model diversity to the forecast combination, we employ both linear (pooled-regression) and nonlinear (LightGBM) cross-learning models in our framework. The proposed framework also accounts for the static and dynamic exogenous variables available in the retail domain by modelling the key driving factors of demand.

The base global models of our framework are trained in parallel to minimise the overall computational cost of our solution. Compared to existing GFM based demand forecast models, which are mainly developed using deep Neural Networks, our proposed framework is a robust, highly computationally efficient, and highly scalable GFM architecture that can be easily deployed in any production environment.

## Probabilistic Causal Effect Estimation with Global Neural Network Forecasting Models

Presenter: Priscila Grecov
Co-authors: Priscila Grecov;Ankitha Prasanna;Klaus Ackermann;Christoph Bergmeir;Sam Campbell;Deborah Scott

We introduce a novel method to estimate the causal effects of an intervention over multiple treated units by combining Probabilistic Forecasting techniques with global forecasting methods using deep learning models. Considering the counterfactual and synthetic approach for policy evaluation, we recast the causal effect estimation problem as a counterfactual prediction outcome of the treated units in the absence of the treatment. Nevertheless, in contrast to estimating only the counterfactual time series outcome, our work differs from conventional methods by proposing to estimate the counterfactual time series probability distribution based on the past pre-intervention set of treated and untreated time series. This paper presents DeepProbCP, a framework for producing accurate quantile probabilistic forecasts for the counterfactual outcome, based on training a global autoregressive recurrent neural network model with conditional quantile functions on a large set of related time series. The output of the proposed method is the counterfactual outcome as the spline-based representation of the counterfactual distribution. We demonstrate how this probabilistic methodology added to the global deep learning technique to forecast the counterfactual trend and distribution outcomes overcomes many challenges faced by the baseline approaches to the policy evaluation problem. Oftentimes, some target interventions affect only the tails or the variance of the treated units' distribution rather than the mean or median, which is usual for skewed or heavy-tailed distributions. Under this scenario, the classical causal effect models based on counterfactual predictions are not capable of accurately capturing or even seeing policy effects. By means of empirical evaluations of synthetic and real-world datasets, we show that our framework delivers more accurate forecasts than the state-of-the-art models, depicting in which quantiles the intervention most affected the treated units, unlike the conventional counterfactual inference methods based on non-probabilistic approaches.

## NeuralProphet applied to hierarchical forecasting of energy: Aggregate compared to pooled and individual models

Presenter: Oskar Triebe
Co-authors: Oskar Triebe;Mateus Gheorghe De Castro Ribeiro;Rajanie Prabha

We compare four practical approaches to forecasting hierarchical time series: aggregate, individual, locally pooled, and globally pooled models. The evaluation is based on predicting the next 24 hours on a large scale energy dataset. Forecasting is a critical operation in the provision of electrical power. Utilities have historically focused on forecasting load in aggregate. The growth of distributed energy resources, such as home solar, storage and electric vehicles, increasingly causes localized issues, stressing the importance of lower level forecasts. We aim to produce accurate forecasts for all levels, from bottom to complete aggregation. The four approaches are evaluated with an interpretable linear model using NeuralProphet. First, we fit a model on the aggregate sum of each hierarchical level's series. Second, we fit an individual model to each bottom-level series. Third, we fit a globally pooled model for all series in the current level. Fourth, we fit a locally pooled model based on a grouping criteria. We present which approach most accurately forecasts across all levels and discuss any trade-offs at different levels, including the likelihood of extreme forecast errors.

## SETAR-Tree: A Novel and Accurate Tree Algorithm for Global Time Series Forecasting

Presenter: Christoph Bergmeir
Co-authors: Christoph Bergmeir;Geoffrey Webb;Rakshitha Godahewa

Threshold Autoregressive (TAR) models have been widely used by statisticians for non-linear time

series forecasting during the past few decades, due to their simplicity and mathematical properties. On the other hand, in the forecasting community, general-purpose tree-based regression algorithms (forests, gradient-boosting) have become popular recently due to their ease of use and accuracy. In this study, we show the close connections between TAR models and regression trees. These enable us to use the rich methodology from the literature on TAR models to define a hierarchical TAR model as a regression tree that trains globally across series, which we call SETAR-Tree. In contrast to the general-purpose tree-based models that do not primarily focus on forecasting, and calculate averages at the leaf nodes, we introduce a new forecasting-specific tree algorithm that trains global Pooled Regression (PR) models in leaves allowing the models to learn cross-series information and also uses some time-series-specific splitting and stopping procedures. The depth of the tree is controlled by conducting a statistical linearity test commonly employed in TAR models, as well as measuring the error reduction percentage at each node split. Thus, the proposed tree model requires minimal external hyperparameter tuning compared to the state-of-the-art tree-based forecasting algorithms and provides competitive results under its default configuration. We also use this tree algorithm to develop a forest where the forecasts provided by a collection of diverse SETAR-Trees are combined during the forecasting process. In our evaluation on six publicly available datasets, the proposed tree and forest models are able to achieve significantly higher accuracy than a set of state-of-the-art tree-based algorithms and forecasting benchmarks across four evaluation metrics.

## A forecasting framework for strategic electricity demand forecasting in a dynamic environment

Presenter: Jana Breedt
Co-authors: Jana Breedt;Louis Louw

The landscape of energy utilization is experiencing rapid evolution, driven by the need to integrate renewable energy resources in the migration towards a sustainable carbon-free footprint. Progression in technology associated with renewable electricity sources introduces new opportunities, increased system complexity and uncertainty. These changes in the energy industry pose significant challenges to forecasting the peak demand for electricity supplied by the electrical networks. Forecasting the electricity demand for a utility serves as the foundation of the network planning of the utility grid. The availability of forecasting frameworks was investigated through literature and a case study analysis to gain real-world insight. Literature provided several frameworks for short term energy forecasting, and limited research was found on long-term forecasting frameworks, with none including the influence from renewable integration. A distinct need was presented to identify factors and interlinking concepts to forecast a utilities' demand and combine these analytically within a forecasting framework. The proposed Integrated Strategic Electricity Demand (ISED) forecasting framework aims to assist an electricity demand forecaster in formulating strategic forecast scenarios in line with business strategy and goals for a utility. The framework was verified and validated with literature and industry experts and implemented it in a case study for the electricity utility of South Africa, Eskom Holdings SOC limited.

## The Value of Probabilistic Forecasts for Bidding and Scheduling in Energy Markets

Presenter: Mario Beykirch
Co-authors: Mario Beykirch;Tim Janke;Florian Steinke

Probabilistic Forecasting in combination with stochastic programming is a key tool for handling the growing uncertainties in future energy systems. Derived from a general stochastic programming formulation for the optimal scheduling and bidding in energy markets we examine several common special instances of local energy systems containing uncertain loads, energy prices, and variable renewable energies. We analyze for several setups theoretically whether only an expected value forecast, marginal or bivariate predictive distributions, or the full joint predictive distribution is required Additionally, we analyze selected cases quantitively in simulation studies. A key results is that for many energy system setups employing a full joint forecast distribution in the stochastic optimization can not improve the results in the actual considered

task. We find that for market schedule optimization, the expected price forecasts are sufficient in almost all cases, while the marginal distributions of renewable energy production and demand are required in most cases. For bidding curve optimization, pairwise or full joint distributions are necessary except for specific instances. Empirical results further show that even in setups in which a probabilistic forecast is theoretically beneficial, the actual value added in the stochastic optimization differs significantly depending on the considered energy system setup. This work helps practitioners choose the simplest type of forecast that can still achieve the best theoretically possible result for their problem and researchers in the field of energy forecasting to focus on the most relevant instances.

## The Short-term Electricity Consumption Forecast Competition Under COVID-19 Lockdown Conditions

Presenter: Irina Chuchueva
Co-authors: Irina Chuchueva;

Imagine you are responsible for the day-ahead electricity consumption forecast and, from tomorrow, your country is quarantined — a new reality is coming. You are asking yourself: Do I need to change my forecast approach? How do I "explain" to the model that tomorrow is the first day of a new reality? What should I do if the quarantine is prolonged? This is my story: I was responsible for the outcome of the short-term electricity consumption forecast competition. The beginning of the competition coincided with the start of the COVID-19 lockdown in Russia. The forecast service I developed won the competition. In this paper, I draw conclusions and give my answers to the three questions above.

