43nd International Symposium on Forecasting

June 25-25, 2023. Charlotsville Virginia, USA

General Chair: Yael Grushka-Cockayne, Program Chair: Matthew Schneider

Book of Abstracts

Differences In College Major Premia Between Urban And Non-Urban Environments

Presenter: Andrew Kim
Co-authors: Andrew Kim

The labor market returns to a college degree vary greatly across different college majors. In addition, labor markets vary in the level of earnings and the kinds of occupations favored across urban and non-urban environments. This paper uses College Scorecard data to examine how college major premia differ between urban and non-urban environments. The distribution of earnings is very similar in both environments, except for Business and Economics majors and for Science majors, both of which perform better relative to other majors in urban environments than non-urban environments.

What The American Public Thinks About Artificial Intelligence And The Future Of Digital Technology

Presenter: Lee Rainie
Co-authors: Lee Rainie

This presentation will focus on the most recent general population surveys by the Pew Research Center of Americans about their views of artificial intelligence (AI). The surveys have covered general attitudes about AI and some of the particular contexts of AI use in health care, social media, facial recognition, workplace settings, human enhancements like brain chips and exoskeletons, and self-driving vehicles. The surveys are nationally representative and cover striking differences in public attitudes by age, race/ethnicity, gender, and socio-economic class. Among other things, they cover the degree to which the public is concerned about bias and discrimination in AI systems. More generally, they cover the aspects of AI advances that they feel are exciting and the advances that worry – even terrify – them. The surveys also reveal significant insights about public opinion differences tied to awareness of AI – or lack of awareness. There are sizeable portions of the population who have heard nothing about the broad concept of AI and also have heard nothing even about some widely-used applications like face recognition. Those who are aware and knowledgeable about AI have strikingly distinct views from those who are not aware of AI. Additionally, the surveys have captured emergent public attitudes about large language models and their applications. For instance, surprisingly few people have heard about ChatGPT and there are varying public views about the possibly helpful and harmful impact of LLMs.

3 Dimensional Correlated Predictive Analytics Risk Modeling And Management Platform

Presenter: William Genovese
Co-authors: William Genovese

The changing consumer experience at exponential scale is one contributing factor in creating risk and compliance disruption in financial services. The "merging" of banks and non-banking entrants (including the "BigTech" e-commerce giants) through marketplace platforms and new business models will introduce regulatory changes from data privacy to the core credit, market, and operational risk pThe required assessment to determine systemic risk exposure will require advancements in financial innovation for the global financial system, such as blockchain with smart contracts with the convergence of AI + Quantum Computing and predictive analytics to determine exposures before they happen. The sheer interconnectedness brought about by new business models and merging of industries will require a "better mousetrap" for identifying much larger complex systemic risk across these digital financial marketplaces. The proposed solution here is to develop a 3 dimensional or 3 pillar (Basel based as a starting point) predictive analytics risk framework and model to cover: 1. Credit Risk 2. Market Risk 3. Operational Risk. The model will leverage data from "black swan" events such as key results and outcomes from the 2008 Financial Crisis, and mortgage industry crashes, Hurricanes such as Floyd, Katrina, and Harvey in the US, the 2011 Japan Earthquake and Tsunami, and other major natural disasters. The model would be constructed on a blockchain using smart contracts that are representative of normal transactional and network data that show impact to credit, market, and operational risk dimensions. It could potentially be developed and built to represent impact for large public institutions (public chain) and smaller institutions (private enterprise, "behind the data center firewall" sidechains). Ideally, existing archetype network topologies based on each institution type, along with transaction types could be used to set up the blockchain structure and nodes. Then by analyzing network data and traffic with AI and Quantum Computing for the black swan events, as compared to normal traffic and processing (as represented in the smart contracts), outliers could be identified–for example–was there an overload of abnormal transactions outside the smart contract parameters that caused error conditions and faults? This information could be found and used by AI to update contracts on the blockchain, and predict with better potential certainty the impact of black swan events.

The Effect Of Dynamic And Interactive Visualizations On Judgmental Forecasts

Presenter: Johanna Giesecke
Co-authors: Johanna Giesecke;Benjamin Buchwitz;Sebastian Knoche

The mode of presentation of time series data in judgmental forecasting tasks has received relatively little attention (Reimers & Harvey, 2023 , p.15). The few existing studies, however, suggest that judgmental forecasts are significantly influenced by the format they are displayed in. Static graphical formats in the type of bar-, line- or scatterplots are known to have various effects on the location and quality of the judgmental forecasts. This raises the question of whether a more dynamic and interactive set of tools affect the location and quality of the judgmental forecasts in a comparable fashion. In this study, first effects of using an interactive line chart that enables forecasters to rescale and rapidly reconfigure their predictions on the location and quality of their judgmental forecast are being examined. An online survey is being conducted to collect forecasts from both non-experts and specialists in data analysis. All participants are presented with data visualizations of synthetic time series and asked to make forecasts based on the information provided via an R Shiny Web App. Moreover, this study also comments on the hypothesis that individuals with greater domain knowledge, and/or profound experience in data analysis and decision-making react differently to the interactive toolset and thus generate different and potentially more accurate forecasts compared to non-experts using static charts. By this, we provide valuable insights into the capacity of individuals to make accurate forecasts based on interactive data visualization and the factors that influence this capacity.

Long-Term Forecasting Of Seasonal Goods Using De-Biased Expert Judgment: Beyond The Pandemic

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

The problem of forecasting demand for seasonal products became even more challenging for retailers amid the Covid-19 pandemic. They are now faced with the difficult task of forecasting demand in the "new normal". On the one hand, the pandemic's future trajectory is hard to predict. On the other hand, an increase in supply constraints forces retailers to confirm orders several seasons in advance. Given this increased uncertainty in future demand, we develop a forecasting model that combines de-biased expert judgment to utilize their knowledge of shifts in consumer behavior, the effects of inflation and other global factors on product demand, and statistical methods well suited for extrapolating repeated patterns from the past. Using data from a premium bicycle manufacturer, we show that accurate demand forecasts for the next three years can be obtained by integrating experts' estimates of the category growth rate with a seasonal decomposition of the pre-pandemic demand curve.

Utility Of Human Judgment Ensembles During Times Of Pandemic Uncertainty

Presenter: Mark Orr
Co-authors: Mark Orr;Juan Cambeiro;Bryan Lewis;Srinivasan Venkatramanan;Justin Crow;Madhav Marathe

Responding to a rapidly evolving pandemic like COVID-19 is challenging, and involves anticipating novel variants, vaccine uptake, and behavioral adaptations. Human judgment systems can complement computational models by providing valuable real-time forecasts. We report findings from a study conducted on Metaculus, a community forecasting platform, in partnership with the Virginia Department of Health, involving six rounds of forecasting during the Omicron BA.1 wave in the UnitedStates from November 2021 to March 2022. We received 8355 probabilistic predictions from 129 unique users across 60 questions pertaining to cases, hospitalizations, vaccine uptake, and peak/trough activity. We observed that the case forecasts performed on par with national multi-model ensembles and the vaccine uptake forecasts were more robust and accurate compared to baseline models. We also identified qualitative shifts in Omicron BA.1 wave prognosis during the surge phase, demonstrating rapid adaptation of such systems. Finally, we found that community estimates of variant characteristics such as growth rate and timing of dominance were in line with the scientific consensus. The observed accuracy, timeliness, and scope of such systems demonstrates the value of incorporating them into pandemic policymaking workflows.

Incorporating Parameters Uncertainty In ETS

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

Exponential smoothing (ETS) is one of the popular forecasting models, used widely both in academia and practice. It produces accurate point forecasts, but several studies showed that it produces narrower prediction intervals than expected. One of the reasons for this is that ETS does not take the parameters uncertainty into account when generating the predictions, it assumes that the parameters are known. This is a strong and a fundamentally wrong assumption, which is violated every time the model is applied to data. While there were some attempts to address this issue in the literature, the proposed approaches, such as ETS with bagging, rely on techniques that have their own assumptions, which can be violated as well. In order to resolve the issue comprehensively, we develop a simulation procedure propagating the parameters uncertainty to the states of the model and then to the final forecasts. The procedure works for any type of ETS model and is embedded in the ETS framework, not requiring any additional time series manipulations

in contrast with the existing approaches. We show how it works and compare its performance with the existing approaches on real data.

Cross-Sectional State-Space Forecasting With Partial Pooling

Presenter: Paul Sangrey
Co-authors: Paul Sangrey;Mike Bedard;Matt Johnson

We propose a novel architecture for time series models built upon state-space methods.We jointly estimate many, potentially multivariate, distributions defined using state-space models by partially pooling their parameters across the cross-section.These joint distributions define a novel recurrent neural network.By combining state-space methods and neural networks, we leverage the interpretability of state-space models and the scalability and flexibility of neural networks.This lets us build an accurate, flexible, and scalable forecast that is not a black box.We implement this architecture by building a library on the deep learning library MXNet to leverage state-of-the-art scalable optimization techniques including automatic differentiation and computation graphs to estimate the parameters governing the state-space models. This library abstracts over a large class of state-space models allowing users to estimate almost arbitrary model specifications without code changes.We forecast weekly business formation by state, which is obtained from the FRED economic database at the St. Louis Federal Reserve bank.We show that this forecast is more accurate than state-of-the-art neural network approaches (DeepState) and more accurate than state-of-the-art univariate generalized linear models (Prophet) when the data are particularly volatile.

Pfr: An R Package For Particle Filtering

Presenter: Taylor Brown
Co-authors: Taylor Brown

pfR is a new R package for particle filtering. In an attempt to promote a more interactive use through an R session, it provides an R interface to the C++ template library pf. All details of the particle filtering algorithms are abstracted away, so no technical expertise in particle filtering is required to specify and use any state-space model. Moreover, it is extensible so that small changes in the code can change the filtering algorithm or any of its tuning parameters. Two things make pfR unique. First, models are specified directly into C++ code, so there is no need to learn a domain-specific language. Second, pfR makes heavy use of class templates. This means almost every customizable parameter of a particle filtering algorithm can be specified at compile time, which can lead to dramatic speed gains. For example, the number of particles in a particle filter, the dimensions of vectors, the resampling strategy, and even the precision of floating points are all compile-time constants.A variety of examples using financial time series are given. Different stochastic volatility models are specified, and performance comparisons of filtering and (approximate) likelihood evaluations are provided. Last we demonstrate a few computationally expensive algorithms that make use of particle filters as a building blocks.

Cross-Temporal Probabilistic Forecast Reconciliation

Presenter: Daniele Girolimetto
Co-authors: Daniele Girolimetto;George Athanasopoulos;Tommaso Di Fonzo;Rob J Hyndman

Forecast reconciliation is a post-forecasting process that maps a set of incoherent forecasts into coherent forecasts which satisfy a given set of linear constraints for a multivariate time series. In this paper we extend the state-of-the-art cross-sectional probabilistic forecast reconciliation to the cross-temporal framework, where temporal constraints are also considered. We develop a non parametric bootstrap and a parametric Gaussian approach to draw samples from an incoherent cross-temporal distribution.

The multi-step residuals are used for a better estimation of the covariance matrix, specifically in the time dimension where the usual one-step residuals fail. To address the high-dimensionality issues, we propose four alternatives for the covariance matrix by exploiting the two-fold nature (cross-sectional and temporal) of the cross-temporal structure and we introduce the idea of overlapping residuals. A simulation study is performed to investigate the theoretical and empirical proprieties of the different approaches. Finally, we consider two empirical forecasting experiments using the Australian GDP and the Australian Tourism Demand datasets to evaluate the feasibility and the performance of the proposed procedures. For these applications, the optimal cross-temporal reconciliation approaches significantly outperform the base forecasts according to the Continuous Ranked Probability Score and to the Energy Score. These findings demonstrate the effectiveness of the proposed techniques in improving the accuracy of probabilistic forecasting models. Overall, the paper expands and unifies the notation for cross-sectional, temporal and cross-temporal reconciliation and investigate the probabilistic cross-temporal framework in more detail.

Multivariate Random Forest-Based Hedging Policy For Seaborne Forward Freight Markets Under Correlated Demand And Price

Presenter: Burakhan Sel
Co-authors: Burakhan Sel;Stefan Minner

Volatilities of freight rates and shipping demand constitute financial risks for charterers and ship owners. The two main players of shipping markets, charterers and ship owners, use Freight forward agreements (FFA) to hedge their financial risk by fixing freight rates for specified cargo amounts in advance. FFAs protect the parties from adverse price changes in freight markets, but they might cause opportunity costs when favorable price changes occur. Moreover, charterers face risks of uncertain shipping demand to be secured in FFAs. Therefore, using FFAs requires careful assessments of future freight rates and demand with forecasting and stochastic optimization methodologies. In this study, we consider the freight procurement problem of a charter that minimizes the total expected freight procurement cost in multiple periods under correlated price and demand. Considering the state-dependent base stock characterization of the optimal policy, we propose a data-driven approach that determines procurement quantities for FFAs using the probabilistic demand and freight rate forecasts instead of point forecasts. A multivariate random forest model estimates possible future freight rate and shipping demand scenarios, enabling decision-makers to approximate optimal procurement quantities with FFAs. With the proposed approach, we contribute to the new and growing literature on prescriptive analytics and forecast-based decision-making, which is limited in hedging decisions, especially under correlated price and demand. Numerical evaluations show that the proposed approach outperforms the policies using point forecasts, data-driven policies from the commodity procurement literature, and baseline policies. Moreover, probabilistic demand forecasts are found to be more critical than probabilistic price forecasts regarding total expected cost minimization. We evaluate the performance of the policies on the market data of 15 shipping routes used by Capesize, Panamax, Supramax, Handysize, Clean Tanker, and Dirty Tanker vessel types. The probabilistic forecast-based policy outperforms the point forecast-based policies. The proposed approach provides the lowest costs in 8 out of 15 routes, the second lowest cost in 5 routes, and the third lowest cost in 2 routes among 10 data-driven and baseline benchmark policies.

Dynamic Harmonic Regression With Simulation For Real-Time Estimate Of Likelihood Of Hitting Sales Target

Presenter: Robert Robison
Co-authors: Robert Robison

Context is key for interpreting incoming sales numbers. Our client desired to estimate the probability that total sales in the month or quarter would meet their forecasts at any point in the time period. We built a Dynamic Harmonic Regression on the daily level that uses simulation to generate an estimate

for the likelihood of the cumulative forecasts to exceed a given threshold. It is updated every day with the newest sales numbers. Since it captures both yearly and weekly seasonality, any major deviance from normal behavior that has an impact on meeting sales targets is immediately identified.A custom seasonal trend decomposition method was also devised to give the client insight into how the dynamic harmonic regression was making its forecasts. First, the seasonal component was removed by isolating the fourier term component of the forecasts for the period of interest. The trend was calculated from the deseasonalized data using a 90 day moving average with a centered Tukey window.

Robustifying The Hamilton Filter And The Beveridge-Nelson Decomposition Using Smooth Robust Forecasting Devices

Presenter: Andrew Martinez
Co-authors: Andrew Martinez

We examine the empirical and theoretical differences and similarities between the Hamilton Filter and the Beveridge-Nelson Decomposition. We show that the differences are largely driven by the initial jumping-off point and whether or not the data has a unit root. Recent advances to the Beveridge-Nelson decomposition have brought these measures closer in line albeit with the disadvantage of less transparency. We also show that both measures can benefit from the use of a smooth robust forecast devices to robustify against structural breaks. We apply our results to quarterly real GDP in the United States and examine how estimates of the output gap implied by these measures evolved following the Great Recession and the COVID-19 pandemic.