## Privacy-Preserving Renewable Energy Probabilistic Forecasting using Differentially Private Federated Learning

Presenter: Jean-François Toubeau
Co-authors: Jean-François Toubeau;Yi Wang

This work presents a new privacy-preserving framework (based on deep learning) for the multi-horizon Probabilistic Forecasting of dispersed renewable energies. This task is indeed becoming increasingly important for cost-effectively managing modern electricity networks. However, traditional forecasting tasks are carried out centrally, by gathering raw data of end-users in a single database that exposes their private information. To avoid such privacy issues, this work relies on two main pillars. First, no exchange of raw information is performed (i.e., measurements are kept local throughout learning and test procedures), which is achieved through a distributed learning scheme, known as federated learning. Second, the learning procedure is augmented with differential privacy. The principle is to inject noise into the training, which is calibrated in such a way that the privacy leakage of any sensitive information can be bounded and quantified. In particular, the methodology is tailored to achieve user-level privacy, i.e., enforcing that the dataset of any stakeholder has a limited impact on the learned model, thus preventing inference of local raw information. The main challenge when computing privacy guarantees for iterative procedures is to keep track of the total privacy loss, which is accumulating along the training. Here, Rényi differential privacy is therefore implemented due to its ability to accurately track cumulative privacy losses in presence of Gaussian noise. In addition to these privacy consideration, the forecasting problem is framed using cross-series learning, which allows to smoothly integrate any new stakeholder joining the collaborative learning task (i.e., cold-start forecasting) without being plagued by data scarcity.

## End-to-End Learning of Coherent Probabilistic Forecasts for Hierarchical Time Series

Presenter: Tim Januschowski
Co-authors: Syama Sundar Rangapuram;Lucien Werner;Konstantinos Benidis;Pedro Mercado;Jan Gasthaus;Tim Januschowski

We present a novel approach to forecasting of hierarchical time series that produces coherent, probabilistic forecasts without requiring any explicit post-processing step. Unlike the state- of-the-art, the proposed method simultaneously learns from all time series in the hierarchy and incorporates the reconciliation step as part of a single trainable model. This is achieved by applying the reparameterization trick and utilizing the observation that reconciliation can be cast as an optimization problem with a closed-form solution. These model features make end-to-end learning of hierarchical forecasts possible, while accomplishing the challenging task of generating forecasts that are both probabilistic and coherent. An extensive empirical evaluation on real-world hierarchical datasets demonstrates the advantages of the proposed approach over the state-of-the-art.

## Forecasting student exam results based on online activity and self-reported self-regulation: A partially interpretable machine learning approach

Presenter: Filotas Theodosiou
Co-authors: Filotas Theodosiou;Yves R. Sagaert;Liam Bossant;Tom Madou

Using learning analytics to model and forecast the student's learning behaviour has been considered highly beneficial. It has been shown to help in designing curricula and reduce the overall dropout rate. During the 2020-2022 pandemic, its broad implementation became more and more relevant due to the significant increase in the amount of students' online behavioural data and the reduction of face-to-face interactions. Despite the advantages, in practice, it is not common to consider learning analytics in higher education. Since teachers lack the specialised knowledge needed to interpret predictions from complex models and have limited faith in unexplained forecasts, they prefer to rely entirely on their judgement. In this work, we combine the expertise of educators with the predictive capabilities of machine learning. A machine learning model is fitted to predict the exam outcomes striving for a balance between accuracy and interpretability. We generate predictions halfway through the semester. Forecasting near the exam period allows for additional data which possibly enhances the overall model's accuracy. However, it leaves insufficient time for students to adjust their learning habits. To improve the learning capacity of the predictive model and make up for the limited information, we link online activity data with self-reported information on students' self-regulation and learning strategies. We use different Gradient Boosting algorithms, enriched with artificially engineered features based on experts' knowledge. Each forecast is coupled with a partial explanation based on the contribution of the specialised predictors, providing a quantitative interpretation of the input-output dependencies, bringing it one step closer to explainable AI. Educators can translate the partially explained predictions into valuable insights that in turn result in personalised student-oriented strategies.

## Probabilistic Time Series Forecasting with Implicit Quantile Networks

Presenter: Adèle Gouttes
Co-authors: Adele Gouttes;Kashif Rasul;Mateusz Koren;Johannes Stephan;Tofigh Naghibi

Deep-learning-based time series forecasting models have become popular due to their end-to-end training, the ease of incorporating covariates, and their automatic feature extraction abilities. Those models are typically recurrent Neural Networks (RNN) or attention-based models. It is often desirable for the outputs to be probability distributions, in which case forecasts provide uncertainty bounds. In the deep learning setting the two main approaches have been to either model the data distribution explicitly or to use Bayesian Neural Networks. The former methods rely on some parametric density function, such as that of a Gaussian distribution, which is often based on computational convenience rather than on the true underlying distribution of the data.

We present IQN-RNN, a deep-learning-based univariate time series method that learns an implicit distribution of outputs. Our approach does not make any a-priori assumptions about the underlying distribution of our data. The probabilistic output of our model is generated via Implicit Quantile Networks (Dabney et al., 2018) (IQN). The major contributions of our paper are: 1. model the data distribution using IQN which allows the use of a broad class of datasets; 2. model the time series via an autoregressive RNN where

the emission distribution is given by an IQN; 3. demonstrate competitive results on real-world datasets in particular when compared to RNN-based probabilistic univariate Time Series Models.

The IQN-RNN model generalises quantile regression, by minimising the integrand of the Continuous Ranked Probability Score. At training time, a quantile is randomly sampled for each observation, time step and epoch. It is then passed to both the quantile loss and the model. The model learns a representation of the quantile value via an embedding layer, and thus a mapping between a quantile and the corresponding quantile value of the forecasted distribution. During inference, we analogously sample a new quantile for each time step of our autoregressive loop. Sampling a larger number of trajectories allows estimating statistics for each observation such as mean, quantiles, and standard deviation. This strategy addresses potential quantile-crossing issues, since the IQN-RNN architecture does not otherwise guarantee monotonicity with respect to the quantile.

## Temporal Hierarchies in the context of AI

Presenter: Nikolaos Kourentzes
Co-authors: Nikolaos Kourentzes;Filotas Theodosiou

Using multiple temporal aggregation levels to forecast time series has been shown to perform well in a wide variety of applications, with Temporal Hierarchies providing a methodology to achieve that. Temporal Hierarchies operate by first constructing temporally aggregate views of a time series, modelling each independently, and subsequently blending the modelled information into a single coherent forecast. The last step can be understood as a constrained linear combination of the independent forecast. This introduces two challenges, the calculation of the covariance matrix between the forecast errors of the independent forecasts that is necessary to obtain the combination weights, and the correlation between forecasts which is introduced by construction, as they are modelling the same target time series, albeit at different aggregation levels. We investigate the use of deep learning to overcome these and improve the performance of temporal hierarchies further. The problem of forecast reconciliation is explored from two perspectives. First, we use an auto-encoder-like architecture to obtain a joint representation of the temporal hierarchy forecasts. Second, we investigate an end-to-end deep network to produce temporal hierarchy forecasts that rely on convolutional layers and LSTMs. We demonstrate the benefits and usefulness of the alternatives and explore their data requirements: when are AI-based hierarchies preferable to conventional ones? Finally, we conclude by proposing a general-purpose architecture for neural networks that embed temporal hierarchies.

## Benchmarking Big Data with mixed-frequency multivariate Time Series Models

Presenter: R. Andrew Butters
Co-authors: R. Andrew Butters;Scott Brave;Ezra Karger;Michael Fogarty;Ross Cole

We outline a signal extraction framework that benchmarks Big Data with traditional stratified random sample surveys in order to retain its appealing features while still maintaining adherence to a broader population. We apply our procedure to mixed-frequency dynamic factor model and mixed-frequency vector autoregression applications utilizing high-frequency Big Data measures related to retail sales and consumer sentiment. These examples illustrate how our framework can be combined with existing statistical methods for multivariate time series to benchmark the timelier and more granular information in Big Data with a traditional data source in a binding or non-binding manner into an estimate of an underlying signal of interest.

## Hierarchical Regularizers for Mixed-Frequency Vector Autoregressions

Presenter: Marie Ternes
Co-authors: Marie Ternes;Alain Hecq;Ines Wilms

Mixed-frequency Vector AutoRegressions (MF-VAR) jointly model the dynamics between variables

recorded at different frequencies. However, as the number of series and high-frequency observations per low-frequency period grow, MF-VARs suffer from the "curse of dimensionality". We curb this curse through a novel convex regularization method that accounts for covariates at different (high-frequency) lags being temporally ordered. Our key modelling tool is a group lasso with nested groups which allows for a hierarchical sparsity pattern that prioritizes the inclusion of coefficients according to the recency of the information the corresponding series contains about the state of the economy. Additionally, we investigate the presence of nowcasting relations by sparsely estimating the MF-VAR error covariance matrix. The sparsity pattern provides evidence on those high-frequency variables (i.e., the series and their particular time period) one can use to build coincident indicators for the low-frequency main economic indicators. In a simulation study, we find that the hierarchical regularizer performs well in terms of estimation accuracy and variable selection when compared to alternative methods. Furthermore, we accurately retrieve nowcasting relations between the low- and high-frequency variables by sparsely estimating the error covariance matrix. We apply the proposed method to a high-dimensional MF-VAR for the U.S. economy to study predictive Granger causality relations through a network analysis. Moreover, we investigate which high-frequency series nowcast quarterly U.S. real GDP growth, use those to construct a reliable coincident indicator of GDP growth and evaluate its performance pre and post Covid-19.

## The forecasting performance of the factor model with martingale difference errors: A comparative study

Presenter: Luca Mattia Rolla
Co-authors: Luca Mattia Rolla;Alessandro Giovannelli

The present paper studies the predictive ability of the factor model with martingale difference errors (FMMDE), recently introduced by Lee and Shao (2018). The FMMDE makes possible to retrieve a transformation of the original series so that the resulting variables can be partitioned into distinct groups according to whether they are conditionally mean independent upon past information or not. In terms of prediction, this allows to achieve optimal results (in the mean squared error sense) by restricting the modelling effort to the subset of factors that exhibit some form of serial dependence in the conditional mean. A further contribution offered by this work is the introduction of a new methodology for determining the true dimension of the FMDDE factor space based on the test for the martingale hypothesis proposed by Wang et al. (2021). The forecasting performance of the FMMDE, equipped with the novel procedure for factor selection, is compared with that of the standard principal-component model defined in Stock and Watson (2002). The empirical analysis is composed of two sections. First, we perform a simulation exercise to understand under which conditions the two approaches differ in terms of forecasting performance. Second, we compare empirically forecasts from these methods using the classic monthly FRED-MD dataset focusing on different forecasting horizons. The results show that for medium to long horizons the FMMDE model appears to have an advantage over the SW model. This result is particularly evident for the output sector while for the price sector the two approaches show the same performance.

## Predicting off-centered steel strips using deep learning

Presenter: Juhee Bae
Co-authors: Juhee Bae;Joakim Ebervik;Gunnar Mathiason;Alexander Karlsson

It is crucial to understand and control the behavior of the dynamic rolling process for high-quality products in the steel manufacturing industry. In this study, the rolling mill is a reversing roughing mill that reduces the thickness of cast stainless-steel slabs from typically 200 mm to 25mm in 5 to 9 passes through the mill. For each pass, the mill sets up a new thinner roll gap, and the material is accelerated into the mill. The process has high dynamics and is sensitive to material straightness and equipment alignment. Improved straightness enables higher material yield and stability in downstream processes. The off-centered measurement of the finished rolled-out slab strip is recorded for each time step under a sensor table. To understand the rolling process and the factors contributing to better-quality workpieces, we propose a deep learning model based on a convolutional neural network (CNN) and a long-short-term memory network

(LSTM). The CNN extracts features from the time-series sequence and the LSTM layer interprets these extracted features across time steps. In this study, we use historical production data that includes multiple input features with 7 features of slab properties, including length and thickness of the slab, and the 15 time-series features from the rolling mill, including the rolling force and speed for each time step. The prediction of off-centered time-series measurement of a target strip is challenging; therefore, we predict the labels from k-means clustering, which groups similar target time-series together. We apply methods to balance datasets, which are inevitable with real-world data. Based on the proposed model, we find the predictors and explain the outcomes of the process based on domain knowledge.

## Interpretable Mixture of Experts for Structured Data

Presenter: Sercan Arik
Co-authors: Sercan Arik;

With the growth of machine learning for structured data, the need for reliable model explanations is essential, especially in high-stakes applications. We introduce a novel framework, Interpretable Mixture of Experts (IME), to provide interpretability for structured data, while preserving the accuracy. IME consists of an assignment module and a mixture of interpretable experts such as linear models where each sample is assigned to a single interpretable expert. This results in an inherently interpretable architecture where the explanations produced by IME are the exact descriptions of how the prediction is computed. During training, IME utilizes past errors made by different experts to improve its assignment process. In addition to providing a standalone inherently interpretable architecture, IME can also be integrated with existing Deep Neural Networks (DNNs) to offer interpretability to a subset of samples while maintaining the accuracy of DNNs. Experiments on various structured datasets show that IME is more accurate than a single interpretable model and performs comparably to existing state-of-the-art deep learning models in terms of accuracy while providing faithful explanations.

## Training an ML model to make outlier adjustments and forecasts simultaneously

Presenter: Casey Lichtendahl
Co-authors: Casey Lichtendahl;Haoyun Wu;Jill Zhou;Weijie Shen;Chris Fry

Without any preprocessing to clean a time series of its outliers (such as spikes and level shifts), a neural network's forecasts are often poor. To improve our neural network's forecasts, we introduce a standard autoencoder as the first two layers in the network. We also introduce the evaluation of a backcast into our loss function. Neural Networks applied to time series data typically evaluate only forecasts in their loss functions. The combination of an autoencoder and the loss function's backcast evaluation adjusts for outliers and improves the model's forecasts—both point forecasts and quantile forecasts. Because our ML-model is a hybrid model, we are able to decompose a series into its outlier, trend, seasonal, and residual components. This decomposition allows us to visually confirm the adjustments made for outliers in the time series history.

## Capacity Planning for a Persistent Service with Growing and Volatile Demand

Presenter: Stijn De Waele
Co-authors: Stijn De Waele;Raja Velu

Making capacity plans for a persistent service when demand is both growing and volatile can be a serious challenge. In the tech industry, we encounter this challenge in several places, including capacity planning for Cloud services. The planner needs a good forecast of the service's long-term growth trend, especially when there are long lead times involved in adding new capacity and long stretches between additions. The planner also needs a good forecast of the volatility around the service's trend. For a given amount of capacity in place between capacity additions, the long-term trend in demand and the instantaneous volatility around that trend determine the service level the end user experiences within that

time period. Thus, the problem reduces to the following: plan capacity so as to meet a service level objective over a future period between additions. To solve this problem, we introduce a two-stage forecasting model using low-frequency data in the first stage and high-frequency data in the second stage. In the first stage, we forecast the long-term trend using the low-frequency demand data. In the second stage, we forecast the high-frequency volatility around the trend given our trend forecast.

## Making and Evaluating Forecasts for Type-2 Service Level Objectives

Presenter: Casey Lichtendahl
Co-authors: Casey Lichtendahl;

To achieve a type-2 service level objective (SLO) over some time period, the capacity of a persistent service may be set so that the ratio of expected censored demand to expected uncensored demand over the time period equals some desired percentage (e.g., 99.9%). Rather than a ratio of expectations, a type-2 SLO may involve an expected ratio of censored to uncensored demand. We call the capacity set to achieve either kind of type-2 SLO a capacitile. A capacitile is a statistical functional in the sense of Gneiting (2011), similar to a quantile or expectile. From the type-2 SLO itself, we derive a new loss function, called the hook loss function, using Osband's principle. The capacitile is the unique minimizer of the expected hook loss, when the expectation is taken with respect to the distribution of the demand process over the SLO's relevant time period. When making and evaluating forecasts to inform capacity plans and to achieve type-2 SLOs, we encourage the use of a capacitile and the hook loss function.

## Forecasting time series of probability density functions

Presenter: Nachiketas Waychal
Co-authors: Nachiketas Waychal;Arnab Kumar Laha;Ankur Sinha

In this work, we propose a new method for modeling and forecasting functional time series data. We focus on probability density functions (PDFs) evolving in discrete-time and explore two related approaches. In the first approach, we use dynamic linear combinations of basis functions in the Bayes Space (B^2) for modeling the functional time series. In the second approach, we first apply transformations, such as the centered log-ratio (CLR) transform, to the given data before proceeding with the modeling. The functional forecast is obtained by extrapolating the estimated dynamic parameters and forecasting the set of coefficients over the forecast time horizon. We demonstrate the proposed method on a real-world dataset of functional time series of PDFs.

## LASSO Principal Component Averaging – a fully automated approach for point forecast pooling

Presenter: Katarzyna Maciejowska
Co-authors: Katarzyna Maciejowska;Bartosz Uniejewski

This paper develops a novel, fully automated forecast averaging scheme, which combines LASSO estimation method with Principal Component Averaging (PCA). LASSO-PCA (LPCA) explores a pool of predictions based on a single model but calibrated to windows of different sizes. It uses information criteria to select tuning parameters and hence reduces the impact of researchers' at hock decisions. The method is applied to average predictions of hourly day-ahead electricity prices over 650 point forecasts obtained with various lengths of calibration windows. It is evaluated on four European and American markets with almost two and a half year of out-of-sample period and compared to other semi- and fully automated methods, such as simple mean, AW/WAW, LASSO and PCA. The results indicate that the LASSO averaging is very efficient in terms of forecast error reduction, whereas PCA method is robust to the selection of the specification parameter. LPCA inherits the advantages of both methods and outperforms other approaches in terms of MAE, remaining insensitive the the choice of a tuning parameter.