Improving Models And Forecasts After Equilibrium-Mean Shifts

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

Equilibrium-mean shifts can result from changes in intercepts with constant dynamics, or be induced by shifts in dynamics with non-zero data means, or both. Induced shifts distort parameter estimates and create a discrepancy between the forecast origin and the equilibrium mean, leading to forecast failure, requiring modifications to previous forecast-error taxonomies. Step-indicator saturation can detect induced shifts, but that does not correct forecast failure. To discriminate direct from induced equilibrium-mean shifts, we augment the model by multiplicative indicators which interact all selected step indicators with the lagged regressand. Forecasts can be markedly improved after induced shifts by including these interactive indicators.

Forecasting Uk Inflation Using Historical Evidence On The Role Of Energy In Productivity And Prices

Presenter: Jennifer Castle
Co-authors: Jennifer Castle;David Hendry;Andrew Martinez

We model UK price and wage inflation, productivity and unemployment over a century and a half of data, selecting dynamics, relevant variables, non-linearities and location and trend shifts using indicator saturation estimation. The four congruent econometric equations highlight complex interacting empirical relations. The production function reveals a major role for energy inputs additional to capital and labour, and although the price inflation equation shows a small direct impact of energy prices, the substantial rise in oil and gas prices seen by mid-2022 contribute half of the increase in price inflation. We find empirical evidence for non-linear adjustments of real wages to inflation: a wage-price spiral kicks in when inflation exceeds about 6–8% p.a. We also find an additional non-linear reaction to unemployment, consistent with

involuntary unemployment. Scenario forecasts for price inflation are produced for 2023 and 2024 by solving out the four conditional models, allowing for analysis of the composition of inflation and the degree of uncertainty based on assumptions regarding the exogenous drivers of inflation.

Comparing Midas And Bayesian Var Models For Gdp Forecasting: Insights From Simulation And Empirical Studies

Presenter: Samir Safi
Co-authors: Samir Safi;Olajide Idris Sanusi

This paper examines the sensitivity of forecasting approaches to model misspecification through a comprehensive simulation and empirical studies, using quarterly time series data for Palestine economic indicators from 2000 to 2020. The efficiency of Bayesian Vector Autoregression (BVAR) models is compared to mixed-data sampling (MIDAS) throughout the simulation processes, using the root mean squared error (RMSE) as a measure of forecasting criterion. The MIDAS and BVAR approaches are evaluated in the context of mixed-frequency data, such as annual, monthly, and quarterly time series data. The study discusses the validity of the models, explores the method of model selection, evaluates the precision of various MIDAS regression and BVAR models forecast. The results indicate that the Bayesian VAR model outperforms the MIDAS and ARDL models, showing the important implications for policy makers and stakeholders who rely on accurate GDP forecasts in decision making. This study fills a significant gap in empirical research on this topic within Palestinian economic indicators with implications for policy makers and stakeholders in the region, highlighting the importance of accurate forecasts in decision making, and offering valuable insights into the best practices for forecasting in mixed-frequency data settings.

Tracking Economic Activity With Alternative High-Frequency Data

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

Most macroeconomic series failed to capture the sharp fluctuations during the COVID-19 pandemic. Also, it proved difficult to extract business cycle information from alternative high-frequency data. We present a Bayesian mixed-frequency dynamic factor model with stochastic volatility for measuring GDP growth at high-frequency intervals. Its novelty is an additional state-space block, in which the sparse observations in the mixed-frequency data are augmented to a balanced panel with observed and estimated latent information. The dynamic factor is then estimated conditional on the augmented data. We apply the model to a set of daily, weekly, monthly, and quarterly time series and extract a dynamic factor, which is identified as the weekly growth rate of GDP. Our model exploits the information in rich datasets well, tracking GDP timely and accurately during volatile periods.

Predicting Recessions, Depth Of Recessions And Monetary Policy Pivots: A New Approach

Presenter: Azhar Iqbal
Co-authors: Azhar Iqbal;Nicole Cervi

To combat decades-high inflation, the FOMC has raised rates at the fastest pace since the 1980s. Will the Committee's actions result in a recession or a soft landing? We present a new framework to predict the shape of the landing. Specifically, our work helps decision makers predict recessions, the depth of recessions (mild or severe) and monetary policy pivots.The first phase introduces three ways to predict recessions. The first method evaluates the effectiveness of a few inverted U.S. Treasury yield curves. The second approach identifies a threshold between the 10-year Treasury yield and federal funds

rate (10-year/FFR). The final tool employs Probit modeling. The second phase of our analysis utilizes the 10-year/1-year Treasury yield spread to predict the duration of a recession. Historically, twelve consecutive months of a negative 10-year/1-year spread is associated with deeper recessions. Evaluating the tools with data going back to 1955, the 10-year/1-year spread has predicted all 10 recessions with an average lead time of 12 months. The 10-year/FFR threshold has predicted all recessions and monetary policy pivots with an average lead time of 18 months. At present, all three tools are signaling a recession and/or Fed policy pivot is more likely than not within the next year. The 10-year/1-year spread breached the recession-prediction threshold in August 2022. Through February 2023, the spread has remained negative for eight consecutive months, which highlights the risk that an upcoming recession may not be mild. Meanwhile, our Probit framework suggests there is a 57% chance of a recession during the next four quarters, and the 10-year/FFR threshold breached back in March 2022. Essentially, our proposed early warning system will help decision makers, in real time, to determine whether an upcoming recession will be mild or severe and whether changes in the current monetary policy stance are imminent.

Forecasting The Power Demand Of Households Within A Mexican Community.

Presenter: Jorge Ángel González-Ordiano
Co-authors: Jorge Ángel González-Ordiano;Karla Obermeier-Velazquez;J. Emilio Quiroz-Ibarra;Lázaro Bustio-Martínez;Daniel Alejandro Perez-Delamora;Guillermo Fernandez-Anaya

The challenges of the energy transition, such as the volatile generation of some renewable energy systems, has made the forecasting of power demand essential in maintaining the necessary balance between energy supply and demand. Obtaining forecasting models requires power demand time series with various resolutions. As this type of data is not commonly available in Mexico, we have been collecting power demand data of various household within a Mexican community using Internet of Things (IoT) sensors. Our goal is to offer the data as part of an open dataset as a way of pushing the energy transition and energy forecasting topics within Mexico and its academic community forward. We noticed, however, that obtaining accurate forecasting models for the time series collected is not a trivial task, due to a lack of consistent patterns and noise. Therefore, we speculate that using hierarchical forecasting methods that forecast the aggregated time series and then derive the forecasts for the individual households, might be useful. In this work, we present the results of using top down hierarchical forecasting methods with our time series data and compare it to methods that forecast the demand of the individual households separately.

Mastering the Flow of Time: Building Robust Time-Series Forecasting Pipelines with a Weather Twist

Presenter: Sankalp Gilda
Co-authors: Sankalp Gilda;Jack Pfeiffer;Srishti Sehgal;Erik Hasse;Christian Brown

In this engaging and in-depth presentation, we delve into the core components of developing robust and industrial-grade time-series forecasting pipelines. Drawing from lessons learned in both academia and industry, we expose common pitfalls and illuminate best practices at every stage of the process. Topics range from data preparation, appropriate cross-validation techniques, and advanced feature engineering to effectively handling data leakage, addressing concept drift, and ensuring effective performance evaluation. We then shift our focus to a more specific application: the strategic use of single-site weather data as exogenous variables. The challenges of integrating such data and making it a meaningful component of forecasting models are explored in detail. The theories and concepts discussed are anchored by a hands-on case study, demonstrating their tangible impact on forecast accuracy and overall model performance. This presentation serves as a comprehensive guide for professionals seeking to enhance their time-series forecasting skills, bridging the gap between academic theory and practical industry application.

Risk Management in Wholesale Electricity Markets: A Signal Processing Approach

Presenter: Ritvana Rrukaj
Co-authors: Ritvana Rrukaj;Benjamin P. Fram;Leif K. Sandal

Medium-term load forecasting is a critical tool utilized by power market system operators for system planning and load-serving entities for procuring power supply contracts. Techniques from the field of signal processing are employed to model seasonal load in the New York wholesale electricity market (NYISO). We begin by using mathematical filtration techniques to smooth raw NYISO load and price data that spans 2006 to 2018. The resulting filtered data series express load and prices as percentage deviations from their annual moving averages. Next, we develop a nonlinear load model in price and time by using the 90-day moving averages of the filtered price and load data and time-dependent periodic terms to capture the cyclicality of the load. In the final step, we model the noisy residuals using an autoregressive (AR) process. We conclude by performing out-of-sample forecasts of our model using test data from 2019 and 2020. The forecasting results reveal that our model can predict seasonal load with a high degree of accuracy and potentially assist market participants with their medium-term planning objectives.

Forecasting Duration Of Hospital Stays In Spain Caused By The Impact Of Increasing Atmospheric Co2 Concentrations Using Gamlss Models

Presenter: Roberto Morales-Arsenal
Co-authors: Roberto Morales-Arsenal;Ángela Rabadán-Navarrete

Climate change is producing adverse effects that are being ignored by the health insurance industry. This study develops a methodology that evaluates the effect of CO2 concentration levels on the duration of hospital stays and, therefore, on hospital spending in Spain. A regression analysis has been applied using the generalized additive model for position, scale and shape with Box-Cox power exponential distribution whose main advantage is its flexibility. The model explains 94% of the variability of CO2 levels, which in turn determine the inter-annual variability of stays by 85%. Additionally, a novel CO2 forecasting method has been developed whose predictive capacity was satisfactory. Annual data from 2005 to 2019 disaggregated by autonomous communities have been used. The results show the importance of evaluating the effects of pollution that can modify the expected behavior of the underwriting risk.

Detecting Optimal Cut-Off Points In Medical Studies: Improve Survival Prediction Accuracy By Stratifying Hazard Ratios

Presenter: Hae Rim Kim
Co-authors: Hae Rim Kim;Jieun Shin;Young Jun Cho

In survival analysis, hazard ratios are commonly used to measure the effect of a variable on the risk of an event, such as death or disease progression. By stratifying the hazard ratio, we can account for these differences and obtain a more accurate prediction of survival. In this study, we compared the F1-score of machine learning models with and without stratified hazard ratios in predicting survival in a thyroid cancer dataset. To address the imbalanced data, the SMOTE algorithm, which generates synthetic samples, was used.This study provides the significance of accounting for patient heterogeneity when predicting survival in medical research. Our analysis indicates that by identifying optimal cut-off points and stratifying hazard ratios, the accuracy of survival predictions can be significantly improved. These results suggest that the use of stratifying hazard ratios could be a valuable tool in personalized medical diagnosis and treatment, potentially increasing the effectiveness of medical interventions.The PELT and linear regression tree algorithms were used to find the best cut-off points in continuous data. Log-rank test and Kaplan-Meier

analysis were evaluated to ensure that all clusters were significantly stratified. And bootstrap was used to confirm the stability of the cut-off point.

Forecasting Pediatric Heart Donors

Presenter: Michael Porter
Co-authors: Michael Porter

Pediatric heart transplantation remains the only chance at survival for children with end-stage heart failure or inoperable congenital heart defects. Children in need of a heart transplant enter a waiting list according to their urgency (Status 1A, 1B, or 2). When a donor heart becomes available, a ranked list of waitlisted candidates is generated based on the distance between the donor and candidate, the compatibility between donor and candidate blood types, the candidates' status and time on the waitlist. When a patient receives an offer, their clinicians face the difficult decision to accept the current offer or wait for a better match. While there is a growing body of research on measuring donor suitability, there is limited work on predicting the time till next offer. This talk will discuss our approach for forecasting donor supply, the first step in predicting candidate waiting time. We model donor supply as a marked spatio-temporal point process where the marks refer to certain donor characteristics that determine compatibility with a candidate (e.g., age, weight, blood type, cause of death). We identify seasonal patterns, dependence between marks, effects of COVID-19, and forecast the time between donors and donor counts over multiple forecast horizons and donor hospitals. These forecasts are used as inputs to another system which predicts the time till next offer for all waitlisted patients.

Multivariate Modeling Of Tourism Expenditure By Applying Vine Copulas. An Application To The Case Of Fribourg - Switzerland

Presenter: Miriam Scaglione
Co-authors: Miriam Scaglione;Juan Gabriel Brida;Leonardo Moreno

This study models the joint distribution of tourism expenditure disaggregated by category simultaneously with a set of covariates associated with the tourist trip (including destination, length of stay, season of the year, among others).The dependence structure, among all variables, is established by means of a family of paired copulas (regular vines), which allows fitting a high-dimensional multivariate statistical model.As an application, the model is calibrated on a database of the tourist region of Fribourg (Switzerland). The good fit of the model to the data is observed both in the marginal distributions and in the dependency structure.The empirical results provide different examples of the usefulness of the model in answering different questions. In particular, the study shows that the key variables to understand the associations between them are the destination in the region and the place of origin of the tourist.

Assessing The Competitiveness Of The European Airbnb Sector In Times Of Disruptions

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

The COVID-19 pandemic has had an unprecedented impact on all sectors of the tourism and hospitality industry. Airbnb, one of the major players in the accommodation domain, was not exempt from these disruptions. In light of changing consumer accommodation preferences as a consequence of COVID-19, this study will delve into assessing the competitiveness (i.e., efficiency) of the European Airbnb sector in three different periods: pre-, during, and post-pandemic. More specifically, the aim is to assess the efficiency of the Airbnb sector of European countries by employing interactive data envelopment analysis (DEA)

approach between 2018 and 2022. Selection of input and output variables for DEA modeling is grounded in the literature, and the attempt will also be made to include uncontrollable variables such as COVID-19 Stringency Index. Given the analytical focus and arguably, undeniable stakeholders' interest in maximizing demand, the output-oriented BCC (Banker, Charnes, Cooper) modeling is a priori deemed the most fitting for this study. The contributions lie in a novel, fully-fledged efficiency analysis and benchmarking of the Airbnb sector in both longitudinal and national contexts, resulting in a detailed assessment of the sector's performance in periods of different levels of disruptions. Ultimately, the efficiency scores of DEA can also be used in the subsequent analyses for modeling Airbnb demand and for producing reliable forecasting scenarios.

Impacts Of Multiple Crises On The British Tourism Market: A Counterfactual Forecasting Perspective

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

The tourism industry is a significant contributor to the UK economy, while the current British tourism market has been profoundly shocked by Brexit, the COVID-19 pandemic and the cost-of-living crisis. The overlaps between events create challenges in isolating the impacts of the individual crisis. Although the situation is looking brighter moving forwards with the fade of the COVID-19 pandemic, tourism practitioners and policymakers are still facing an uphill battle and trying to adjust to the fast-changing situation. To support effective tourism management, the impacts caused by these three events should be isolated and evaluated independently. A few methods such as the Difference-in-Differences and regression discontinuity design have been widely used in economics to assess policy or crisis impact in a semi-experimental setting with intervention and control groups. As the pandemic and hyperinflation have widely affected most of the major economies, it is difficult to identify the control group and classical causal inference methods failed to estimate the impact of those crises. To fill in this gap, the present study applies the Bayesian structural time series (BSTS) model to make counterfactual forecasting. A basket of countries whose market structure of volume closely match that of the UK before the Brexit referendum are introduced as the covariate to build a counterfactual UK that did not leave the EU. The COVID-19-related covariates are selected to disentangle the costs of the COVID-19 pandemic on the British tourism market. In this way, the tourism recovery pattern with the absent cost-of-living crisis can be obtained. Isolating the impacts of Brexit, the COVID-19 pandemic and the cost-of-living crisis on British travel markets, this study includes the monthly data of the major European countries and several Anglosphere countries. The variable selection also considers the public health and government restriction indexes such as death record, reproduction rate, and public stringency index. This research will generate new knowledge on the understanding of the British travel market across the challenging times and provide a valuable analytical framework for isolating the impacts of multiple major events on tourism.