## Bayesian autoregressive mixture models based on context trees

Presenter: Ioannis Papageorgiou
Co-authors: Ioannis Papageorgiou;Ioannis Kontoyiannis

Novel prediction techniques are introduced for real-valued time series, based on a new rich class of nonlinear autoregressive (AR) mixture models. These mixture models are defined in terms of partitions of the state space, with a different autoregressive model associated to each region of the partition. The state space partitions are defined in terms of a discretized version of the most recent samples, which is extracted from the real-valued observations using a simple quantizer from the real line to a finite alphabet. The partitions considered in this work are represented as discrete context trees, and they are shown to be meaningful and easily interpretable, while at the same time capturing important aspects of the underlying structure present in the data. The resulting model class is a rich class of flexible autoregressive mixtures, which generalizes popular mixtures of AR models like the Threshold Autoregressive (TAR) models and the Mixture Autoregressive (MAR) models. Along with the modelling framework, a collection of methodological and algorithmic tools are also developed, that allow for exact Bayesian inference within this vast model class, in a computationally very efficient manner. Building on ideas and techniques from the recently introduced Bayesian Context Trees (BCT) framework, it is shown that the maximum a posteriori probability (MAP) model can be identified exactly, including the MAP context-tree partition and the MAP parameters of the AR model in each region. Also, all the corresponding inferential procedures can be performed sequentially, something which facilitates effective online prediction in applications where data are observed on the fly. The performance of the proposed methods in forecasting is illustrated through real-world applications from finance (stock price changes) and economics (modelling the US unemployment rate), considering both single-step ahead and multi-step ahead forecasts. Overall, the proposed methods are found to outperform several commonly used state-of-the-art approaches.

## Wisdom of the Crowd vs. Wisdom of the Few: Citizen Forecasting in Quebec, Canada

Presenter: Brian Thompson-Collart
Co-authors: Brian Thompson-Collart;Adrien Cloutier;William Poirier;Yannick Dufresne

Traditional vote intention polling suffers from a fundamental yet unresolvable issue of sampling validity. Its predictions depend on randomly sampling from a population that does not exist at the time the polls are taken (Jackson & Lewis-Beck, 2020). Recent research suggests that citizen forecasting shows promise as an alternative forecasting method (Murr, Stegmaier, & Lewis-Beck, 2021). Several studies show that estimates made from citizen forecasts can be improved by delegating the forecasting task to the most competent citizens and weighting their responses more heavily (Murr, 2015, Mongrain, 2021). Additionally, citizen forecasts from non-representative samples of the electorate can accurately predict certain election outcomes such as who will lead government or what parties will gain representation in parliament (Graefe, 2016). This paper would (a) elicit forecasts from non-representative samples of citizens with different combinations of demographic variables (b) compare these forecasts to a randomly drawn sample used as a benchmark, and (c) test whether different weighting schemes improve forecasts drawn from non-representative samples.

References Graefe, A. (2016). Forecasting proportional representation elections from non-representative expectation surveys. Electoral Studies, 42(1), pp. 222-228. Jackson, N., Lewis-Beck, M. & Tien, C. (2020). Pollster problems in the 2016 presidential election: Vote intention, vote prediction. Italian Journal of Electoral Studies, 83(1), pp. 17-28. Mongrain, P. (2021). A technocratic view of election forecasting: Weighting citizens' forecasts according to competence. International Journal of Public Opinion Research, 33(3), pp. 713-723. Murr, A., Stegmaier, M. & Lewis-Beck, M. (2019). Vote expectations versus vote intentions: Rival forecasting strategies, British Journal of Political Science, (1), pp. 60-67. Murr, A. (2015). The wisdom of crowds: Applying Condorcet's jury theorem to forecasting US presidential elections. International Journal of Forecasting, 31(1), pp. 916-929.

## Forecasting for lead-time period by temporal aggregation: Whether to combine and how

Presenter: Thanos Goltsos
Co-authors: Thanos Goltsos;Bahman Rostami-Tabar;Shixuan Wang

Temporal aggregation (TA) refers to transforming a time series from higher to lower frequencies (e.g. monthly to quarterly). There are two different types of aggregation: overlapping and non overlapping; which, when added to the option of using original time series, present the forecaster with three approaches to produce required forecasts over the lead-time period: i) non-overlapping aggregation (NOA) ii) overlapping aggregation (OA); and iii) bottom-up to aggregate forecast (BU). Forecasters then need to decide which approach should be used and whether and if so how to combine the forecasts generated by the three approaches. Very often, practitioners are encouraged to aggregate the time series to a frequency relevant to the decisions the eventual forecasts will support, such as lead-time period. Literature also suggests that forecast combination leads to overall improved forecasting accuracy. Forecast combination refers to combining different forecasts, however they are derived, into one. While these two fairly well populated streams of literature present the individual merits of aggregation and forecast combination, to the best of our knowledge, the effect of combining forecasts derived from time series aggregated in different ways has not yet been studied. Using an extensive data set of M4 competition, we design and execute a full factorial experiment exploring the effect of different initial frequencies, data aggregation levels, forecast and combination methods. We are surprised to find that neither temporal aggregation strategies have an overall positive effect on forecasting accuracy. Equally concerning is the fact that straight (average) combinations of these forecasts are similarly of no benefit to the accuracy. To extract the benefits of both well-supported individual forecasting practices of temporal aggregation and combination, we propose a framework that aims to borrow and combine temporal aggregation and forecasting combinations. We find considerable improvements in forecasting accuracy, and discuss areas where the framework is expected to perform best in the future. We also present an agenda for further research on forecasting by temporal aggregation.

## The Impact of Sampling Variability on Estimated Combinations of Distributional Forecasts

Presenter: Ryan Zischke
Co-authors: Ryan Zischke;Gael Martin;David Frazier;Donald Poskitt

We investigate the measurement of sampling variation in estimated forecast combinations, with particular attention given to the combination of forecast distributions. Estimation is implemented via optimisation according to a scoring rule, which may be chosen to reward the form of forecast accuracy that matters for the problem at hand. Forecast performance is also measured using the out-of-sample expectation of a scoring rule, and we study the sampling variability of this performance measure. Importantly, we provide some novel insights into the behaviour of estimated forecast combinations by linking this behaviour to existing theoretical results on one-stage versus two-stage estimation of combinations of models. Our key findings are two-fold. First, the forecast performance of the one-stage estimator – in which the constituent models and the combination function are estimated jointly – is shown to be superior to that of the two-stage estimator. Second, in the limit, all sampling variability in the performance of the two-stage estimator derives from estimation of the constituent models, with no contribution from the estimated weights; a result that has implications for the so-called 'forecast combination puzzle'. These theoretical insights are demonstrated numerically, both in simulation settings and in an extensive empirical illustration using a time series of S&P500 returns.

## Model combinations through revised base rates

Presenter: Anastasios Panagiotelis
Co-authors: Anastasios Panagiotelis;Fotios Petroopoulos;Evangelos Spiliotis

Standard selection criteria for forecasting models focus on information that is calculated for each series independently, disregarding the general tendencies and performances of the candidate models. In this paper, we propose a new way to statistical model selection and model combination that incorporates the base-rates of the candidate forecasting models, which are then revised so that the per-series information is taken into account. We examine two schemes that are based on the precision and sensitivity information from the contingency table of the base rates. We apply our approach on pools of exponential smoothing models and a large number of real time series and we show that our schemes work better than standard statistical benchmarks. We discuss the connection of our approach to other cross-learning approaches and offer insights regarding implications for theory and practice.

## A score-driven model of short-term demand forecasting for retail distribution centers

Presenter: Henrique Helfer Hoeltgebaum
Co-authors: Henrique Helfer Hoeltgebaum;Denis Borenstein;Cristiano Fernandes;Álvaro Veiga

Forecasting is one of the fundamental inputs to support planning decisions in retail chains. Frequently, forecasting systems in retail are based on Gaussian models, which may be highly unrealistic when considering daily retail data. In addition, the majority of these systems rely on point forecasts, limiting their practical use in retailing decisions, which often requires the full predictive density for decision making. The main contribution of this paper is the modeling of daily distribution centers (DCs) level aggregate demand forecasting using a recently proposed framework for non-Gaussian time series called score-driven models or Generalized Autoregressive Score (GAS) models. An experimental study was carried out using real data from a large retail chain in Brazil. A log-normal GAS model is compared to usual benchmarks, namely Neural Networks, linear regression, and exponential smoothing. The results show the GAS model is a competitive alternative to Retail Demand Forecasting in daily frequency, with the advantage of producing a closed form predictive density by construction.

## Panel: What's new in Retail Demand Forecasting?

Presenter: Robert Fildes
Co-authors: Robert Fildes;Stephan Kolassa

Following the presentation from Robert Fildes, the panel will consider the exciting developments and challenges that have emerged in machine learning, the availability of novel data bases and of course the problems posed by the Covid pandemic and other significant events. Without prejudging the panel's conclusions, the increased complexity of the retail environment will require new organizational arrangements to ensure that the potential advances are realized. But we need to answer the question as to the circumstances where the benefits outweigh the costs.