Forecasting European Airbnb Occupancy During The Pandemic: The Benefits Of Panel-Data And Markov-Switching Models

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

This study evaluates the forecast accuracy of various models applied to monthly Airbnb occupancy during the COVID-19 pandemic. Data for 43 European countries are employed. The models are estimated over a training set ranging from January 2017 to December 2020, thus explicitly including the pandemic period. Out-of-sample forecasts are produced for a test set ranging from January 2021 to December 2021. The seven rival models consist of two econometric models employing the COVID-19 Stringency Index as a measure of response strategies against the pandemic (panel-data models with and without macroeconomic controls), two naïve models (naïve-1 and seasonal naïve), ETS, SARIMA, and seasonal Markov-switching

autoregression. Forecast accuracy is evaluated in terms of RMSE, MAE, MAPE, and SMAPE for the total of countries, as well as for several sub-samples. Due to the pandemic, accurate forecasts are more difficult to produce than during "normal" times. However, Markov-switching and panel-data models allowing for the pandemic are typically preferred over "standard" models such as seasonal naïve or ETS across accuracy measures and sub-samples.

Making Your Models Sticky: How Aes Deploys Forecasting At Scale In A Mercurial World

Presenter: Sean Otto
Co-authors: Sean Otto

A sticky AI model is one that users find informative, useable and accurate…. but often times we find that no model survives impact with the user or the world. As a renewable energy company, forecasts are becoming core to delivering energy to the grid in a market that is becoming more intense and mercurial. From weather impact forecasts to asset maintenance forecasts to energy delivery forecasts the kaleidoscope of challenges continues to amaze and confuse all while business demands consistency and mitigation of risk. Welcome to that true challenges of applied models in a data driven world where stickiness is only as good as yesterday's success. Learn about our journey and join us for a conversation around how AES is creating its forecasting models, scaling those models and engaging its business users.

Right-Sizing A Commodities Price Forecasting Framework Using Automatic Machine Learning

Presenter: Jesse Luebbert
Co-authors: Jesse Luebbert

Time series prediction is critical in many domains, including finance, healthcare, and industries. The success of these models often depends on the quality of the features used in the modeling process. Therefore, a compelling feature engineering framework is crucial for developing accurate time series prediction models - a process that became easier as well as more powerful with the advent of automated machine learning (AutoML) tools like H2O Driverless AI.This project presents a case study on using H2O Driverless AutoML to predict PET (Polyethylene terephthalate used for plastic bottles) price changes in the next 12 months using more than 15 years of historical data and 30 different global financial indexes as exogenous features. By focusing on optimizing and controlling the feature engineering process utilizing a genetic algorithm, the proposed methodology outperforms other feature engineering approaches in terms of directional accuracy and mean absolute error, even with a relatively small dataset and simpler models.The success of this approach highlights the importance of tailoring the input variables to the specific prediction task for capturing patterns of trends and seasonalities. Moreover, the study discusses the challenges encountered during the project, such as dealing with missing data as well as preventing overfitting while creating thousands of new features. It was possible to compare the results with other time series prediction models and highlight the benefits of using AutoML for this task.The findings have broad implications for anyone interested in developing accurate time series prediction models with a framework that can be applied to other prediction tasks across different sectors and industries.

How's The Weather Tomorrow? An Introduction To Vre Forecasting

Presenter: Andres Figuerola
Co-authors: Andres Figuerola

Forecasting Variable Renewable Energy (VRE) is of fundamental importance for Energy market participants. Given the sudden growth in the VRE world, a scalable and quickly replicable solution is key in sustaining the rapid expansion of the sector. VRE output depends greatly on the weather conditions which can be difficult to forecast. Additionally, the accuracy of VRE forecasting can have a significant economic impact on the outcomes and performance of energy-trading decisions.In this talk, we will start with some of the basics of energy markets and how they are influenced by VRE generation. I will then share some of the main challenges we face while creating these forecasts, depending on the technology. Lastly, we will talk about scaling approaches at AES and some unusual suspects that can impact the energy output.

Deploying Best-Of-Breed Metalearning And Forecasting Frameworks With H2o's Tidal Pulse.

Presenter: Jon Farland
Co-authors: Jon Farland

The interest surrounding time series forecasting from both academia and industry alike has increased in the last decade and has correspondingly spawned a wide set of tools and frameworks for producing forecasts. Some are purely focused on forecasting particular quantities such as energy demand, while others focus on business operations such as supply chain. Whereas these tools are often specialized and commercial grade pieces software, recent advances in automatic machine learning (AutoML) allows for more general ecosystems to be built for time series modeling. Tidal Pulse is a new time series modeling and forecasting engine available in the H2O AI Cloud that combines cutting edge AutoML techniques with a scalable and temporally partitioned database. The underlying data model forms a foundational layer for visualizations of trend and seasonality components across time series, as well more advanced analytics. Forecasters can interact with their data, applying pre and post processing layers, as well as generate both point and probabilistic forecasts from models ranging from traditional statistical methods to the more recent applications of deep learning architectures. Time series can be segmented by various clustering techniques, and AutoML tools like H2O Driverless AI can be leveraged for automatic feature engineering and metalearning. Rather than being limited to a single forecasting framework, this new engine enables forecasters to combine advanced open source frameworks with commercial grade AutoML tools in a single forecasting ecosystem.

Forecast Multivariate Time Series Using Lower Dimensional Components

Presenter: Yangzhuoran Fin Yang
Co-authors: Yangzhuoran Fin Yang;Rob J. Hyndman;George Athanasopoulos;Anastasios Panagiotelis

High-dimensional multivariate forecasting is often restricted to univariate models that do not consider inter-series relationships or multivariate models that are difficult to estimate. Dynamic Factor Model gained popularity by using factors to capture common signals but is limited to linear transformations of series. We propose a framework to forecast multivariate time series using potentially nonlinear components, which capture common patterns shared across series. This procedure does not limit the choice of forecast model but serves as additional steps. It is applicable even to components without an inherent back-transformation by training a back-transformation model using bootstrap samples and expanding windows. We demonstrate its usage with principal component analysis in a simulation example. In the application of Australian tourism data, using principal components and ISOMAP to forecast has shown superior performance in short-term forecasting while Laplacian Eigenmaps have shown competitive results for longer forecast horizons. The forecast performance is found to be robust to the number of components.

Estimation And Forecasting Methods For Integer Autoregressive Processes

Presenter: Pashmeen Kaur
Co-authors: Pashmeen Kaur;Peter F. Craigmile

A popular and flexible model for time series of counts that are Markov in time is the integer autoregressive process of order p, INAR(p). These processes are defined using thinning operators based on past values of the process and an independent in time discrete innovation process. These models are especially useful in epidemiology, where we are interested in modeling and forecasting disease counts over time. Through the choice of thinning operator and innovation process, these models have been shown to be well suited for modeling time series of low counts, and can allow for under- or overdispersion. While there are many choices of thinning operators and innovation processes given in the literature, less focus has been spent on comparing statistical inference and forecasting procedures for these models. We provide an extensive study of exact and approximate inference and forecasting methods that can be applied to a wide class of INAR(p) processes. We summarize the applicable theory, provide simulations to compare small sample performance, and illustrate this methodology using a disease surveillance series. We close with a discussion of possible extensions of this work.

Exploring New Horizons In Human Activity Research: Time Series Features Of Points Of Interest Visitation Patterns

Presenter: Jesse Piburn
Co-authors: Jesse Piburn

Traditional strategies for using visitation patterns at Points of Interest (POIs) in human dynamics research primarily rely on comparative analyses. While these methods have offered valuable insights, they often face constraints that may limit our full understanding of human activity patterns.In this light, our presentation shares an ongoing exploration into a promising alternative: time series feature analysis for understanding POI visitation patterns. By focusing on time series features, this approach doesn't rely on the more typical relative comparisons and can more easily be integrated into existing AI/ML models. This method considers multivariate analysis and retains the diversity of individual visitation behaviors within a dataset, aiming to provide a more comprehensive picture of the underlying patterns.Our ongoing work seeks to enhance our understanding of human activity by investigating visitation dynamics at POIs in greater detail. We will discuss the potential techniques and advantages of time series features for visitation pattern comparison in our presentation, demonstrating how this approach can offer industry professionals a more nuanced perspective on human behavior.

Strategies For Learning Inherently Interpretable Additive Forecasting Models

Presenter: Oskar Triebe
Co-authors: Oskar Triebe;Karl Richter;Christoph Bergmeir

Trust is a prerequisite for the adoption and use of a forecast. Trust is built with repeated confirmations over time, or, by understanding the forecasting system and its reasoning. Forecast interpretability is especially of relevance in domains where decisions are of high stakes and involve humans (eg. energy load forecasting). Additive forecasting models that combine multiple time series component functions are a common approach to creating inherently interpretable forecasts. While the composition of simple time series components (eg. trend, seasonality) commonly has a unique optimal solution and is therefore inherently interpretable, we have found that this condition does not hold for compositions of more complex time series components that overlap in their expressiveness (eg. trend and autoregressive component). In an unconstrained setting, one can observe that the learned component functions are not stable across repeated runs, nor are there any guarantees that the learned functions adhere to the theoretical definition

of its time series component. While most additive forecasting models employ some form of mechanism that effectively stabilizes the learned model in practice, we found that no previous work explicitly discusses the underlying problem, solution approaches or provides reasonings why one mechanism was chosen over another. In this work, we review strategies for interpretability that can be found in current forecasting and time series decomposition methods. Further we define three desiderata for additive forecasting models: Component interpretability, prediction accuracy and explanation stability (SAI). Lastly, we perform a quantitative evaluation of each strategy with respect to the SAI dimensions. We believe that this work is a meaningful contribution that (a) increases the awareness of forecasting practitioners on the interpretability strategies and their implications for the model they are using and (b) provides forecasting researchers with a structured overview of interpretability strategies that can be used in any additive model.

Support Vector Regression: Risk Quadrangle Framework

Presenter: Stan Uryasev
Co-authors: Stan Uryasev;Anton Malandii

The Support Vector Regression (SVR) is investigated in the framework of the Fundamental Risk Quadrangle. Both formulations of SVR, ε-SVR and ν-SVR, correspond to the minimization of equivalent regular error measures (Vapnik error and CVaR norm) with a regularization penalty. These error measures, in turn, define corresponding risk quadrangles. By constructing the risk quadrangle corresponding to SVR, we show that SVR is the asymptotically unbiased estimator of the average of two symmetric conditional quantiles. Furthermore, the quadrangle approach shows the equivalence of ε-SVR and ν-SVR. Additionally, SVR is formulated as a regular deviation minimization problem with a regularization penalty by applying Error Shaping Decomposition of Regression. Finally, the dual formulation of SVR is derived in the risk quadrangle framework.

New Robust Inference For Predictive Regressions

Presenter: Rustam Ibragimov
Co-authors: Rustam Ibragimov;Jihyun Kim;Anton Skrobotov

We propose two robust methods for testing hypotheses on unknown parameters of predictive regression models under heterogeneous and persistent volatility as well as endogenous, persistent and/or fat-tailed regressors and errors. The proposed robust testing approaches are applicable both in the case of discrete and continuous time models. Both of the methods use the Cauchy estimator to effectively handle the problems of endogeneity, persistence and/or fat-tailedness in regressors and errors. The difference between our two methods is how the heterogeneous volatility is controlled. The first method relies on robust t-statistic inference using group estimators of a regression parameter of interest proposed in Ibragimov and Muller, 2010. It is simple to implement, but requires the exogenous volatility assumption. To relax the exogenous volatility assumption, we propose another method which relies on the nonparametric correction of volatility. The proposed methods perform well compared with widely used alternative inference procedures in terms of their finite sample properties.

Change Point Detection In Time Series Using Mixed Integer Programming

Presenter: Alexander Semenov
Co-authors: Alexander Semenov;Artem Prokhorov;Anton Skrobotov;Peter Radchenko

We use recent advances in mixed integer optimization (MIO) methods to develop a framework for identification and estimation of structural breaks in time series. The framework requires a transformation of the the classical structural break detection problem into a Mixed Integer Quadratic Programming problem.

MIO is capable of finding provably optimal solutions to this problem using a well-known optimization solver. The framework allows to determine the unknown number of structural breaks. In addition to that, we demonstrate how to accommodate a specific required number of structural breaks, or a minimal required number of breaks. We demonstrate the effectiveness of our approach through extensive numerical experiments on synthetic and real-world data. We examine optimal and sub-optimal solutions of the problem, and the effect of tuning the parameters. We show how to choose the tuning parameters and compare our results with established econometric methods.

Bi-Objective Cost-Sensitive Machine Learning: Predicting Stock Return Direction Using Option Prices

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

We design a new framework for predicting the sign of stock return using cost-sensitive example-dependent learning where option prices contain signal about the asymmetric costs of misclassification. We show how this approach permits significant improvements in learning based on gradient boosting and logistic loss. We also provide evidence of massive economic benefit of the approach.

Handling Concept Drift In Global Time Series Forecasting

Presenter: Rakshitha Godahewa
Co-authors: Rakshitha Godahewa;Ziyi Liu;Kasun Bandara;Christoph Bergmeir

Machine learning (ML) based time series forecasting models often require and assume certain degrees of stationarity in the data when producing forecasts. However, in many real-world situations, the data distributions are not stationary and they can change over time while reducing the accuracy of the forecasting models, which in the ML literature is known as concept drift. Handling concept drift in forecasting is essential for many ML methods in use nowadays, however, the prior work only proposes methods to handle concept drift in the classification domain. To fill this gap, we explore concept drift handling methods in particular for Global Forecasting Models (GFM) which recently have gained popularity in the forecasting domain. We propose two new concept drift handling methods, namely: Error Contribution Weighting (ECW) and Gradient Descent Weighting (GDW), based on a continuous adaptive weighting concept. These methods use two forecasting models which are separately trained with the most recent series and all series, and finally, the weighted average of the forecasts provided by the two models are considered as the final forecasts. Using LightGBM as the underlying base learner, in our evaluation on three simulated datasets, the proposed models achieve significantly higher accuracy than a set of LightGBM baselines across four evaluation metrics.

Short-Term Load Forecasting With Global Models: A Comparative Analysis Of Neural Network Architectures

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

Short-term load forecasting is a critical component of managing electricity systems and ensuring that future demand can be met without compromising any part of the system. As a result, researchers have devoted significant effort in developing neural network methods for accurately estimating future load. While a large number of architectures have been proposed, the vast majority of models have been trained on data from a single location, ignoring recent advances in general-purpose forecasting that encourage the use of cross-learning. In this study, we explore a different approach to short-term load forecasting using a “global”

neural network that is trained with load data from multiple countries. We evaluate six architectures with varying degrees of complexity and compare them to models trained with country-specific data, which is the typical approach. Additionally, we investigate the use of exogenous variables. Our evaluation framework uses the actual forecasting horizon that system operators follow for day-ahead forecasts, including 24 European countries and a year of forecasts. Our results demonstrate that our approach produces significantly more accurate forecasts, with the deep fully-connected architecture we developed outperforming all other models and benchmarks.

Forecasting Large Sets Of Economic Data Using Global Models And Time Series Descriptive Information

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

Recent studies suggest that "global" models, trained using large data sets that consist of numerous time series, can result in more accurate forecasts compared to models trained on each series separately. While these models can exploit the richer information that diverse series offer, they often lack knowledge about the particular properties of the individual series, such as their category and metric unit. If such descriptive information was available in addition to historical observations, global models could become more generic and capable of accurately forecasting time series of different domains and characteristics. In this work, we evaluate the potential benefits of this notion by considering a data set of economic series, accompanied with indicative covariates that could be used for improving the specialization of global models. Our approach is compared to both standard global models that do not involve any covariates and existing approaches that indirectly employ some sort of time series classification, providing useful findings and suggestions for future research.

Pre-Trained Deep Networks Outperform True Models When Predicting Time Series Processes

Presenter: Pablo Montero Manso
Co-authors: Pablo Montero Manso;Marcer Scharth

We will present experimental results on pre-trained networks, models that forecast a new timeseries without modifying their parameters (zero-shot learning).We first train deep networks on a large set of simulated time series coming from well-known time seriesprocesses such as AR, ARIMA, State Space or Volatility models.We measure the forecasting accuracy of these networks on new time series comingfrom these processes. The pre-trained networks outperform the classic approachof estimating the true model of each time series (e.g. via maximum likelihood) and then forecasting withthe model it found. For example, a neural network that is trained on a varietyof AR(5) processes is more accurate at predicting a new AR(5) time series than fitting an AR(5) to it.We provide scaling laws in this setting, measuring the effect of the network architecture,size of the training set, process complexity and length of the series.The implications of these results go far: they hint at the potential of moreaccurate models that are simple and easy to use (no parameter tuning, no model selection), tools that are faster (orders of magnitude faster forecasting when running on hardware such as GPUs), more energy efficient (trained once, not once per series!), trivial to extend to new types of time series processes.We provide a new understanding of global models or cross-learning: as approximations of complete forecasting algorithms (estimate and predict), rather than predictive functions that are shared across series.We believe that the experimental setup will also help design new time-series specifc network architectures.

When Does Information On Forecast Variance Improve The Performance Of A Combined Forecast?

Presenter: Matthias Hartmann
Co-authors: Matthias Hartmann;Christian Conrad

In this paper, we show that the consensus forecast can be biased if some forecasts minimize an asymmetric loss function and the DGP features conditional heteroscedasticity. This result still holds if cross sectional heterogeneity in the loss function is allowed for, including the case where a share of the cross section of forecasts is produced under a symmetric (squared) loss objective. In this setting, the time-varying bias depends on the variance of the process. As a consequence, the information from the ex-ante variation of forecasts can be used to improve the predictive accuracy of the combined forecast. We consider two widely employed measures for the ex-ante forecast variance, namely the average over the variances of individual cross sectional units on the one hand and the cross sectional dispersion of point forecasts ("disagreement") on the other hand. Both statistics are shown to be informative. The average individual variance provides the largest predictive content. Forecast survey data from the Euro area and the U.S. confirm the implications of the theoretical model.

Polarized Expectations, Polarized Consumption

Presenter: Rupal Kamdar
Co-authors: Rupal Kamdar;Walker Ray

This paper argues that political polarization plays a key role in shaping the economic expectations and consumption behavior of households. Using a combination of survey and consumption data of U.S. households, we document five facts. First, household beliefs are well-described by a single factor, which behaves like sentiment. Second, at any given time there is wide dispersion in household sentiment, largely driven by political affiliation. Third, household sentiment is highly persistent, with one exception: following elections when the White House switches parties, optimistic households become pessimistic and vice versa. Fourth, the magnitude of this switching behavior has increased over time. Fifth, consumption responds differentially along party lines following changes in the White House. We show that standard theories of expectation formation struggle to simultaneously rationalize these facts.

Partisan Bias In Professional Macroeconomic Forecasts

Presenter: Jane Ryngaert
Co-authors: Jane Ryngaert;Benjamin Kay;Aeimit Lakdawala;Michael Futch

Combining forecasts from the Wall Street Journal's Economic Forecasting Survey with voter registration and political contribution data, we investigate the role of partisan bias in professional forecasting. We document that macroeconomic expectations of professional forecasters are systematically related to their affiliation with the political party in control. Democrat-affiliated forecasters have GDP growth forecasts that are 0.4 percentage points higher than Republicanswhen the Democrats control the White House (relative to when Republicans control it). There is suggestive evidence that changes in control of the House and the Senate also affect professional forecasts in a manner like control of the executive office. We discuss the implications of our results for the large literature that uses professional forecasts to study the role of deviations of expectations from the commonly used friction-less rational expectations benchmark.

Practice Makes Perfect: Learning Effects With Survey-Based Point And Density Forecasts Of Inflation

Presenter: James Mitchell
Co-authors: James Mitchell;Hana Braitsch

This paper shows how both the accuracy of, and the narrative associated with, the point and density forecasts from a well-known survey of households' inflationary expectations (the New York Fed's Survey of Consumer Expectations) depend on the tenure of the survey respondents. Both point and density forecast accuracy increase with experience. Density forecasts also become increasingly unimodal and symmetric for repeat respondents. When using expectations data to forecast inflation, our results suggest that greater weight should be given to the responses of repeat respondents, in contrast to how results are commonly communicated. But as a gauge of inflationary expectations from the population as a whole, new respondents' expectations are more representative. This is confirmed when we compare the expectations of new respondents to the SCE with other surveys of households' inflation expectations that do not allow for repeat respondents.

U.S. State Tax Revenue Forecasting Challenges: Incorporating Innovations In Bayesian Vector Autoregression (Bvar) Forecasting

Presenter: Melissa Mcshea
Co-authors: Melissa Mcshea

Given large unforeseen shocks instigating the last two recessions, forecasters must better account for such events. I consider American states to be the laboratory for exploring how tax revenue forecasting can be improved. Unlike the federal government, states cannot legally run budget deficits to generate revenues to fund their expenditures. Naturally, reliability of forecasting models is imperative. For large states that rely on model-derived forecasts, Bayesian Vector Autoregression (BVAR) models appear promising as they have been fairly effective in macroeconomic forecasting at the national level. I will assess the effectiveness of select innovations in BVAR models in forecasting the following categories – total, sales, and personal and corporate income tax revenues– for California, New York, and Virginia. The sample will cover 1985Q4 to 2022Q4. The explanatory variables will include measures of the U.S. and state economies, and demography. Real-time macroeconomic data will come from the Chicago Fed National Activity Index and the Philly Fed Coincident Index, for example. Other data will come from the FRED database, Census Bureau, Bureau of Economic Analysis, and Department of Labor. I expect to use at least six to up to as many as two dozen variables of various frequencies as explanatory variables to exploit the advantages of a larger BVAR as demonstrated by Brave, Butters, & Justiniano (2019). Also, I will consider using a suite of hyperparameter settings of the BVAR model as Brave et al. (2019) do. Because forecast presentation is what policymakers and the public alike rely on rather than knowing the nitty-gritty of model construction and output, I will produce point forecasts and the more meaningful forecast densities as do Chiu et al. (2017).I anticipate that incorporating innovations in BVAR and presentation of density forecasts will significantly enhance forecast accuracy as well as relaying the potential effects of black swan events. What I am not sure of is whether doing so compensates for other ills related to data integrity to the extent never before seen: unemployment statistics were grossly inaccurate as it was difficult for government surveyors to do their work amidst widespread business disruptions early in the pandemic (Casselman, 2020).

Global Inflation Connectedness

Presenter: Kamil Yilmaz
Co-authors: Kamil Yilmaz

This paper analyzes inflation connectedness over a large set of countries from 1970 to 2022 using the

Diebold-Yilmaz methodology. The recent surge in inflation worldwide appears to result in a substantial increase in inflation spillovers, especially among developed economies but also from developed to developing economies. We show that inflation connectedness gradually increased along with the integration of global markets through increased trade flows over time and is much higher today than it was in the late 1970s and early 1980s. The second set of regressions reveals that the inflation connectedness from one country to the other is an inverse function of the distance between the two countries, as predicted by the gravity equation. Furthermore, the pairwise inflation connectedness has a higher elasticity with respect to the source-country income than the target-country income. We also show that distance elasticity does not decrease over time, as would be predicted by increased trade integration.

Advancing Forecast Accuracy Analysis: A Partial Linear Instrumental Variable And Double Machine Learning Approach

Presenter: Christoph Schult
Co-authors: Christoph Schult;Katja Heinisch;Fabio Scaramella

This study expands upon work by Engelke et al (2019) by exploring the relationship between forecast accuracy and forecast assumptions using German data and a novel empirical approach. We employ Partial Linear Instrumental Variable (PLIV) regression models combined with Double Machine Learning (DML) methods to address issues of high-dimensional nuisance parameters and endogeneity. This innovative PLIV-DML framework enables a more complex understanding of the relationships between forecast assumptions and forecast accuracy than traditional OLS-based analysis. Our evaluation sample ranges from 1992 to 2019 and includes 1460 annual GDP forecasts and various assumptions for oil, exchange rate, and world trade. The PLIV-DML model's inherent flexibility allows us to examine two possible violations of assumptions of the model used in Engelke et al (2019): rationality of forecasters and linearity of the data generation process.This research contributes to the field of forecasting by providing a more robust and flexible analysis of forecast accuracy determinants. For instance, we contribute to the discussion regarding weak instruments and instrumental variables validity in macroeconomic models. Further, we found evidence of serious differences between OLS-based estimates, as proposed by Engelke et al. (2019), and results based on DML estimates. In particular, we report a constant underestimationof OLS estimators of the impacts of squared assumption errors of oil price and world trade on squared forecast errors of GDP. The insights gained from this study have potential implications for improving the accuracy of economic forecasts and enhancing the understanding of underlying forecasting processes.

Energy Forecasting Competitions: Past, Present, And Future

Presenter: Tao Hong
Co-authors: Tao Hong

Energy forecasting competitions date back to 1990s, of which some of the early contestants later became Nobel laureates. The popularity ramped up in 2010s through three Global Energy Forecasting Competitions and many other smaller scale competitions organized by industry and academic institutions. In this presentation, we will discuss some lessons learned from the major energy forecasting competitions in the past three decades. We will also give an outlook of the next competitions.

Bigdeal Challenge 2022: An Introduction

Presenter: Shreyashi Shukla
Co-authors: Shreyashi Shukla;Tao Hong

The BigDEAL Challenge 2022 attracted hundreds of participants worldwide, who contributed many

novel ideas to the under-researched problem of peak load forecasting. This paper introduces the three tracks of the challenge, Peak Load Magnitude, Peak Timing, and Peak Shape forecasting, with details on the aspects of the problem, the data, and a summary of the methods used by selected top entries. We also provide our reflections on the challenges and motives of hosting this competition. The framework of the competition and participation statistics are presented. Furthermore, the competition dataset and problem formulation are thoroughly described.

Using Conditional Invertible Neural Networks To Perform Mid-Term Peak Load Forecasting

Presenter: Benedikt Heidrich
Co-authors: Benedikt Heidrich;Oliver Neumann;Matthias Hertel;Veit Hagenmeyer;Ralf Mikut

Electricity time series forecasting is essential for the future electrical system since it enables the implementation of applications required to increase the share of renewable energy sources. Depending on the applications, other properties than the overall accuracy of the forecast may be important, such as the daily peak magnitude, the daily peak position, or the shape of the load curve around the daily peak. Thus, the BigDeal Challenge dealt with the prediction of these properties by introducing corresponding metrics. The challenge asked to perform mid-term forecasts (one to three months). Consequently, models that require lag information from previous days to forecast the electrical load for the next days are not suitable. Moreover, since weather forecasts and calendar information for each future time step is provided, this task is related to conditional time series generation tasks. Thus, we use a conditional Invertible Neural Network (cINN) since it shows promising performance on time series generation tasks, their bijectivity avoids problems such as mode collapse, and we experience that cINNs are easy to train. We train our cINN to provide a bijective mapping from the space of the time series to a normal distribution, using the available weather and calendar information as conditioning information. The bijective mapping allows to sample from the normal distribution to get forecasts of the time series. An advantage of using a generative model is that it is designed to create new time series while considering their specific properties. Consequently, the cINN should perform well in tasks where properties such as the shape are important.A disadvantage of the cINN is its fixed input and output size. Thus, since we create day-ahead forecasts for each time step, this results in overlapping forecasts. To cope with these overlapping samples and to increase the robustness, we temporally align the forecast samples and take the median for each time step.Our talk at ISF23 will present our solution for the BigDeal challenge in depth, how our implementation achieved first places in the peak shape and peak positioning track, and point out further ideas to use generative models in time series forecasting.

Daily And Intraday Application Of Various Architectures Of The Lstm Model In Algorithmic Investment Strategies On Bitcoin And The SandP 500 Index

Presenter: Robert Ślepaczuk
Co-authors: Robert Slepaczuk;Katarzyna Kryńska

This thesis investigates the use of various architectures of the LSTM model in algorithmic investment strategies. LSTM models are used to generate buy/sell signals, with previous levels of Bitcoin price and the S&P 500 Index value as inputs. Four approaches are tested: two are regression problems (price level prediction) and the other two are classification problems (prediction of price direction). All approaches are applied to daily, hourly, and 15-minute data and are using a walk-forward optimization procedure with numerous IS and OOS periods. The out-of-sample period for the S&P 500 Index is from February 6, 2014 to August 26, 2022, and for Bitcoin it is from February 1, 2014 to August 26, 2022. We discover that classification techniques beat regression methods on average, and that intraday models perform much better in case of classification approach, while daily ones produce outperforming results in case of regression methods. The research cover s3 types of ensemble models: through frequencies, assets,

and the combination of both of them. We come to the conclusion that the ensembling of models positively affects their performance only on the condition of specific characteristics of the component parts. Finally, a sensitivity analysis is performed to determine how changes in the main hyperparameters of the LSTM model affect strategy performance. It reveals that we can distinguish the specific hyperparameters which can increase the performance of LSTM model for all tested cases.

Optimism, Pessimism, And Future Stock Returns: Enhancing The Predictive Power Of Short Interest

Presenter: Mark Schneider
Co-authors: Mark Schneider;Soroush Ghazi;Jack Strauss

Short interest is one of the strongest known predictors of future stock returns. Short interest is commonly interpreted as a measure of informed trader pessimism. We investigate if a theory-based measure of market optimism further enhances the predictive power of short interest. We find support for this hypothesis in both in-sample and out-of-sample tests. The effect is strongest at the quarterly horizon where the out-of-sample R-squared increases from 4% with short interest to over 7% for short interest combined with optimism. The largest gains in predictability over time occur during the great recession and during the COVID pandemic. The largest gains in predictability across industries occur in high risk industries, especially in the high tech sector. Our analysis sheds new light on the sources of the predictive power of short interest and on the theoretical explanation for its return predictability.

Value At Risk Forecasting For Stock Market: Based On Textual Information And A Hybrid Garch-Lstm-Based Model

Presenter: Yangfan Cao
Co-authors: Yangfan Cao;Xi Luo;Choo Wei Chong

Deep learning models and natural language processing (NLP) have been receiving considerable attention recently since they have been successfully used in diverse fields. By using text mining techniques and deep learning models, this work aims to build a more accurate VaR model in three steps. Firstly, to properly describe the risks in the stock market, we use a kind of natural language processing method to extract the dynamic risk and sentiment factors from online stock news headlines and assign them as independent variables. Secondly, by including the forecasts of the three GARCH-type models that are based on distinct economic characteristics, we examine whether the proposed long and short-term memory (LSTM) and convolutional neural network (CNN-LSTM) model significantly improves the accuracy of VaR. Finally, this study combines the LSTM-based model with the estimates from one to three GARCH-type models to estimate VaR. In this study, the S&P500 historical data and the corresponding stock news headline data from 2012 to 2023 are employed to verify our assumption empirically. The findings show that the VaR model produces more accurate forecasts with the use of textual elements. Besides, VaR estimation benefits from the use of LSTM models. Furthermore, the results highlight that a hybrid LSTM model with input from forecasts of various GARCH types has a superior VaR estimation capability.

Unlocking The Black Box Of Sentiment And Cryptocurrency: What, Which, Why, When And How?