Panel: Chair, Robert Fildes Panelists: Dr Stephan Kolassa, SAP; Dr. Trevor Sidery, Tesco, UK; Professor Kai Hoberg, KLU, Hamburg

## Equity Risk Premium Prediction: The Role of Technical Trading with Mixed Frequency Models

Presenter: Kuok Sin Un
Co-authors: Kuok Sin Un;

Academic research often focuses on evaluating the value of technical trading strategies generated at a specific frequency (minutely, hourly, daily, weekly or monthly). This paper fills a gap by comparing technical indicators sampled at different frequencies. We develop an aggregate technical indicator by employing the Mi(xed) Da(ta) S(ampling), or MIDAS model to bring high-frequency (daily or weekly) technical indicators to a common low-frequency (monthly). Then the aggregated technical indicators are compared in forecasting monthly U.S equity risk premium. Our findings suggest that the aggregated

high-frequency technical indicators (daily or weekly) can add value for momentum trading strategies compared to low-frequency (monthly) technical indicators, with a net-of-transactions-costs annualized certainty equivalent return gain up to 1.54%. While there is no clear evidence of additional economic value for moving average trading strategies.

## An introduction to time-varying lag autoregression

Presenter: Philip Hans Franses
Co-authors: Philip Hans Franses;

This paper introduces a new autoregressive model, with the specific feature that the lag structure can vary over time. More precise, and to keep matters simple in this introduction, the autoregressive model sometimes has lag 1, and sometimes lag 2. Representation, autocorrelation, specification, inference, and the creation of forecasts are presented. A detailed illustration for annual inflation rates for eight countries in Africa shows the empirical relevance of the new model. Various potential extensions are discussed.

## Internal Migration and Climate in Turkey

Presenter: Sule Akkoyunlu
Co-authors: Sule Akkoyunlu;

The present study investigates the impact of economic and non-economic factors on internal migration for 81 Turkish provinces and for the year 2011. The empirical estimates indicate that internal migration is determined, on the one hand by the economic variables such as income, unemployment rate and the employment growth, on the other hand by non-economic variables such as socio-economic factors as well as climate indicators. The coastal variable explains the out-migration as well as the net-migration. In addition, while humidity increases the out-migration, it decreases the net-migration. The results are robust to inclusion of socio-economic-demographic variables as controls.

## Modelling and forecasting minimum and maximum temperatures in the Iberian Peninsula

Presenter: Esther Ruiz
Co-authors: Esther Ruiz;Gloria Ginzalez-Rivera;Vladimir Rodriguez-Caballero

In this paper, we propose a novel methodology to model and forecast minimum and maximum teperatures based on State Space Models for interval valued time series. In doing so, we allow the center and range of temperature to be related and obtain the uncertainty for the temperature trend and dispersion using bootstrap techniques. The methodology is implemented to monthly maximum and minimum temperatures in four localizations in the Iberian peninsula choosen to represent different climate conditions. Namely, we consider temperatures in Barcelona, Coruña, Madrid and Seville. Additionally to modelling interval temperatures at each of these four locations, we also fit a multivariate model to extract potential commonalities among them.

## D for Diversification: Efficient ensembling of Neural Networks for time series forecasting

Presenter: Artemios-Anargyros Semenoglou
Co-authors: Artemios-Anargyros Semenoglou;Evangelos Spiliotis;Vassilios Assimakopoulos

Combining the predictions produced by different forecasting models is considered to be one the most successful strategies for improving accuracy in time series forecasting, as proven in numerous studies and competitions. When it comes to aggregating the output of multiple models, a large number of techniques

have been suggested in the literature, ranging from simple averaging the contributing forecasts to sophisticated combination schemes. Regardless of the chosen combination technique, considering models that lead to truly diverse predictions can remove individual models' biases and increase the ability of the ensemble to accurately forecast a variety of different series. Furthermore, combining the forecasts of multiple models typically results in more robust forecasts overall, especially when the ensemble consists of Neural Networks (NN) that are heavily dependent on randomly initialized parameters. Motivated by these insights , in this study, we explore in greater detail the combination of forecasts produced by feed-forward networks. Due to the opaque nature of NN training, it is not clear which set of network and training parameters should be considered in order to get to a more diversified ensemble that leads to improved accuracy and robustness. More specifically, we focus on the impact that different "high-level" parameters (such as the size of the input vector, the forecasting horizon and the loss function) have on forecasting accuracy. We evaluate all these variants and the corresponding ensembles on the M4 Competition dataset, considering time series of different frequencies. Our results suggest that, using the insights of our analysis, it is possible to create an accurate, robust, easy-to-implement and computationally cheap ensemble of models that outperforms most state-of-the-art forecasting techniques.

## Reinforcement Learning for Portfolio Optimization: An Introduction with Applications

Presenter: Anastasios Kaltsounis
Co-authors: Anastasios Kaltsounis;Evangelos Spiliotis;Vassilios Assimakopoulos

Portfolio Optimization (PO) is a financial engineering problem that typically requires forecasting the prices of particular assets and dynamically allocating wealth among them according to an investment strategy to maximize returns. Although studied for several decades, PO remains a challenging problem, mainly due to the high volatility of the markets and the weak dependence of future prices on historical data and statistics. Reinforcement Learning (RL) has been proven a powerful AI-based tool for analyzing patterns, identifying efficient investment practices and constructing artificial agents that predict prices and allocates wealth accordingly. The successful application of RL in other areas than Trading, such as Robotics, Natural Language processing (NLP) and Gaming, highlights its potential and motivate its use in the field. We provide a brief but complete introduction to RL algorithms for PO, covering key takeaways from the literature, presenting fundamental modeling setups and steps, and analyzing the impact of certain parameters to the performance of the algorithms.

## Classifying time series patterns and modeling segments to aid automated forecasting

Presenter: Youngjin Park
Co-authors: Youngjin Park;Mahesh Joshi

To forecast large-scale time series data, you typically need to run a time series forecasting program to find the best model automatically. But even automatic modeling and forecasting of time series can take a long time, because the pool of possible candidates is so large. If you try to fit a model to each time series by using a detailed, interactive modeling process, it can take even longer when the number of time series is large. One way to tackle this problem is to cluster time series into disjoint groups by using a similarity measure, which is a metric for the distance between two sequences. When the time series are clustered into groups, you need to map the right model pool for each group. However, there is no easy way to identify the characteristics of every group that such similarity-based clustering creates, so it is difficult to automate the decision of which candidate model types to map to each group. This paper describes a different method to reduce the time for automatic modeling and forecasting of time series data. The method consists of three phases. First, the time series are classified by patterns. The patterns can be short, retired, intermittent, seasonal, year-round, etc. The paper provides definitions of these patterns. In the second phase, the series are segmented into groups. In the third phase, the right model pool is mapped to each segment that has been made. Because of the well-defined patterns associated with each segment, it is easy to decide which

candidate model types to use for each segment and then to automate this mapping. For example, a time series in a segment that has an intermittent pattern can be modeled by a temporal aggregation model, and a time series in a segment that has a seasonal pattern can be modeled by a seasonal model. This method helps improve the automated modeling and forecasting of a large number of time series.

## Forecasting Macroeconomic Tail Risk with Big Data Quantile Regressions

Presenter: Jan Prüser
Co-authors: Jan Prüser;Florian Huber

Modeling and predicting extreme movements in GDP is notoriously difficult and the selection of appropriate covariates is key in obtaining precise forecasts. In this paper, our focus is on using huge datasets in quantile regression models to forecast the conditional distribution of US GDP growth. To capture possible non-linearities between output and the regressors we include non-linear transformations of the data. This, in combination with the fact that, by definition, the number of observations in the tails is rather small, we propose a Bayesian approach to shrink our large-scale model towards a simpler model. Since Markov chain Monte Carlo simulation becomes impractical in these dimensions, we rely on fast Variational Bayes approximations to the posterior distribution of the coefficients and the latent states. Our results suggest that huge information sets in combination with priors that introduce substantial shrinkage pays off for forecasting the tail risk of GDP.

## Bias, Information, Noise: The BIN Model of Forecasting

Presenter: Ville Satopaa
Co-authors: Ville Satopaa;Marat Salikhov;Philip Tetlock;Barbara Mellers

A four-year series of subjective-probability forecasting tournaments sponsored by the U.S. intelligence community revealed a host of replicable drivers of predictive accuracy, including experimental interventions such as training in probabilistic reasoning, anti-groupthink teaming, and tracking-of-talent. Drawing on these data, we propose a Bayesian BIN model (Bias, Information, Noise) for disentangling the underlying processes that enable forecasters and forecasting methods to improve – either by tamping down bias and noise in judgment or by ramping up the efficient extraction of valid information from the environment. The BIN model reveals that noise reduction plays a surprisingly consistent role across all three methods of enhancing performance. We see the BIN method as useful in focusing managerial interventions on what works when and why in a wide range of domains.