Presenter: Jack Strauss
Co-authors: Jack Strauss;Erik Erik Mekelburg;Donyetta Bennett;T.h. Williams

We evaluate the impact of a large set of daily sentiment measures for predicting Ethereum (ETH) returns using Machine Learning (ML) methods. We open the black box of ETH predictability to identify

5 W's: What, Which, When, Why, and hoW. What ML methods work best? Which variables robustly predict ETH returns? When and Why does predictability occur? And how can we improve predictability? We use sentiment measures from Refinitiv MarketPsych Analytics (RMA), formerly Thomson Reuters, a novel database that scans over four thousand news and social media sites. We show that a single sentiment index or a few principal components do not adequately summarize the broad spectrum of different sentiment measures. In our sample, twenty-seven components capture only 50% of the variance. The relationships further between sentiment variables evolve over time. Results document that expanding sentiment measures beyond a simple sentiment index is relevant for ETH predictability. Our paper models periods of high absolute confidence measured by increases in the absolute value of positive and negative sentiment, and shows they have substantially lower mean-squared forecast errors than periods of low sentiment. We find that periods of high confidence, similar to bad days, lead to substantially higher out-of-sample R2. This occurs because periods of high positive and negative sentiment periods generate autocorrelation and predictability.We introduce an ML ensemble procedure that exponentially weights forecasts from traditional ML methods based on recent MSFE criteria. By discounting past model performance, our ensemble procedure accommodates time variation in model selection and generates investment gains and significant out-of-sample predictability. Lastly, we highlight the growing importance of transaction costs and build them into our investment model.

What Is The Value Of Congruous Forecasts Across Time?

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

Forecasts of future demand are necessary for inventory management. Typically, the focus is on producing accurate forecasts, which are desirable in the statistical sense. On the other hand, the limited work that looks jointly at forecasting and inventory control has identified low out-of-sample bias to be more important than accuracy. Shrinkage estimators and forecast combination shift the attention from in-sample fit to better generalization in the future. These result in less volatile forecasts, with typically better predictive distributions, specifically at quantiles of interest, and less out-of-sample bias. Moreover, companies often prefer forecasts that may be suboptimal in the statistical sense, but change less across time periods, putting less strain in production planning and inventory management, even though this may harm accuracy, attempting to minimize total costs. Arguably this increased congruous forecasts across time periods points to a different objective than accuracy. This is also reflected in recent views on forecast evaluation, where metrics closer to the relevant decision making are seen as desirable, albeit difficult to operationalize, and has been speculated to relate to the trustworthiness of forecasts.In this work we investigate the impact of increased congruous forecasts across time periods in terms of forecasting, and supported decisions, such as inventory management. We do this by employing a variety of approaches that can decrease the volatility of forecasts across origins: shrinkage and M-estimators, and forecast combination. We simulate the inventory position of a company, investigating the connection between the incurred costs, forecast congruity, bias, and accuracy. Our objective is to characterize the trade-off, to support better specification of forecasting models in the application context, while potentially avoiding relying on difficult to operationalize performance metrics.

Demand Forecasting And Inventory Management With Leading Indicators

Presenter: Yves R. Sagaert
Co-authors: Yves R. Sagaert;Nikolaos Kourentzes

Accurate forecasts are central to effective supply chain management. To generate accurate forecasts macro-level leading indicators can be useful, particularly in highly volatile market conditions, where univariate forecasting methods may not perform well. However, including such explanatory variables may be difficult at the product level, where noise in the demand signal and other idiosyncratic effects, such as

promotions or stockouts of products, can mask the effect of macroeconomic variables. Instead, one can model this information at company or product group levels, where many of the product-specific effects are attenuated, with empirical evidence of the benefits of forecasting accuracy. To be able to transmit this information at the product level, and therefore aid inventory management, we use the implied hierarchical structure of the supply chain. At the aggregate levels of the hierarchy, we generate forecasts with models that rely on leading indicators, while at the more disaggregate levels, we make use of univariate forecasts. An additional complication is that different leading indicators may be relevant for forecasts of different horizons, and therefore, over the relevant lead times for inventory management, forecasts from multiple models need to be generated and hierarchically reconciled. Our methodology addresses that, as well as the estimation of probabilistic coherent forecasts for the disaggregate product level demand that is needed for inventory management. We investigate the level of aggregation that leading indicators become useful and demonstrate the efficacy of our methodology both in terms of forecast accuracy and inventory management.

Unfulfilled Demand For Promotional Items In Online Retail: Investigating The Connection Between Backroom Inventories And Retail Shelves Using Efficient Forecasting Techniques

Presenter: Amila Thibbotuwawa
Co-authors: Amila Thibbotuwawa;Achala Hasini Perera;H. Niles Perera;Priyanga Dilini Talagala

Online retail has grown rapidly around the world, buoyed further by the COVID-19 lockdowns. Our study investigates the unfulfilled demand for promotional items. We primarily aimed to examine the connection between backroom inventories and retail shelves using efficient forecasting for stock-keeping units that are subjected to a price-related promotion. Our research considers retail sales data encompassing 5 months at a leading supermarket chain in South Asia. Our analysis indicates that around 12 percent of the stock keeping units are unfulfilled while they are being promoted. This implies the failure to accurately forecast demand during periods where stock-keeping units are promoted with a price-related discount. This is despite the fact that the Phi coefficient and the Pearson correlation coefficient suggest a weak correlation between retail inventory and price-based promotions. This indicates retail operator has not prepared sufficient inventory for price-based promotions. The results further indicate the importance of connecting retail inventory into forecasting to ensure a more holistic perspective in demand planning. Further, inspired by the literature, we considered Deep Learning techniques to improve forecasting to this end. This study introduces a novel model based on two different algorithms in Deep Learning: the Multi-Layer Perceptron (MLP) and the Long Short-Term Memory Network (LSTM). Thus, this research contributes to the body of knowledge by addressing the gap on inventories as a driver for improving forecasts and customer-oriented research in product promotions. The practical nature of our research, which was conducted at a supermarket chain and involved investigating messy data in online retail operations, indicates opportunities for applying such models in the realworld. By improving forecasting accuracy, retailers can reduce the likelihood of unfulfilled demand for promotional items, which can lead to increased customer satisfaction and sales. In conclusion, our study sheds light on the importance of inventory management for accurate demand forecasting for price-based promotions in online retail. Leveraging Deep Learning techniques can improve forecasting accuracy and contribute to more efficient and effective demand planning.

Inventory Control For Periodic Intermittent Demand

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

Inventory control of many stock keeping units is challenging because demand is intermittent. Intermittent demand patterns are not well modeled by a single distribution of demand per period. It is common to model the time between demand occurrences and the size of demand occurrences separately. Such models implicitly assume that times between demand occurrences are Markovian. Data from practice,

however, indicate that the time between demand events is often not Markovian but—contrary to implicit model assumptions— displays periodicity. Consequently, the time since the last demand occurrence is an important predictor for future demand. We use the discrete compound renewal process to model such periodic intermittent demand. In a periodic review inventory system, we show that the optimal inventory policy is a state-dependent base-stock policy, where the order-up-to-levels are non-decreasing in the time since the last demand, regardless of the nature of the distribution of time between demand occurrences. 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.

Business Decision-Making: How Our Biology Leads Us Astray

Presenter: NA
Co-authors: John Burkhardt

Forecasting and modern business decision-making in general draw heavily from classical and neoclassical economic traditions. While this provides a large toolbox and allows for straightforward models, there are notable shortcomings in this philosophy. Advances in recent decades in neuroscience and psychology demonstrate that the human brain is not built for business decisions, and in fact cannot be meaningfully optimized towards such. Human decision-making is driven by imperatives not considered in traditional economic thinking; this leads to systematic errors and gaps in most forms of quantitative decision-making. Given this limitation, how can we adapt business decision-making to accommodate the biological hardware we possess?This session will explore:-Behavior, judgements, and decisions in an evolutionary context-How biology interacts with the use and utility of business models-Approaches to mitigate naturally-occurring cognitive bias

Mitigating Biases To Improve Forecast Accuracy

Presenter: Jeff Baker
Co-authors: Jeff Baker

Demand distortion impacts supply chain profitability. There are significant costs associated with either over-forecasting or under-forecasting. Unfortunately biased forecast overrides are the number one distorter of the demand signal.This session talk will discuss forecast bias sources, and strategies to mitigate them. Particular emphasis will be placed on algorithm aversion, and its impact on non-technical participants in the consensus forecasting process.

Your Bias Is Showing: Understanding Cognitive Bias In Practice

Presenter: Jonathon Karelse
Co-authors: Jonathon Karelse

Decades of research has shown conclusively (Goodwin, Hyndman) that effectively integrating business intelligence and judgement into statistical forecasts can be highly effective, and in times of exceptional uncertainty, vital. We also know from the work of Behavioural Economics pioneers like Kahneman, Tversky and Thaler that the human capacity for estimation and judgement under uncertainty is extremely limited. In times of crisis or extreme change, historic models offer much less utility, so human inputs are more critical than ever. How do we get the benefit of human input into forecasts without undermining them with the inherent biases and heuristics that come along with them?This session will explore:- Prevalence of cognitive biases in practice, with particular focus on forecasting and demand planning- Mitigation strategies to help maximize the net benefit of human judgments

Demand Forecast At Wayfair - An Overview Of Our Pipeline

Presenter: I-Chen Lee
Co-authors: I-Chen Lee

The demand forecasting team at Wayfair aims to generate accurate and timely forecasts in order to enable efficient inventory planning and positioning. We outline our forecasting framework, which is based on top-down / bottom-up hierarchical approaches and their ensembles. We use econometric and machine learning models to capture time series and cross-sectional signals over millions of items, individually or at predefined aggregation levels. We augment our point forecasts with forecast distributions to account for uncertainty and to assist with safety stock planning. We also forecast sales regional shares and estimate sales lifts if we position items closer to our customers for faster delivery. Our forecasts are further adjusted by downstream business teams to account for business intelligence and decisions for selected items and for short-term horizons. We refresh our pipeline on a regular cadence in order to better serve our customers and suppliers across global markets.

Forecasting At Thrasio

Presenter: Philip Brooks
Co-authors: Philip Brooks

Thrasio is a next-generation consumer goods company reimagining how the world's most-loved products become accessible to everyone. Forecasting is at the center of this mission serving as an integral connection point between demand and supply systems, ensuring the right product is available at the right place, at the right price. In this talk we'll walk through Thrasio's end-to-end forecasting system with an emphasis on how we create feedback loops with upstream demand lever systems, integrate human feedback in a structured and traceable format, and flexibly estimate our uncertainty throughout to power probabilistic supply chain management and optimization. We'll address not only the technical challenges associated with building integrated systems to operate a large consumer goods company, but also the organizational challenges that arise throughout and how we've structured ownership of key systems to incentivize collaboration across a growing team of scientists and engineers.

Customer Demand Forecasting In Retail

Presenter: Arash Sangari
Co-authors: Arash Sangari

One of the key signals to many planning and execution engines in any retailer is the customer demand forecast. Unlike many other time-series forecasting problems, the target signal, i.e., customer demand, is not always fully observable. Due to several constraints in the retail supply chain, the sales may not be reflecting the true customer demand. In addition, in many business use-cases it is desired to show how much different decisions may interact with the customer demand and influence it. This talk will cover the broader impact of demand forecasting for retailers and the challenges and opportunities that are unique to the retail customer demand forecasting.

A Novel Non-Parametric Time Series Structural Model

Presenter: André Ramos
Co-authors: André Ramos;Davi Valladão

Structural time series models are widely used for forecasting and extracting non-observable components such as level, slope, and seasonality. However, these models rely on Kalman filtering and non-linear

optimization, which often result in estimates that are highly dependent on user-dependent initial guesses, leading to suboptimal forecasts.In this work, we propose a novel non-parametric structural time series model that leverages the power of machine and statistical learning to provide a more robust and reliable estimation process. We take an innovative approach by reformulating these models as high-dimensional regressions and introduce a regularization scheme to efficiently estimate the coefficients, extract non-observable components, and generate out-of-sample forecasts.Our proposed model also incorporates the ability to detect outliers and can be used to initialize a traditional Gaussian structural model. We validate our model using 10000 monthly series from the M4 competition, and our empirical results demonstrate a superior forecasting performance when compared to a traditional Gaussian structural model.Overall, our non-parametric structural time series model offers a promising alternative to traditional models providing a more accurate and reliable approach for forecasting and extracting non-observable components.

An Efficient Algorithm For Approximating Arma Model Fitting In Large-Scale Time Series Data

Presenter: Ali Eshragh
Co-authors: Ali Eshragh;Luke Yerbury;Fred Roosta;Asef Nazari;Michael Mahoney

In the current age of big data, non-uniform sampling schemes for matrix compression, like leverage score sampling, provide superior alternatives to naive computations, with high-quality numerical implementations and robust theoretical guarantees. This presentation introduces the Sequential Approximate Leverage Score Algorithm (SALSA), which employs randomized numerical linear algebra (RandNLA) techniques to estimate the leverage scores of any general matrix. The talk demonstrates that the approximations obtained through this algorithm are bounded favorably with high probability. The SALSA algorithm is then utilized to efficiently estimate leverage scores within the LSARMA algorithm, facilitating the fitting of ARMA models in the context of large-scale time series data. The effectiveness of these two innovative algorithms is demonstrated using large-scale simulations and real-world data.

A Similarity-Based Approach To Covariance Forecasting

Presenter: Mark Jennings
Co-authors: Mark Jennings;Chao Zhang;Mihai Cucuringu;Alvaro Cartea

Forecasting the realised covariance matrix of a multidimensional time series is a ubiquitous problem in various financial domains, including risk management, asset pricing, and portfolio optimisation. In this paper, we introduce a novel framework which leverages recurrent latent structures inherent in the data to tackle the problem. Specifically, our framework calculates similarity scores between the test input and each individual training input using their recent histories. We use these scores to filter the training data by relevance, and consequently, training data that do not share the observed dynamics of the test input are excluded from the regression process. This reduces the complexity of the forecasting task. We then produce a forecast based only on the relevant training data. The estimators we use are a modified version of k-nearest neighbours regression which compensates for unbalanced distributions of nearest neighbours around the test input (an approach we call iterative assembly) and heterogeneous autoregression. To improve the accuracy and robustness of our forecasts, we furthermore propose a dynamic ensembling scheme inspired by empirical similarity approaches ([Golosnoy et al., 2014]). The scheme combines our forecasts with forecasts produced by existing models and uses information about the recent forecasting performances of different models to generate a weighted average forecast.We demonstrate that our framework produces computationally efficient, interpretable, and accurate forecasts of the realised covariances of US equity returns and outperforms existing widely-used benchmark models. The framework has several desirable properties that make it well-suited to this problem. It is scalable to high-dimensional settings. It functions well in scenarios where the number of available training observations is limited because it does not rely on complex, highly parameterised models. Furthermore, the ensemble scheme adjusts to rapidly changing

market conditions by giving less weight to models that are unsuitable for the current market dynamics, which minimises the impact of model specifications. This property leads to improvements in robustness against turbulent conditions compared to popular alternative models. We also evaluate the economic value of our forecasts by applying them to the task of portfolio optimisation, and show that our framework generates a higher Sharpe ratio than those of competing models.

A Mixed-Frequency Var Model With Closed Form Solution

Presenter: Heiner Mikosch
Co-authors: Heiner Mikosch;Maurizio Daniele;Stefan Neuwirth

We develop a mixed-frequency vector autoregression (VAR) model using a stacked vector approach. We then show how to transform the model from its stacked vector form into a closed form in which the model can be estimated analytically via multivariate least squares (MLS). Also, we develop a Bayesian Minnesota prior for analytical shrinkage estimation of the model. Our mixed-frequency VAR does not involve modeling of latent variables and falls in the class of observation-driven models. It contrasts with previous literature that proposed paramater-driven mixed-frequency VARs using a latent-variable state space framework. Monte Carlo simulations yield that both the MLS and the Bayesian estimators of our mixed-frequency VAR are consistent and fast. Which of the two estimators to prefer depends on the size of the parameter space and the available observation size. In an empirical out-of-sample forecasting exercise with quarterly, monthly, and weekly macroeconomic and financial data we find our mixed-frequency VAR to outperform a standard quarterly-frequency VAR.

Carpe Diem: Can Daily Oil Prices Improve Model-Based Forecasts Of The Real Price Of Crude Oil?