## Technological interdependencies predict innovation dynamics

Presenter: Francois Lafond
Co-authors: Francois Lafond;Anton Pichler;J. Doyne Farmer

We propose a simple model where the innovation rate of a technological domain depends on the innovation rate of the technological domains it relies on. Using data on US patents from 1836 to 2017, we make out-of-sample predictions and find that the predictability of innovation rates can be boosted substantially when network effects are taken into account. In the case where a technology's neighborhood future innovation rates are known, the average predictability gain is 28% compared to simpler time series model which do not incorporate network effects. Even when nothing is known about the future, we find positive average predictability gains of 20%. The results have important policy implications, suggesting that the effective support of a given technology must take into account the technological ecosystem surrounding the targeted technology. The full paper is available at https://arxiv.org/pdf/2003.00580.pdf/subsection%7BMoments of TGARCH models with skewed innovations}Presenter: M. Angeles Carnero
Co-authors: M. Angeles Carnero;Angel Leon;Trino Manuel Ñíguez

This paper extends the results in He et al. (2008) and Francq and Zakoian (2010) by obtaining the

moments and some cross-moments of the unconditional distribution of the TGARCH(1,1) model under different specifications of the conditional mean. We find analytical expressions for the unconditional variance, skewness, kurtosis and cross-correlations of a time series, generated by a TGARCH assuming that innovations follow different non-normal ditributions such as a Gram-Charlier, SNP or Skewed-t distribution. Our results show that when innovations are distributed as the Gram-Charlier: (i) the skewness of the TGARCH is mainly driven by the skewness of the innovations, having the excess kurtosis of the innovations a smaller effect. However, this is not the case for the kurtosis, since apart from the kurtosis of the innovations, its skewness also affects significantly the kurtosis of the series, (ii) if the conditional mean is not constant, returns can be asymmetric although innovations are symmetric; and (iii) skewed innovations can generate cross-correlations different from zero, indicating leverage effect, even when the volatility model is symmetric. These results are illustrated using nine series of daily stock index returns finding evidence of skewed unconditional distributions for six of them.

## Forecasting in unplanned healthcare service: A Literature Review

Presenter: Mingzhe Shi
Co-authors: Mingzhe Shi;Bahman ROSTAMI-TABAR;Daniel Gartner

The ability to foresee unplanned healthcare events is critical for ensuring a constant level of care quality. Various instances of poor performance, such as long patient waiting times, are a result of poor planning based on patient demand forecasts. To address this issue, predicting unplanned care activities is a critical component of healthcare delivery planning. Numerous review articles have discussed forecasting issues in a single service, but only a few review articles consider the entire unscheduled care system. This paper seeks to bridge that gap. By outlining the present state of the art and opportunities for future research, this paper intends to provide academics and healthcare managers with a better understanding of how forecasting can be employed in unplanned care systems.

A systematic review of the literature was carried out based on a comprehensive search of electronic databases, and included topics ranging from forecasting approaches to healthcare decision-making levels. A total of 2,096 articles were retrieved and 123 studies were identified, and they were classified and analyzed using our classification methodology.

The current review showed that studies on forecasting unscheduled care patient demand primarily focus on emergency care and inpatient care, with comparatively few studies on primary care and surgical care. Additionally, most of the models used in forecasting unplanned care activities were either regression models or Time Series Models, and the usage of the AUC-ROC curve was wide-ranged, reported in 51 studies. When it comes to forecasting decision levels, the majority of studies focus on operational planning, whereas tactical and strategic planning receive less attention.

This review evaluates trends in the research literature and makes recommendations for future research. The primary findings suggest that future research should focus on forecasting across multiple levels of unscheduled care. Meanwhile, given the scarcity of studies utilizing Bayesian and judgmental forecasting models, it is advisable for researchers to consider utilizing these forecasting methodologies. Additionally, future researchers should address the long-term implications of their forecasting algorithms when applying them to healthcare services. Emergency healthcare managers should also be aware of the model limitations associated with various planning horizons.

## Forecasting length of stay in trauma network

Presenter: Zihao Wang
Co-authors: Zihao Wang;Bahman ROSTAMI-TABAR;Jane Haider;Mohamed Naim

Patients' length of stay (LOS) is widely recognized as a critical indicator in evaluating healthcare planning and management in different hospital systems. In recent years, the forecasting of emergency department and paediatric patients' LOS has received considerable scholarly attention. However, the prediction of LOS in some more hierarchical and complex health care systems such as trauma networks has

not been fully discussed. LOS within the trauma system is influenced by a combination of factors such as triage in the prehospital stage, transfer between trauma facilities, timeliness of a series of definitive care and ongoing ward care etc. Moreover, we define the overall LOS in the trauma network as the time from the patient is admitted to a specified trauma facility to the time they discharge. A proper prediction of the LOS could provide insight to effectively allocate and utilize the medical and human resources in the trauma network and scheduling trauma patients' discharge plan. Therefore, the objective of this study is to firstly identify and evaluate timeliness, medical and social predictors of LOS by modelling patient flow based on eight years of admission data in the South Wales Major trauma network. Secondly, to apply machine learning with input relevant predictors to generate a LOS forecast of patients. Our research outcome provides a method for predicting trauma patients' LOS and help improve the overall operational management performance of the trauma network.

## Probabilistic Forecasting of Length of Stay in an Emergency Department

Presenter: Siddharth Arora
Co-authors: Siddharth Arora;James Taylor

The aim of this study is to model the total length of stay (LOS) for patients in emergency departments (EDs). We define LOS as the total time a patient spends in the ED. The decline in ED attendances during the pandemic can partly be attributed to patient concern regarding contagion in the ED and while travelling to seek care. It has been shown that prolonged waiting times are associated with a higher risk of acquiring secondary infections, and an increased risk of mortality and hospital readmissions. Providing personalized forecasts of LOS at an early stage of treatment could improve patient satisfaction, reduce dropout rates, and assist patients in selecting an ED from a network of EDs. Moreover, LOS estimates can assist first responders in prioritizing patients to streamline patient flow, which can help EDs mitigate adverse financial implications, including penalties. Our modelling framework is both probabilistic and personalized. Using data for more than 200,000 patients, we generate and evaluate probabilistic forecasts of LOS and a categorical variable for LOS exceeding a crucial threshold. Using explainable machine learning, we provide insight into the key predictors of LOS. As longer ED LOS is associated with a higher risk of admission to the hospital overall, we also investigate if probabilistic estimates of LOS can be used to forecast the probability distribution of admission. The outcomes of this study can help facilitate efficient capacity planning and resource allocation in EDs.

## Multivariate CRPS Learning with Applications to Electricity Price Forecasting

Presenter: Florian Ziel
Co-authors: Florian Ziel;Jonathan Berrisch

Combination methods can significantly improve forecast accuracy, especially if the considered forecasters utilize diverse models. In Probabilistic Forecasting settings, the predictive accuracy of the forecasters may vary across time and the part of the distribution, e.g. some models exhibit superior performance in the center other ones in the tails. CRPS learning provides an optimal forecast combination framework that utilizes the potential varying performance with respect to certain optimality criteria. We discuss CRPS learning extensions to multivariate frameworks in the context of online learning. We consider empirical applications to electricity price forecasting for selected European day-ahead markets. In addition, we also give a brief introduction and some implementation remarks for the Rcpp based R package profoc which provides efficient multivariate CRPS online learning solutions.

References: Berrisch, Jonathan, and Florian Ziel. "CRPS learning." Journal of Econometrics (2022+), doi.org/10.1016/j.jeconom.2021.11.008

## Aggregating distribution forecasts from deep ensembles

Presenter: Benedikt Schulz
Co-authors: Benedikt Schulz;Sebastian Lerch

The importance of accurately quantifying forecast uncertainty has motivated much recent research on Probabilistic Forecasting. In particular, a variety of deep learning approaches has been proposed over the past years, with forecast distributions obtained as output of Neural Networks. These neural network-based methods are often used in the form of an ensemble based on multiple model runs from different random initializations, resulting in a collection of forecast distributions that need to be aggregated into a final probabilistic prediction. With the aim of consolidating findings from the machine learning literature on ensemble methods and the statistical literature on forecast combination, we address the question of how to aggregate distribution forecasts based on an ensemble of deep Neural Networks. Using theoretical arguments, simulation experiments and a real data study on probabilistic wind gust forecasting, we systematically compare aggregation methods on the scale of probabilities and quantile functions for three variants of neural network-based approaches to Probabilistic Forecasting with different forecast distribution types as output. Our results show that combining forecast distributions can substantially improve the predictive performance. We propose a general quantile aggregation framework for deep ensembles which shows superior performance compared to a linear combination of the corresponding forecast densities. Finally, we investigate the effects of the ensemble size and derive recommendations of aggregating distribution forecasts from deep ensembles in practice.

## On the uncertainty of a combined forecast: The critical role of correlation

Presenter: Andrey Vasnev
Co-authors: Andrey Vasnev;Jan Magnus

The purpose of this paper is to show that the effect of the zero correlation assumption in combining forecasts can be huge, and that ignoring (positive) correlation can lead to confidence bands around the forecast combination that are much too narrow. In the typical case where three or more forecasts are combined, the estimated variance increases without bound when correlation increases. Intuitively, this is because similar forecasts provide little information if we know that they are highly correlated. Although we concentrate on forecast combinations and confidence bands, our theory applies to any statistic where the observations are linearly combined. We apply our theoretical results to explain why forecasts by Central Banks (in our case, the Bank of Japan) are so frequently misleadingly precise. In most cases, a correlation above 0.7 is required to produce reasonable confidence bands.

## Inventory Forecasting and Short Term Fill Rates: Managerial Insights, Forecasts Evaluations and Computational Procedures.

Presenter: Jakub Wojtasik
Co-authors: Jakub Wojtasik;Joanna Bruzda;Babak Abbasi

Flexibility is one of the most important factor in most of the logistic strategies. Combined with more commonly used lean management methodologies inclines to analyzing inventory forecasting in a shorter time horizon. Moreover, current market situation exposes danger of unstable Supply Chains and long term forecasts seem to be highly imprecise. For this reason, we consider inventory forecasting in periodic review order-up-to inventory systems in the presence of commitments concerning the average fill rate. We assume that the performance review period includes multiple replenishment periods (with the basic case when performance review period is equal to the replenishment period). We discuss inventory forecasting in this context within the decision-theoretic framework and introduce the corresponding Murphy diagrams, enabling Forecast Evaluation without focusing on a particular loss function from the family of consistent loss functions. This approach to inventory Forecast Evaluation is illustrated with empirical data and confronted with standard methods such as, e.g., efficiency curves. Furthermore, we consider parametric and non-parametric estimation techniques to set the order-up-to level and, using extensive numerical illustrations, we demonstrate that our computational procedures enable managing risk of not fulfilling an average fill rate requirement, even without assuming any knowledge about the shape of the demand distribution. Several managerial implications are obtained. As a byproduct, we also present direct formulas

to compute optimal base stock levels for common types of distributions used in inventory modelling.

## Using Uncertainty Estimation in Demand Forecasting for Optimizing the Inventory Planning of a German Wholesaler

Presenter: Julia Schemm
Co-authors: Julia Schemm;Nico Beck

Accurate demand forecasts are a major advantage for wholesalers across the board. However, point forecasts are not enough to make optimal ordering decisions because they ignore forecast uncertainty. As most wholesalers try to avoid stock-outs in order to not lose revenue and customers, defining a reorder level or safety stock is crucial. On the other hand, e.g. storage capacity and capital tie-up can be restricting factors. Reliable dynamic uncertainty estimation in demand forecasting can both help prevent stock-outs and at the same time reduce unnecessarily large safety stocks. We apply multiple statistical, machine learning and deep learning Time Series Models (both univariate and multivariate) for forecasting the demand of a German wholesaler in the area of HVAC and plumbing on an article level. Our training and evaluation procedure includes hyper-parameter optimization on a validation set. In addition, we implemented multiple methods from literature for estimating the forecast uncertainty. We evaluate the accuracy of both the point forecast and the uncertainty estimation on historical data of product sales. For deriving the optimal purchasing decision with respect to service level and restrictions like storage capacity, forecasting results serve as input for a mathematical optimisation model.

## Supply Chain demand forecasting under a backordering context.

Presenter: Juan R. Trapero
Co-authors: Juan R. Trapero-Arenas;María Eugenia Babiloni-Griñón;Ester Guijarro-Tarradellas;Diego Pedregal-Tercero

Demand forecasting is a strategic activity of Supply Chain management. Production planning and inventory control decisions are usually driven by those forecasts. Many companies utilize sales as a demand estimation. However, sales and demand are not exactly equal and in case of stockouts, sales become a demand biased estimator. Essentially, there exist two extreme situations in the presence of such stockouts. First, if the client decides to wait until the product is available again, it is called a backordering context. On the other hand, if the client is not willing to wait, then, we have a lost-sales situation. In reality, a mixed of both situations occurs, where a part of the demand is backordered and another part goes to lost-sales. Interestingly, when selecting the "best" demand forecasting model/method, usually, the backordering/lost-sales context does not play an important role in that decision. In general terms, forecasting model selection can be done by either identifying unobservable components as trend, seasonality and cycles in past data or minimizing a loss function or an information criteria. However, how the inventory assumptions (backordering/lost-sales context) influence the forecasting demand model selection has not been clearly investigated. The objective of this paper is to explore how the different forecasting models should be modified to incorporate the backordering process when using sales as an input for the forecasting support system. To do this, we will explore two well-known forecasting models as ARIMA and Exponential Smoothing expressed in a State Space framework. In addition, the inventory system simulated will be a (R, S) system.

## Bond portfolio optimization in turbulent times: a dynamic Nelson-Siegel approach with Wishart stochastic volatility

Presenter: Richard Schnorrenberger
Co-authors: Richard Schnorrenberger;

Modeling and forecasting the time-varying volatility of bond yields play a prominent role in many finance applications. However, amid periods of financial turmoil, managing interest rate risk on a daily basis

is rather a challenging task due to extreme realizations and sudden changes in bond yields that can easily lead to implausible density forecasts. To reduce forecasting uncertainty and account for structural instability in volatile bond markets, the predictive performance of flexible yield curve models with time-varying VAR parameters and Wishart stochastic volatility is investigated under a Bayesian MCMC scheme. A bond portfolio optimization and Value-at-Risk forecasting application to daily US Treasury yields also highlight the potential gains of modeling frameworks with factor Wishart stochastic volatility. The results clearly indicate that the proposed modeling features are economically motivated due to their outperformance in terms of portfolio allocation and risk management during turbulent times including the Great Recession and COVID-19 pandemic.

## Dow Jones Industrial Average and SandP 500 Index Prices: A Time Series Modeling and Forecasting Analysis

Presenter: John Guerard
Co-authors: John Guerard;Dimitrios Thomakos

The Dow Jones Industrial Average, DJIA, is the oldest and most quoted stock index in the United States. The authors received DJIA daily closing prices, since its inception in 1896, until 2020, from the S&P Dow Jones Indices, to celebrate its 125th birthday. The research database included the S&P 500 Index data, since-its inception in 1928. We estimate traditional ARIMA, ARMA with saturation variables, and Adaptive Learning Models, based our 2019 International Journal of Forecasting article on Adaptive Learning Models, and our robust Time Series Models, to forecast the DJIA, since inception, and the DJIA and S&P 500 Index data, in the overlapping periods to estimate Time Series Models of its closing prices. Can we forecast daily stock prices better than random walks? Economists and statisticians have addressed this topic since Alfred Cowles, in 1933, including Paul Cootner and Andrew Lo.

## Cross-Sectional Dependence in Growth-at-Risk

Presenter: Shaoni Nandi
Co-authors: Shaoni Nandi;

This paper explores the impact of cross-sectional dependence in panel data models that link vulnerabilities of future GDP growth with macro-financial and uncertainty indicators. Using a long quarterly panel dataset for 24 countries, we find that including cross-sectional dependence enhances the model performance in both in-sample and out-of-sample evaluations. In a unique finding we note that in the presence of cross-sectional dependence, the indicators commonly associated with GDP catastrophes have limited significance on 5 per cent GaR up to a forecast horizon of 12 quarters, which is relevant for economic policymaking. Encouraged by superior out-of-sample performance, we analyse predicted GaR and estimate a range of measures to quantify risks. We find several meaningful signals of risk and GDP slowdowns from the forecasts of GaR that are relatable to observed data at various points in time. We additionally find that the factors that are used to represent the cross-sectional dependence determine the direction of GaR. Also, these factors have a time-varying impact i.e., a positive role in normal times and exert further downward pull in times of distress.

## European Economic Convergence in the 21st century. What can happen in the COVID19 crisis?

Presenter: Aránzazu de Juan Fernández
Co-authors: Aránzazu de Juan Fernández;Ana Bango Samaniego;Antonio Martín Arroyo

There have been three main events that may have affected the economic convergence in the European Union in the current century: the Eastern enlargement in 2004, the 2008 crisis and the pandemic crisis. The effects on economic convergence of this last event cannot be analyzed today but we can foresee what may happen if we analyze the effects on European economic convergence of the other two events. To

do this, we apply the most powerful inverted beta test (de Juan and Arroyo, 2006) to GDP per capita database for the European countries in order to test the convergence hypothesis. We consider different periods, 1970-2004, 2004-2021, 1970-2007 and 2008-2021, in order to analyze if the economic convergence between the EU members has been affected by these events. We also check for the existence of convergence clubs by applying cluster analysis to the normalized periodogram distances, as suggested in Caiado, Crato and Peña (2006). This in-depth analysis can help to understand how the economic convergence was in the previous events and to foresee what can happen in the COVID19 crisis with the Next Generation EU funds.