Presenter: Stephen Snudden
Co-authors: Stephen Snudden;Reinhard Ellwanger;Amor Aniss Benmoussa

The standard approach in the literature is to compute model-based forecasts of the real price of crude oil with monthly average prices. We show how existing model-based forecasts approaches can be extended by incorporating the underlying daily oil prices. For both univariate and multivariate methods, alternative disaggregated methods are compared, including bottom-up, Period-End-Price Sampling (PEPS), and Mixed Data Sampling (MIDAS). We find that using unaveraged oil prices, especially end-of-month prices, yields large gains across a variety of forecast approaches. In some cases, forecast error is more than halved compared to existing specifications. Contrary to models estimated with monthly average prices, models that utilize the underlying daily prices can outperform the random walk forecast of daily prices at short horizons. The analysis informs how to efficiently construct forecasts for other series expressed in real terms to avoid information loss from temporal aggregation.

An Analysis Of Co2 Emissions In Spain Using Many Macroeconomic Predictors

Presenter: Pilar Poncela
Co-authors: Pilar Poncela;Esther Ruiz;Aranzazu De Juan

The aim is to analyze how the (macro) economic activity in Spain affects CO2 emissions. To do so, we build an extensive database of macroeconomic predictors and proceed in two ways: on the one hand, we use variable selection techniques to find a small set of economic predictors with a higher correlation with CO2 emissions and analyze the resulting relations. On the other hand, we build a dynamic factor model with a large dataset of macroeconomic predictors. The extracted common factors capture macroeconomic conditions in Spain. In the second step, we check the relation of the common factors with CO2 emissions.

We also estimate joint models that incorporate both the common factors and the individual predictors as explanatory variables of CO2 emissions. Our results indicate that private consumption and maritime transport are the most significant variables in order to explain C02 emissions. Once these individual predictors are considered, the information contained in the macroeconomic data set only has negligible explanatory power for emissions. Finally, we evaluate the forecasting performance of the different models.

An uncertainty - resilient multi-variate hybrid wavelet VAR neural forecaster for macroeconomic policy variables

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

Forecast of real macroeconomic statistics for emerging economies like Brazil, Russia, India, and China has been at the forefront of a wide range of policy debates among central bankers. Questions like, can movements or dynamic shifts in some macro-economic policy variables hold valuable information about future movements in inflation, have been traditionally answered using a specific set of mathematical tools, which can potentially model the interplay of various other macro-economic factors like oil price shocks, movements in real GDP, import prices and exchange rates to determine and answer questions on structural volatilities in inflation numbers across geographies. This study proposes a novel multi-variate wavelet–based vector auto-regressive neural network algorithm (W-VARNN) that can factor in various endogenous factors and is capable of producing forecasts in a multi-variate (for all the endogenous factors) set up for CPI Inflation and other associated endogenous policy variables like real GDP, exchange rates, etc. for the BRIC countries. The algorithmic framework is reasonably flexible and scalable and can accommodate exogenous factors like economic policy uncertainties and geo-political risk while producing forecasts for all the endogenous variables. The proposal leverages various forms of measure to study the causal association between variables (which is an important pre-requisite for any traditional VAR model) and gain a deeper understanding of the underlying structure of endogeneity among macroeconomic policy variables. In this study, we introduce various other measures like the non-linear granger – causality test, convergent cross mapping, etc. to determine the directional causality. In this algorithmic framework, we use wavelet transformation to decompose the endogenous variables into wavelet details and smooth components and these components along with the set of exogenous variables (uncertainty indices) are then modeled using an auto-regressive neural network to generate h-period forward forecasts for the inflation series. The proposal's performance is compared with various other SOTA VAR architectures like VAR, DeepVAR, VARMAX, etc. Finally, the study introduces conformalized prediction intervals for the generated forecast to study the robustness and to construct the "uncertainty bounds" around the forecast values.

Systematizing Macroframework Forecasting: High-Dimensional Conditional Forecasting With Accounting Identities

Presenter: Sakai Ando
Co-authors: Sakai Ando;Taehoon Kim

Forecasting multiple macroeconomic variables with accounting identity restrictions, also known as macroframework, is useful for presenting an internally consistent economic narrative. For example, a forecast of GDP alone does not tell us why GDP fluctuates, but a forecast of a macroframework consisting of the GDP expenditure components (i.e., consumption, investment, government spending, and net exports), and GDP as their sum, can pinpoint the underlying drivers of the predicted aggregate GDP growth. Because of such benefits, macroframework forecasting is widely used in policy institutions, including central banks and international financial institutions.Macroframework forecasting, however, is challenging. A macroframework can contain hundreds of variables, but forecasters often have information about only a subset of these (known variables for short). As a result, forecasters often take the known variables as given and use rules of thumb and ad-hoc adjustments to extend the information about known variables

to the unknown variables. Despite the wide usage of macroframework forecasting in policy institutions, a systematic way to forecast the unknown variables from the set of known variables has not yet been developed. Macroframework forecasting, thus, remains resource-intensive and involves numerous ad-hoc adjustments.We propose a novel 2-step method to forecast unknown variables conditional on known variables, which reflects historical correlations and satisfies accounting identities. The method offers (1) the flexibility to incorporate available information in known variables and (2) the convenience to automate the forecasting of unknown variables. The method integrates two strands of literature, namely high-dimensional conditional forecasting and forecast reconciliation of hierarchical time series. Applying our method to forecast GDP subcomponents in an advanced and emerging market country, we show that it improves upon alternative forecasting techniques.

Step By Step - A Quarterly Evaluation Of Eu Commissions' Gdp Forecasts

Presenter: Katja Heinisch
Co-authors: Katja Heinisch

Annual growth forecasts by the European Commission are important figures for policy-making and provide a benchmark for many forecasters. However, they are usually based on quarterly estimates, which do not get much attention and are hardly known. Therefore, this paper provides a detailed analysis of multi-period ahead quarterly GDP growth forecasts for the EU, euro area, and several EU member states with respect to first-release and current-release data. Forecast revisions and forecast errors are analyzed and the results show that the forecasts are not systematically biased. However, for several member states a significant overestimation of short-time horizons is identified. The highest performance is not achieved for the current quarter for all countries, although a high forecast revision occurs in the last step (from one-quarter-ahead forecast to the current quarter). Furthermore, the final forecast revision in the current quarter is generally downward biased for almost all countries. Overall, the differences in mean forecast errors are minor when using real-time data or pseudo-real-time data. The forecast performance also varies across countries, with smaller countries and Central and Eastern European countries (CEEC) having larger forecast errors. The paper provides evidence that there is still room for improvement in forecasting techniques both for nowcasts but also forecasts up to 8 quarters ahead.

Market Ambiguity Attitude And The Risk-Return Tradeoff

Presenter: Mark Schneider
Co-authors: Soroush Ghazi;Mark Schneider;Jack Strauss

The risk-return tradeoff for the aggregate stock market predicts that periods of high market volatility are followed by periods with higher average market returns. Despite being viewed as a fundamental law of finance, the risk-return tradeoff has been difficult to identify in the data, with some studies finding no relation or even a negative relation between market volatility and future excess market returns. The uncertainty-return tradeoff for the aggregate stock market, in which a high variance risk premium is followed by higher average returns has also been recently challenged. Motivated by a representative agent asset pricing model in the presence of Knightian uncertainty, we investigate if market ambiguity attitude can help identify the risk-return tradeoff and the uncertainty-return tradeoff. We obtain theoretical predictions that market ambiguity attitude moderates the risk-return tradeoff and that it has complementary return predictability to the variance risk premium. We introduce a methodology for measuring market ambiguity attitude and then test its relevance for explaining the risk-return tradeoff in in-sample and out-of-sample tests. We find support for the theoretical predictions, both in-sample and out-of-sample. In univariate regressions the relationship between the equity premium and market volatility is unstable, with the slope coefficient changing signs across the two halves of the sample and an out-of-sample R-squared close to zero. Including the interaction term uncovers a positive risk-return relationship at both the monthly and quarterly horizons with an out-of-sample R-squared above 4% at both horizons. Including the interaction

term also produces more stable coefficient estimates for the risk-return tradeoff. Similarly strong results hold for the variance risk premium when combined with market ambiguity attitude. The results are not subsumed by standard return predictors and they are not explained by investor sentiment. Our results highlight the role of market ambiguity attitude in helping to explain time variation in the risk-return tradeoff and the uncertainty-return tradeoff for the aggregate stock market.

Bivariate Long Memories in Stock Returns: An Empirical Investigation of U.S. Lumber Supply Chain Stocks

Presenter: Charles Mutigwe
Co-authors: Rory Cooper;Johnson Kinyua;Anil Gulati; Michael Piggi

In this work we present a prediction vector model based on bivariate Granger causality tests. Using this model, the strength of a bivariate Granger causality relationship over a period is characterized. The notion of structural causal relationship is introduced, these are long memory bivariate Granger causal relationships. As a case study, the prediction vector model is applied to the daily close prices of eight U.S. lumber supply chain related stocks and four economic and financial indicators over a 15-year period. The model reveals that the International Paper stock price is a good predictor of other lumber supply chain-related stocks. In the pre-COVID-19 period the model produced thirteen structural causal relationships. By June 2022, the model produced only one structural causal relationship. This suggests that the economic disruption due to the COVID-19 pandemic has caused structural breaks in many long memory relationships. Additionally, the existence of bivariate relationships with structural causalities lasting as long as 12 to 15 years suggests the existence of some long memory inefficiencies in the pricing of lumber supply chain related stocks.

A Brief Re-Examination To Stock Price Forecasting, Tests Of Market Efficiency And Predictable Profitability

Presenter: John Guerard
Co-authors: John Guerard;Dimitrios Thomakos;Foteini Kyriazi;Konstantinos Mamais

Intelligent people have always been interested in modeling and forecasting stock prices and returns. In the United States, people have talked about the Dow Jones Industrial Average, DJIA, since the creation of the Dow in 1896. We estimate an AR(1) model of DJIA from 1896 – June 2022. We specifically address outliers in the time series with the OxMetrics software, modeling trend saturation variables. We also use rolling windows analysis (RWA) with robust regression of AR(1) models and transfer functions models with the volatility of the Dow and the Leading Economic Indicators, LEI. We report statistically significant forecast error reductions with robust time series and the RWA framework. We show that successful forecasting can be used in generating excess trading returns and expose which kinds of forecasting models are best suited for the purpose of trading.

Key Drivers Of Scenario Planning In Science And Technology Public Scenarios

Presenter: Changhyun Park
Co-authors: Changhyun Park

Scenario planning has been suggested to be a useful foresight tool for visualizing the future in detail and resolving future uncertainty, but less research attention has been paid to science and technology (S&T) scenarios in the public domain. This study aims to study key drivers of S&T scenarios in the public domain, and the methodology is based on two rounds of the Delphi workshop. This study identified three drivers (impact of policy, purpose of policy, participants of policy), one driver (timeline of scenario), and one driver

(technological transformation by technological innovation) at each dimension of policy, time, and technology, respectively. Further, a framework of S&T public scenarios is proposed with the combination of key drivers. This study theoretically contributes to the understanding of what are the key drivers of S&T scenarios and how S&T scenarios can be constructed in the public domain. This study also has managerial implications for foresight experts or policymakers developing future scenarios of S&T public scenarios.

Predicting Food Insecurity In Africa

Presenter: Jade Preston
Co-authors: Jade Preston;William Basener

Food insecurity in Africa is one of the major global-scale humanitarian disasters. The United Nations (UN) World Health Organization estimates that 37 million people in the horn of Africa are in acute hunger and describe the current situation as one of the worst hunger crises in the last 70 years. In this talk we present a machine learning prediction using Neural Networks, Gaussian process, and Bayesian model averaging to predict the probability and quantity of food insecurity for each country in Africa along with proposed locations for new transshipment nodes based on the current food supply chain. Our predictions are based on demographic, climate, environmental, financial, supply chain disruption, and remote sensing crop yield information, trained on historical food insecurity data. Proposed locations for additional transshipment locations is designed to optimize efficiency of the locations within the current supply chain. Results are shared publicly with a HTML interactive GIS.

Beyond Numbers: Understanding And Forecasting Financial Aid Requirements Of Countries

Presenter: Imran Arif
Co-authors: Imran Arif

Financial aid is a vital funding source for many countries, especially developing countries facing economic challenges. To ensure efficient and effective allocation of resources, policymakers and aid organizations must accurately forecast a country's financial aid requirements. Accurate forecasting enables policymakers and aid organizations to plan and mobilize resources proactively, saving lives, reducing poverty, and promoting economic development. Additionally, precise forecasting ensures that aid is targeted to countries with the greatest need and prioritized to marginalized communities within a country. Forecasting can also assist policymakers in planning for contingencies such as economic downturns, natural disasters, or political crises, ensuring that aid is delivered promptly and effectively. However, forecasting financial aid requirements can be challenging due to the complex and multifaceted nature of contributing factors to financial aid needs. We test several forecasting methods to address this challenge, including a regression model that considers various macroeconomic indicators. We use a dataset of historical financial aid requirements from 1960 to 2020 for 150 countries to forecast aid requirements. Our findings help policymakers and aid organizations allocate resources effectively and efficiently.

A Heuristic For Combining Correlated Experts When There Is Little Data

Presenter: David Soule
Co-authors: David Soule;Yael Grushka-Cockayne;Jason Merrick

It is intuitive and theoretically sound to combine experts' forecasts based on their proven skills, while

accounting for correlation among their forecast submissions. Simpler combination methods, however, that

assume independence of forecasts or equal skill, have been found to be empirically robust, in particular, in settings in which there is little historical data available for assessing experts' skill. One explanation for the robust performance by simple methods is that empirical estimation of skill and of correlations introduces error, leading to worse aggregated forecasts than simpler alternatives. We offer a heuristic that accounts for skill and reduces estimation error by utilizing a common correlation factor. Our theoretical results present an

optimal form for this common correlation and we offer Bayesian estimators that can be used in practice. The common correlation heuristic is shown to outperform alternative combination methods on macroeconomic and experimental forecasting where there is limited historical data.

Improving The Wisdom Of A Crowd

Presenter: Joseph Rilling
Co-authors: Joseph Rilling

Forecasting often uses historical data to predict future outcomes. However, there are many situations without easily applicable historical data. In this case, a decision maker will often turn to a panel of judges, and seek to utilize the Wisdom of the Crowds. By combining forecasts from a group of judges, the decision maker hopes to combine the judges' knowledge produce a superior final forecast. However, correlated errors caused by shared information between the judges can cause shared information bias, which does not diminish as sample size increases. Also, skilled experts with unique information can be overshadowed by less informed judges. Recently, researchers have developed several methods to combat shared information bias, and give experts more influence in the final prediction. Three notable methods are connected because they all ask the judges to provide their forecast, and then additionally provide an estimate of the average forecast of their peers. These three methods use the additional information from peer prediction to outperform the simple mean of the individual forecasts. The Surprisingly Popular method (Prelec, Seung, & McCoy 2017) aggregates forecasts where the outcome of interest is categorical. Pivoting (Palley & Soll 2019) and Knowledge Weighting (Palley & Satopaa 2023) can aggregate points forecasts in the continuous case. I propose a novel method that uses the same inputs as the three aforementioned aggregation techniques, and connects the intuition from the Surprisingly Popular method to the case of continuous outcomes. A simulation study demonstrates the stability of my method, and an application to data from twelve real world experiments shows the proposed method outperforms both naïve aggregation approaches and more advanced competitors. Seven of these twelve applied experiments were originally conducted by Palley, Soll, and Satopaa, while the five remaining experiments were conducted by Martinie et. al (2020). I use their data which is made available in the metaggR R package.