## Application of machine-learning models to evaluate air-quality improvement policy: forecasting PM2.5 in South Korea

Presenter: Eunjung Cho
Co-authors: Eunjung Cho;Youngsang Cho

As air pollution has emerged as one of the critical problems that adversely affects human health, many countries, including South Korea, are implementing various policies to improve air quality. Assessing policy effects and devising more effective strategies are equally important as planning and implementing policies. In this study, we quantify the concentration changes of particulate matter with a diameter of less than 2.5μm (PM2.5) in Seoul, South Korea, from January 01, 2019 to December 31, 2019 (policy period) to evaluate the effect of air-quality improvement policies, including the shutdown of old coal-fired plants. We predict business-as-usual (BAU) PM2.5 concentrations while assuming no policy in the policy period using machine- and deep-learning methods, and compared the predicted values with observed ones. For the prediction of PM2.5 concentrations, we use hourly PM2.5, meteorological, and calendar data from January 01, 2015 to December 31, 2018 for training random forest, extreme gradient boosting, and long-short term memory (LSTM) models. Among these models, the LSTM model outperformed the other models as it yields the lowest mean absolute error (1.23 μg/m3) and root mean square error (1.86 μg/m3). Based on the results of the LSTM model, the BAU PM2.5 concentrations over the policy period is predicted to be 28.38 μg/m3. The observed mean PM2.5 concentrations over the same period is 24.66 μg/m3, which implies that the air-quality improvement policies reduce the mean PM2.5 concentration by approximately 3.72 μg/m3. Furthermore, we discovered that state-of-the-art machine- and deep-learning models can be utilized to evaluate the effect of environmental policies, whereas in most previous studies, the effects of air-quality improvement policies are analyzed using conventional atmospheric environment modeling and difference-in-differences methods. The results of this study can be used for developing air-quality improvement policies and related strategies.

## Relationship between GDP and CO2 emissions cycles for the US

Presenter: Lucía Martín
Co-authors: Lucía Martín;

This thesis aims to establish a formal relationship between environmental degradation, measured in CO2 emissions, and the business cycle. To do so we have used emissions data from the US Energy Information Agency as well as economic indicators data mainly from the Federal Reserve of Economic Data of Saint Louis. These economic indicators are used to build two different Composite Indicators that will serve as a proxy for GDP on a monthly basis. In order to build these economic indicators, we are using the Linear Dynamic Harmonic Regression methodology, as it helps us to decompose our variables of interest in their growth and a cyclical component. As a way of establishing relationships between CO2 emissions and the business cycle, we are going to perform an OLS and a Kink-based regression. The main findings of this thesis are that the relationship between emissions and the business cycle exists but only for some of the sources and sectors of emissions that we have taken into account. We have also found an asymmetry in the response of emissions to the different phases of the cycle.

## Forecasting risk appetite using a machine learning sentiment analysis

Presenter: Stuart Royden-Turner
Co-authors: Stuart Royden-Turner;

Sentiment Analysis (SA) is a growing area of academic research and industry focus with a vast array of that includes modelling product review data for user acceptance, voice recording processing to understand products and processes, product and client recommendation systems. My focus is on SA using NLP techniques processing Alternative Data to build a timeseries that can be used in appetite modelling. I review types of natural language data in financial markets (including News and Social Media), modelling methods for both Natural Language Processing (NLP) and SA to generate signals from these Alternative Data. The review of methods starts with simple dictionary-based methods to map words to negative and positive sentiment signals, to more advanced methods that take account of context, such as Lexical methods and finally, the more contemporary approaches using data and advanced Machine Learning (ML) modelling techniques are reviewed. In the next section, I explain my framework for data pre-processing of unstructured financial text data collected from alternative sources. Pre-processed data is further enriched with classic linguistic processing techniques, such as syntactic alternative data sets to provide descriptive analytics, which ultimately modelled with advanced NLP and machine learning techniques. Once the data is fully processed for use in time-series modelling, the study focusses on the methods that are most appropriate for creating accurate economic sentiment timeseries for economic analysis, that can be adopted in a risk-based appetite analysis. Some questions that are asked include:

1. The value in modelling SA with news is well known, does the inclusion of additional sources of information, such as social media, or two set of news data, improve modelling results of sentiment indices?
2. Does the geographic location of an additional data sources predictive power when modelling sentiment indices?
3. How do alternative data, NLP and machine learning techniques fare?
4. Does the inclusion of traditional macro-econometric approaches, such as ARMA improve the modelling results?

## Forecasting the Mobile Market using Statistical and Machine Learning Methods

Presenter: Mohsen Hamoudia
Co-authors: Mohsen Hamoudia;

Machine Learning has gained more importance than ever in both the academic literature and in business and industry. The preponderance of data stemming from the information explosion in all sectors, and the development of various algorithms has facilitated and accelerated the use and application of Machine Learning for time series forecasting. Some studies have shown that Machine Learning could provide key solutions to many issues linked to statistical modeling methods and could be assessed as an alternative to statistical models for time-series forecasting. However, other studies found that statistical methods outperformed popular Machine Learning methods in terms of accuracy measures used and for all forecasting horizons examined. In this paper, we apply and compare several Machine Learning and Statistical methods to the mobile industry and ecosystem. Forecasting the Unique subscribers and the Connections is of high importance to the key players (mobile operators, devices manufacturers, regulation authorities, …). The objective is to improve the forecasts which have proven to be inconsistent in several cases. We also will evaluate the performance of both approaches across multiple forecasting horizons using a dataset of 120 monthly time-series

## A New Bayesian MIDAS Approach for Flexible and Interpretable Nowcasting

Presenter: David Kohns
Co-authors: David Kohns;Galina Potjagailo

In this paper, we propose a new, flexible Bayesian MIDAS framework for nowcasting quarterly GDP

growth, the T-BMIDAS-SVt model. The model incorporates a long-run economic trend (T) and t-distributed stochastic volatility accounting for outliers (SVt) into a Bayesian MIDAS framework. Such a model can quickly become high-dimensional, even with modest amounts of macro indicators, and MIDAS sampling causes a group-correlation structure. We therefore propose the use of a group shrinkage prior which flexibly models between-group and within-group sparsity and nests the popular group-horseshoe prior as a special case. To make the model outcomes interpretable to the policy maker, we propose a group-variable selection algorithm that is motivated by Bayesian decision theory to communicate group-variable importance and its associated uncertainty across the nowcast cycle. We evaluate the T-BMIDAS-SVt in a nowcast application for UK GDP growth, assessing performance of nowcasts over the data release cycle over the period 2002 to 2021, i.e. including the Covid-19 pandemic. We carefully pin down the contributions to nowcast performance of each proposed model component, showing that including either a long-run trend or stochastic volatility substantially improves forecast performance relative to a standard BMIDAS framework, and adding the two in combination provides additional gains. We also compare the T-BMIDAS-SVt to frontier nowcasting methods, showing that our model is competitive prior to the pandemic, while yielding substantial nowcast improvements during the pandemic. The latter relates to the flexible group-variable selection algorithm. Prior to the pandemic the model heavily relied on survey data, mainly production surveys, whereas as the pandemic begins, the model relies more heavily on indicators related to the service sector, consumer spending and mortgages, all of which reflected shifts in consumption behaviour and spending related to lockdowns and therefore contained early information on the large drop in GDP.

## judgyprophet: python package for forecasting with judgmental adjustment using Bayesian informative priors

Presenter: Jack Baker
Co-authors: Jack Baker;Isabell Grübner;Andrew Thwaites

Forecasters are commonly asked to account for known future events. These are large business events which will have an effect on the timeseries e.g. an existing product entering a new market. Not incorporating these events into a statistical forecast can cause significant bias and lead to poor forecast accuracies. We present 'judgyprophet', an open-source python package based on prophet that incorporates these events using Bayesian informative priors. We have found in practice that this method often performs better than using ad-hoc judgmental adjustment or regressors, since the Bayesian methodology is able to update its beliefs post-event. This means that, once the event occurs, judgyprophet is able to correct its prior belief, and if the event deviates significantly from the business predictions, judgyprophet will update its forecast accordingly.

## Forecasting with Prior Wisdom on Group Structure in Panel Date Models

Presenter: Boyuan Zhang
Co-authors: Boyuan Zhang;

Group heterogeneity is recently introduced in the panel data model. We develop a constrained Bayesian grouped estimator that effectively leverages prior information on the latent constant group structure to generate forecasts for panel data. This framework allows us to flexibly incorporate the subjective prior knowledge in the form of pairwise constraints among units. To utilize the pairwise constraints, we revise the standard nonparametric Bayesian prior and propose an intuitive yet valid prior to guide the posterior inference. Interestingly, the resulting Gibbs sample is closely related to the constrained Kmeans algorithm which foreshadows the constrained version of the grouped fixed-effects estimator (Bonhomme and Manresa, 2015). In Monte Carlo experiments, we demonstrate that incorporating prior knowledge generates more accurate estimates and scores predictive gains over unconstrained grouped estimators and standard panel data estimators, even when part of constraints is erroneous. The constrained Bayesian grouped estimator improves the accuracy of forecasting as it optimally extracts the group structure from data, which is in line with prior knowledge. We apply our method to forecast the inflation rate using Phillips curves for various OECD countries. The results show that incorporating prior knowledge into the latent group structure

estimation provides a great deal of flexibility, assists the algorithm in optimally forming the group structure when the data itself is not fully informative, and improves point and density forecasts.