Sas Energy Forecasting Cloud: Making Forecasts More Accessible

Presenter: Arnie De Castro
Co-authors: Arnie De Castro

Accurate and timely energy forecasting is critical for utilities to make operational and trading decisions. However, smaller utilities such as munis, coops, and smart cities often lack the resources to maintain the necessary analytics infrastructure for generating forecasts. To address this issue, we have developed an energy forecasting solution with low-code and no-code technologies that can be delivered as a service in a cloud environment.In this presentation, we will discuss the implementation of energy forecasting as a cloud solution. We will explain how the algorithms are made available to the forecaster and describe some of the commonly used techniques for energy forecasting. Our goal is to present how utilities of any size can make informed near-real-time decisions with more readily accessible energy forecasting software.

An Overview Of Sas® Visual Forecasting

Presenter: Joseph Katz
Co-authors: Joseph Katz

SAS® Visual Forecasting provides a resilient, distributed, time series analysis and scripting environment for cloud computing leveraging the speed, scalability, and elasticity of the SAS in-memory environment. It provides automatic forecast model generation, automatic variable and event selection, and automatic model selection. Users can apply various modeling strategies, create custom modeling strategies, create custom models in Interactive Modeling, compare forecasts, override forecasts, and visualize results. Methods include traditional time series techniques as well as neural networks (NN), recurrent neural networks (RNN), and hybrid techniques (NN + time series) in the forecasting process. In addition, SAS® Visual Forecasting is an open and extensible framework that integrates and scales open-source algorithms to run in parallel.

Advancements In Forecasting Using Machine Learning

Presenter: Charles Chase
Co-authors: Charles Chase

Intelligent automation (IA) driven by AI and ML are disrupting the way companies do business. The rapid deployment of automation is helping us set new standards of efficiency, speed, and functionality. With the introduction of intelligent automation using ML, a large portion of the manual, repetitive activities can be automated allowing demand analysts and planners to be more productive adding real value to the overall process. This session will focus on three areas where IA can have a positive impact. Demand SensingDemand-supply forecasts are sensitive to changing market dynamics caused by unexpected weather events, policies, and, more recently, pandemic situations. This makes maintaining an accurate demand forecast very challenging and traditional forecast models tend to fail in such scenarios. To address this challenge Demand Sensing identifies significant external demand signals and also enhances traditional forecast models with a ML/DL-based hybrid approach to attain stability and accuracy.Assisted Demand Planning Assisted Demand Planning which is an intelligent automation technique that uses ML to boost FVA to learn from past demand planners' manual overrides. It provides demand planners with a digital assistant using ML to provide the direction and range of overrides (as to the need to raise or lower statistical forecasts) at various levels of the business hierarchy. ML analyzes past statistical forecasts and consensus forecast overrides to learn from successful and unsuccessful forecast adjustments to identify the best periods to review candidates for overrides. It then provides guidance to demand planners about where, and by how much, to adjust the business hierarchy forecasts by either raising or lowering the statistical forecast. New Product ForecastingNew Product Forecasting helps overcome the challenge of the lack of historical data required to forecast using any regular forecasting tools. New Product Forecasting is an attribute-based forecasting technique where in the first step the "like" products and substitute products are automatically identified based on the new product's attributes. With the help of the "like" product's history, a forecast for the new product is generated. The model runs automatically and includes advanced predictive techniques and machine learning techniques like decision tree, random forest, neural network, and cluster analysis.

Properties Of The Reconciled Gaussian And Count Distributions

Presenter: Lorenzo Zambon
Co-authors: Lorenzo Zambon;Arianna Agosto;Paolo Giudici;Giorgio Corani

We analyze the properties of the reconciled distributions, following an approach to probabilistic reconciliation based on conditioning. In the Gaussian case, the reconciled distribution is obtained analytically (Corani et al., 2020); the reconciled mean and variance are the same of MinT, despite the different derivation strategy. As a first contribution, we show that the reconciliation decreases the variance of every variable

of the hierarchy, regardless of the amount of incoherence of the base forecasts. In the non-Gaussian case, probabilistic reconciliation via conditioning can be implemented via sampling (Corani et al., 2022 and Zambon et al., 2022). However, the properties of the reconciled distribution have not been formally studied yet. We prove that the variance of the reconciled distribution can be larger than the variance of the base forecasts: this happens if the base forecasts are largely incoherent, as they provide conflicting information. As for the reconciled mean, our discussion is restricted to the case of 1-level hierarchies. The reconciled upper mean can be a combination of the bottom-up and the base upper means, as in the Gaussian case (Corani et al., 2020; Hollyman et al., 2021). However, it can also be lower than both; we call this the strengthening effect. This happens when the distributions of the base forecasts are skewed: if the incoherence is low, the variance decreases after reconciliation as the base forecasts mutually reinforce each other, shortening the right tail and pulling the reconciled means towards zero.We empirically analyze the reconciliation of low-counts time series of economic defaults. We build a simple hierarchical structure by considering five different economic sectors (bottom time series) and their sum (upper time series). The reconciled forecasts clearly outperform the base forecasts on different scores. This confirms the positive results about the effect of reconciliation on intermittent time series (Kourentzes & Athanasapoulos, 2021), extending them to the probabilistic case. Finally, we find an empirical confirmation of our theoretical study.

Probabilistic Cross Temporal Forecasts

Presenter: Ross Hollyman
Co-authors: Ross Hollyman

Recent work developed an approach to cross sectional hierarchical forecasting based on a combinations of forecasts of bottom level series conditioned on base forecasts constructed at each level of a hierarchy. Key ingredients in the recipe for large collections of forecasts are a set of univariate calibration models (1 per series) and a factor model describing the relationship between cross-sectional forecast errors. In this work we extend this approach to cross-temporal forecasts. Multi horizon forecasts for collections of series are linked together with a dynamic (as opposed to a static) factor model enabling serial and cross-sectional correlation in forecast errors to be taken in to account.

Neural Networks And Temporal Hierarchies

Presenter: Nikolaos Kourentzes
Co-authors: Nikolaos Kourentzes

Using multiple temporal aggregation levels in modelling time series can help extract additional information from the data. Conventionally, each aggregated time series is modelled independently, and the resulting forecasts are combined in a single coherent forecast that encapsulates all the relevant information. This approach is at the core of temporal hierarchies, which has appeared with several variants in the literature, both for statistical and AI methods. A challenge in temporal hierarchies, as with all hierarchical methods, is to evaluate whether some aggregation level is beneficial or not to the combination. Although the standard temporal hierarchies do not consider rejecting specific aggregation levels, it is simple to construct examples where aggregate levels may be harmful to the performance of the final forecast, or at minimum computationally expensive without providing any benefit. Focusing on neural networks, we investigate alternative architectures that attempt to address this limitation. Amongst these, we propose a new type of temporal aggregation mechanism that has parallels with the increasingly popular attention mechanism. We empirically demonstrate the merits of the approach and attempt to elucidate the conditions where alternative multiple temporal aggregation approaches are more beneficial.

Statistical Versus Machine Learning Models For Time Series Forecasting: Characterization, Alerts And Recommendations For Use

Presenter: Andreína Alamo
Co-authors: Andreína Alamo;Mario Arrieta;Sergio Calderon

Machine learning techniques are becoming one of the most popular tools in time series forecasting.However, as the M4 competitions show, these are not always the best option when it comes toforecasting and, sometimes, traditional statistical models can be much more accurate. This workpresents recommendations and alerts on the use of these techniques, specifically on Long Short-TermMemory (LSTM) networks applied to time series analysis. This guide is developed through asimulation study in which various configurations for the network and different time series processesare used to evaluate the sensitivity of LSTM neural networks to common time-series features such asnon-stationarity and autocorrelation patterns. A series of recommendations that allow to determinewhether an LSTM model or a more traditional model is the most adequate forecasting tool arepresented, among which, (1) the importance of assessing non-stationarity (because LSTMs arehighly sensitive to it), (2) the need for a large sample size when training LSTM models, and (3) theimportance of analyzing the autocorrelation structure in determining certain hyperparameters ofthe LSTM stand out as key findings.

A Tree-Based Framework For Early And Interpretable Forecasts Of Students' Exam Performance Using Online Activity, Self-Regulation Data, And Behavioral Prompting

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

Education technology (EdTech) is a rapidly growing field that aims to enhance learning outcomes not only for formal education but also for corporate training and lifelong learning. A key component of EdTech is learning analytics, which monitors and analyses the learning of students and workers. One of the major challenges is to provide accurate and explainable predictions of the learners' performances to gain actionable insights. In a school setting, this should also happen in a timely manner, so actions can be taken early on in the semester. In this work, we propose a novel approach that generates interpretable predictions of student exam scores early on in the semester, based on their online activity and self-reported self-regulation information. We use gradient-boosting forecasting models with engineered features based on expert knowledge, to predict student exam results halfway through the semester. This allows enough time for students to adjust their learning habits. Every student is given timely feedback and personalised interventions. We identify the features that most negatively contribute to their predicted score using Shapley Values, and based on these, generate specialized prompts. Our framework increases the insight in important contributors to student success and failure. We evaluate our approach on three years of online data from a single course and show that it achieves higher forecasting accuracy than existing baseline methods, while making predictions halfway through the semester. In our empirical case study, we show how personalised prompts help students to improve their learning habits and ultimately their exam score.

Deep Neural Networks For Shallow Results? An Empirical Evaluation On The Benefit Of Customising Artificial Neural Network Algorithms In Industry Forecasting

Presenter: Sven F. Crone
Co-authors: Sven F. Crone;Kamran Rismanchi;Michelle Huang;Heiko Kausch;Tobias Kempcke

The recent success of Deep Artificial Neural Networks in text, image, and speech recognition as well

as generation has led to their applications in various other disciplines, including time series and causal forecasting. Long-Short term Memory Recurrent Networks (Ma et al, 2015), Deep Belief Networks (Kuremoto et al., 2014), and Structural Autoencoders (Gensler et al., 2016) are but few of the deep architectures promising increases in accuracy over the more established multilayer perceptrons (Crone, 2010; Dudek, 2016). However, a recent survey in forecasting practice indicated that only few companies employ AI in forecasting, and that over 50% of all AI forecasting projects in industry fail (Crone, 2022). This empirical study seeks to assess this discrepancy between the notable hype and lack of implementations of deep neural networks by assessing the empirical accuracy across multiple real-world industry datasets in medical technnology companies. We employ a valid empirical evaluation design across 4 companies and thousands of time series, using reliable error metrics, fixed multi-step horizons and multiple rolling time origins to compare against established statistical forecasting benchmarks in R (including ets, autoarima, theta), recent data science contenders (including Facebook's Prophet, Google's bsts), and machine learning methods including XGBoost and random forests. Our experiments confirm prior studies (e.g. Markidakis and Spiliotis, 2018), which show that the standard implementations of both deep and shallow neural networks architectures fail to outperform established statistical benchmarks of ets, arima and sometimes even Naive on the monthly industry datasets still widely used in practice. However, when presented with carefully engineered feature sets, applying time series feature creation, transformation and selection, and providing algorithm parameter turning for these features, both the shallow and deep networks can be customised to outperform statistical, ml and data science benchmarks methods. We conclude that using „vanilla“ deep or shallow neural network implementations from R or python packages yields poor results, but careful customisation of features can significantly increase their efficacy.

More Efficient Multiple Time Series Forecasting With Time Series Clustering

Presenter: Prince Grover
Co-authors: Prince Grover;Huy Nguyen

In many real world forecasting problems, we deal with tens of thousands of multivariate time series that are related to each other in certain dimensions but also have distinct individual characteristics of trend, seasonality or irregularities. An example is demand forecasting of products sold by an e-commerce in different locations. The demand patterns of sports shoes is likely different than the demand patterns of winter jackets, with latter getting more demand during winter seasons. Since both categories of items are sold by the same e-commerce, they would also have various common patterns like trend of daily active customers, holidays, discounts etc. A single model trained on multiple time series can leverage inter-series dependencies and provide good priors for new related series. Recent progress in forecasting research has led to the exploration of neural network architectures for co-training multiple time series and use residual connections for better and faster training. However, training tens of thousands of related series in single network may also be computationally inefficient and overlook individual level patterns. In contrast, individual series level models may fail to exploit commonalities and transfer learning across series. In this paper, we propose that clustering multiple time series based on their similarity and co-training a model for each cluster can achieve better performance than both single model and individual model approaches. We compare different clustering methods (Dynamic Time Warping (DTW), FastDTW and cosine similarity) and different deep learning forecasting models (LSTM, Transformer, N-BEATS, N-HiTS) on five open source datasets (Electricity, Traffic Usage, Solar, M series). We show that the clustering based approaches outperform the other baselines.

Ensemble Forecasting By Energy Associated Modified By Renyi's Entropy And Statistical Roughness In The Learning Process

Presenter: Cristian Rodriguez Rivero
Co-authors: Cristian Rodriguez Rivero;Julian Pucheta;Gustavo Juarez;Martin Herrera;Daniel Patino;Leonardo Franco

This paper proposes an ensemble method based on Energy associated modified method by Renyi's entropy and statistical roughness to tune recurrent neural networks (E-EAS-LSTM) parameters. The computational models are carried out for univariate chaotic time series with multi-step prediction horizons. The combination of energy associated with series and the relative entropy combines heuristically modified the learning process. The study analyses and compares the relative advantages and limitations of several algorithms, including statistical methods to forecast both benchmark series and rainfall series from 1 to 6 months ahead. Simulation results illustrate that regardless of the effectiveness of the E-EAS-LSTM approach through different series classified by the proposed approach in both, the learning process and the validation test using the SMAPE and NRMSE, and MASE metrics, and by adding fractional Gaussian noise, with a good performance in terms of accuracy.

Blocked Bootstrap Conformal Prediction For Non-Stationary Time-Series

Presenter: Kevin Chen
Co-authors: Kevin Chen;Sankalp Gilda;Oskar Triebe

Reliable uncertainty estimation of model forecasts can greatly benefit real-world applications that make long-term, mission-critical decisions: from energy generation to supply chain management. This uncertainty can be quantified as a prediction interval around the point forecast, for any desired confidence level. Conformal prediction (CP) is a family of prediction interval methods that are both distribution-free and model-agnostic, as they use past data residuals (i.e., "nonconformity" scores) to create the interval width (i.e., q-hat). Conventional CP methods assume that the underlying data is exchangeable, or sequentially invariant, and therefore typically applied on i.i.d. data. More recent CP methods, such as Ensemble Batch Prediction Interval (EnbPI) and Sequential Predictive Conformal Inference (SCPI), orient toward time-series CP by employing non-overlapping block bootstrap (NBB) to narrow prediction interval estimates. In general, however, block bootstrap (BB) is suitable for time-series data only under certain conditions. For example, multiple BB methods assume initial time-series stationarity as they otherwise likely produce bias, by breaking down temporal features like trend and seasonality. NBB, in particular, is challenging to apply to smaller time-series due to the limited number of non-overlapping blocks, unless the length of each block can be further reduced. Nevertheless, this makes setting the block length another essential dataset-dependent choice. We propose the following CP modifications that are suitable for both non-stationary and smaller time-series. First, we stationarize the time-series with a decomposition method before applying bootstrapping. Second, we consider replacing NBB with other more sophisticated BB methods, including moving block bootstrap, stationary block bootstrap, and circular block bootstrap. Third, we explore time-series cross-validation methods as alternatives to bootstrapping entirely. We evaluate whether these modifications can further reduce prediction interval width at all confidence levels without sacrificing their out-of-sample (OOS) coverage rate, in comparison to EnbPI and SCPI. This allows conformal prediction to indeed be applicable to a broader set of real-world time-series.

Modeling Unemployment And Labor Force Participation By Gender And Race

Presenter: Junie Joseph
Co-authors: Junie Joseph

Despite progress for underserved and underrepresented communities over the past decades, persistent gaps remain in key labor-force indicators such as unemployment and labor force participation. The COVID-19 pandemic exacerbated the disparate labor market outcomes for individuals from these communities. While policy makers have expressed interest in understanding how outcomes differ by gender and race/ethnicity, more research on the nature of these gaps is needed. This paper provides insight about these gaps by developing vector autoregressive models in a cointegration framework to ascertain long-run relationships between unemployment and labor force participation rates by gender and race over 1980–2019.

Forecasting from those models into the pandemic helps quantify the consequences of the pandemic on the labor market and highlights the heterogeneity across gender and race that continues to persist. These results contribute to an understanding of the disparities in the labor market and adjustments among demographic intersections by gender and race/ethnicity.

Statistical Properties Of Fomc Forecast Distributions: Evidence From Individual Fomc Participants From 2008 To 2017

Presenter: Jaime Marquez
Co-authors: Jaime Marquez;Joseph Martorana

Conventional explanations of monetary policy decisions in the United States rest on the assumption that the Federal funds rate is determined by a representative central banker (i.e., the Fed) who uses forecasts of prices, economic activity, and unemployment. This assumption is inconsistent with the federalist structure of the Federal Reserve in which the Federal funds rate is determined by a committee made up of the Federal Reserve Board and the Federal Reserve Banks. This inconsistency would be irrelevant if differences in the Fed participants' project from 2008 to 2017ions were small or constant but they are not. Indeed, disparities in their forecasts are large and volatile, across all the forecast horizons used by the FOMC: current-year, one-year ahead, two-year ahead, and longer-run. In other words, analyses or narratives of U.S. monetary policy that proceed as if the conditions for a representative central bank held are inconsistent with the facts. This finding raises several questions: How important is the institutional affiliation of FOMC participants between the Federal Reserve Board and the Federal Regional Banks in accounting for the differences in forecasts? Are FOMC participants relying on the same forecasting framework (i.e., model or rules of thumb) but using different values for the forecast drivers? Or are these participants using the same forecast drivers but relying on different frameworks? Or both? These are the questions that we examine.

Labor Force Participation And Unemployment: Structural Change From The Pandemic?

Presenter: Neil Ericsson
Co-authors: Neil Ericsson;Victoria Tribone;Andrew Martinez

The COVID-19 pandemic resulted in the most abrupt changes in U.S. labor force participation and unemployment since the Second World War, with consequences differing by gender and age. We model the U.S. labor market to help to interpret the pandemic's effects. Specifically, we formulate joint dynamic cointegrated models of disaggregated unemployment and labor force participation rates for 1980–2019. We then use those models to forecast into the pandemic to understand the pandemic's labor market consequences, treating those forecasts as being from an alternative scenario in which the pandemic didn't occur. Heterogeneity across gender and age is particularly prominent at the pandemic's outset. Lower labor force participation persists among many subgroups.

Using The Theory Of Economic Interdependence To Measure And Predict The Regional Development Of The Guangdong-Hong Kong-Macao Greater Bay Area

Presenter: Yijie Wang
Co-authors: Yijie Wang;Wei Chong Choo;Ng Keng Yap

Compared with the world's three major bay area economies, the Guangdong-Hong Kong-Macao Greater Bay Area enjoys unique location advantages and policy dividends. The close economic interdependence between regions has great significance and potential opportunities for developing individuals, enterprises,

and society. This study first analyzes the characteristics of the economic development status of the Guangdong-Hong Kong-Macao Greater Bay Area and the comparative study of other bay areas. Secondly, based on measuring and analyzing the economic dependence of the Bay Area, the composite panel data of 23 cities in Guangdong, Hong Kong and Macao from 2015 to 2022 were used to construct a dummy variable regression model. Through empirical analysis, this research aims to predict and analyze the economic linkage development and market potential investment direction of the Guangdong-Hong Kong-Macao Greater Bay Area. The research shows that the degree of economic interdependence and the spatial distribution of economic development in the Guangdong-Hong Kong-Macao Greater Bay Area is extremely unbalanced, and the tertiary industry is developing rapidly, but there is a large room for transformation and improvement. Moreover, the radiation effect of the central city is significant, and the market investment prospects in foreign investment, resource agglomeration and import and export business are huge.

Do Economic Leading Indicators Work In China?

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

Economic indicators are supposed to provide insight into economic activities and prospects, in hindsight as well as in advance. Leading indicators, such as inflow of new orders, are supposed to anticipate lagging indicators, such as GDP. OECD provides indicators for many countries, among them China, on a monthly basis. Examples of leading indicators are the survey-based BCI (Business Confidence Indicator) and the econometrics-based CLI (Composite Leading Indicator). Our research question is: Is there a leading indicator which can actually anticipate economic development in China?Cross-wavelet methodology provides a means to assess the capability of a series (the leading indicator) to lead another series (the lagging indicator) across time and a selected period band. We show that BCI and CLI lead the GDP series at business cycle length in large Western economies. For China, we find that the CLI is not consistently successful in anticipating Chinese GDP development. The BCI is more successful in China. We also discuss the possibilities of infomation-theoretic concepts derived from stock-market connectedness to construct a leading economic indicator for China.

Performance Of Housing Starts Forecasts

Presenter: Prakash Loungani
Co-authors: Prakash Loungani;Davide Furceri;Hites Ahir;Karan Bhasin

Housing markets have played a critical role in economic growth across several advanced and emerging markets over the last few decades. Several forecasters have tracked housing markets closely, while providing forecasts of housing starts indicating the importance of the sector for macroeconomic forecasting. Despite the growing importance of housing sector, consistent data on housing markets and forecasts are limited to few variables for a small number of countries. This paper focuses on forecasts of housing starts for Australia, Canada, Japan and the US. We rely on housing starts forecast data from Consensus Forecasts at different horizons. Using a Bayesian learning model, we explore the role of forecast revisions and lagged revisions to explain contemporaneous and one year ahead forecast errors. We look at the impact of past forecast revisions on current forecast revision. Further, we look at models that help explain housing starts and compare the performance of modeled forecasts with consensus forecasts.

The Systemic Risk Approach Based On Implied And Realized Volatility

Presenter: Pawel Sakowski
Co-authors: Pawel Sakowski;Rafal Sieradzki;Robert Slepaczuk

We propose a new measure of systemic risk to analyze the impact of the major financial market turmoils inthe stock markets from 2000 to 2021 in the USA, Europe, Brazil, and Japan. Our Implied Volatility RealizedVolatility Systemic Risk Indicator (IVRVSRI) shows that the reaction of stock markets varies across differentgeographical locations and the persistence of the shocks depends on the historical volatility and long-termaverage volatility level in a given market. The methodology applied is based on the logic "the simpler isalways better than the more complex, if it leads to the same results". Such an approach significantly limitsthe model risk and substantially decreases computational burden. Robustness checks show that IVRVSRIis a precise measure of the current systemic risk in the stock markets. Moreover, IVRVSRI seems to be avalid indication of current systemic risk in equity markets and it can be used for other types of assets andhigh-frequency data. The comparison of IVRVRSI with such measures as sRisk, and CATFIN shows that its reactions are quicker and therefore it has a higher predictive power.

Forecasting Volatility Of Stock Market Indices Using Recurrent Neural Networks: A Sequence-Based Approach

Presenter: Florian Sckade
Co-authors: Florian Sckade;Maik Dierkes

Volatility forecasts are essential for a wide range of financial applications, such as portfolio optimization and risk management. Machine learning models are an attractive alternative to traditional econometric methods and are frequently applied in finance research. They are capable of capturing nonlinear behaviors commonly attributed to volatility. We analyze the predictive performance of different recurrent neural networks (RNNs) - simple RNNs, LSTMs, and GRUs - at forecasting realized volatility of three stock market indices. We use RNNs to forecast not just single points or aggregate measures over longer horizons via parallel models, but to directly forecast monthly sequences of volatility as vectors, analyzing the impact of hidden state sizes in single-layer RNNs and the influence of ensembling on results. For this purpose, we use the hidden state vector of the RNNs as inputs to a feed-forward neural network (FFNN) with an output vector, as opposed to a single-point prediction. As these predictions in the FFNN output vector do not explicitly represent the time-series nature of the actual data, we compare the results with Encoder-Decoder RNNs, which allow for easier consideration of the sequential nature of the outputs. The results are mixed and show that RNNs can outperform traditional econometric methods without requiring very deep architectures. However, they do not significantly reduce forecast error when compared to the log HAR model. With sequence-predicting models performing comparatively to individually trained models, we argue for wider adoption of more complex sequence-based RNNs in volatility forecasting.

Light Benchmark - Comprehensive Backtesting Framework For Market Risk Models Comparison

Presenter: Michał Woźniak
Co-authors: Michał Woźniak;Robert Ślepaczuk

The aim of the article is to present LIGHT Benchmark - a real world, multidimensional, comprehensive, and agnostic backtesting framework for market risk models comparison. The algorithm allows for reliable assessment of the quality of the classic and state of the art Value at Risk and Expected Shortfall models in the form of a scoreboard. Our benchmark consists of 375 validation paths, of which two thirds are dedicated to Value at Risk and one third to Expected Shortfall. Each proposed scenario is independent and meets the requirements of the median research on market risk estimation, i.e., meeting the capital requirements imposed by the regulator and optimising day-to-day operations in terms of the amount of reserves held. Taking it into account, we cover following categories of validation techniques during the scoring procedure: fulfilling regulatory requirements, forecasting adequateness and capital effectiveness. Benchmark paths are built on such dimensions as: asset categories, representatives of each asset category, testing periods, testing period sizes, testing horizons and VaR and ES confidence levels. Our scenarios

are built based on our own calculations and an extensive quantitative literature review. In addition, we propose our expert scoring system, which is designed to aggregate individual scores into a final scoreboard. This system has been tested for stability and compatibility with other popular methods such as Model Confidence Set Procedure. The data used in this benchmark has a daily frequency and comes from the period from 2000 to 2021. Finally, we illustrate the use of the LIGHT Benchmark by scoring both classic and promising econometric models for the problem of market risk.

Nested Multiple Seasonalities: Accuracy Vs. Runtime

Presenter: Stephan Kolassa
Co-authors: Stephan Kolassa

As data are collected at higher and higher temporal granularity in hourly or even sub-hourly buckets, time series seasonality becomes more complex. Specifically, hourly data usually exhibits both intra-daily and intra-weekly multiple seasonalities, which are "nested" within each other. Different specialized forecasting methods have been proposed to address this structure, from TBATS over Multiple Seasonal-Trend decomposition using Loess (MSTL), Meta's Prophet model, Multiple Seasonal ARIMA (MSARIMA), Several Seasonalities or State Space ARIMA (SSARIMA) to Double Seasonal Holt-Winters (DSHW). We compare these methods - as well as simple benchmarks like averaging ETS and STL models on the two seasonalities separately, and regression models using harmonics - on hourly demand datasets from different sources, some of which are highly intermittent. Outcome measures are accuracy, but also runtime, as a proxy for overall resource requirements of forecasting models, which become more and more important as non-accuracy attributes of forecasting pipelines, like the CO2 footprint, start drawing attention.

Survey Expectations And Adjustments For Multiple Testing

Presenter: Michael Clements
Co-authors: Michael Clements

Testing hypotheses regarding how individual survey respondents form their expectations is susceptible to the multiple testing problem. Bonferroni-correction controls the family-wise error rate, but may be too conservative. Instead we investigate the impact of controlling the false discovery rate - the expected proportion of the rejected hypotheses that are false - and whether this alters the inferences about how professionals forecast. The effects of adjustments for multiple testing are investigatedfor tests of weak efficiency and the over-reaction hypothesis, for beliefs about the persistence of shocks to output growth, and for forecasters' ability to accurately assess the uncertainty surrounding their forecasts.

Forecasting Performance Versus Computational Cost

Presenter: Fotios Petropoulos
Co-authors: Fotios Petropoulos;Evangelos Spiliotis

The available forecasting algorithms have increased in complexity and computational cost, with direct monetary and environmental implications. In this presentation, we explore the computational cost of forecasting from two perspectives. First, we question the need to consider large families of models. Our argument is that parsimoniously identifying suitable subsets of models will not decrease forecasting accuracy, nor will it reduce the ability to estimate forecast uncertainty. Second, we argue that frequent updates to specify the most "optimal" model form and set of parameters may be irrelevant in practice, also unnecessarily increasing computational cost. We empirically demonstrate our two arguments using a large collection of real-life data. We find that reduced pools of models, as well as intermediate updating scenarios, including the re-estimation of specific parameters but not necessarily the specification of the model form,

can result in similar or even better accuracy with significantly lower computational cost. We translate computational benefits to monetary cost savings and environmental impact and discuss the implications of our results in the context of large retailers.

An Actor-Critic Method For Forecasting Collections Of Related Time Series

Presenter: Joao Filipe Pinto Sousa
Co-authors: Joao Filipe Pinto Sousa;Roberto Henriques

Forecasting extensive collections of related time series (TS) is a relevant task in various domains, e.g. economic or financial predictions, public forecasting, demand models, and consumption plans. More often than not, these problems involve handling high-dimensional multivariate series which relate to each other through potentially complex patterns. Exploiting such global patterns while keeping good statistical accuracy and expressiveness level for locally calibrated predictions (i.e. each individual series) has been a widely studied objective in recent years. The literature already proposes different hybrid models, either combining the benefits of classical TS models (e.g. ARIMA) with deep learning (DL) approaches or combining more complex temporal networks, which are decomposed and regularised differently to capture local and global properties effectively. However, most of these methods are either one-dimensional or are inadequate to adapt fast to changing data distributions or relevant contextual dimensions for the real-world (e.g. cost factors). To circumvent this, we propose a new deep reinforcement learning (DRL) method – T2f. Our approach builds on traditional data-driven ensemble learning fundamentals and innovates by employing an actor-critic method based on the Twin Delayed Deep Deterministic (TD3) algorithm. T2f formulates the prediction task based on a two-step input sequence, which starts by learning a local policy from which local properties are extracted as base model features. The second step layer learns a deep factor based on the global policy, which takes these features as latent variables to control for hidden features in the time series collection. The actor's output layer then generates the combination weights for the two input steps. We present succinct theoretical and empirical evidence for our proposal by putting it against both global and local state-of-the-art methods. This is done on two datasets, one artificial and a large retail sales report. Our results demonstrate the soundness of our solution in terms of accuracy and efficiency in balancing local and global signals dynamically. The latter is achieved in both fast adaptation to changes in the distribution and in consideration of contextually relevant dimensions.

Explaining Which Time Series The Global Forecasting Model Uses To Make Predictions With A Shapley Based Approach

Presenter: Thiebe Sleeuwaert
Co-authors: Thiebe Sleeuwaert;Ruben Crevits

Explainable AI (XAI) is an important topic, in supply chain, and in forecasting in general. Forecast consumers expect more than only an automatically generated accurate forecast. They want to know what drives the outcome of the global forecasting algorithm. Shapley values have been successfully applied to derive how much each feature contributes to the forecast. We expand this idea to not only explain the impact of features to the forecast, but also which time series contributed to the forecast. We create an algorithm using the Shapley principle on the data points instead of the features. We illustrate the value of this approach on a synthetic data set. The examples we show are realistic in a demand forecasting context. Finally, we apply the algorithm on a real dataset.

Global Learning On Heterogeneous Datasets

Presenter: Arnoud Wellens
Co-authors: Arnoud Wellens;Nikolaos Kourentzes

Global forecasting methods have shown to be very competitive on real-world datasets. By fitting a single method over multiple time series, global methods have access to more training data seemingly leading to better generalization than more traditional local forecasting methods that estimate a single method per time series. Surprisingly, these results even hold true for groups of heterogeneous time series. In this paper, we try to understand how heterogeneity complicates the estimation of model parameters and how linear and non-linear global methods deal with this. By simulating homogeneous and heterogeneous groups of time series, we dive deeper into the local and global estimation of the model parameters of linear regression and gradient boosting decision trees. First, we show that estimating the correct model parameters, even for simple time series, requires substantial sample. Second, we show that correctly specified local methods are hard to outperform on heterogeneous datasets by simply providing more training data and lags. Third, we find that non-linear global forecasting methods handle heterogeneity better than their linear counterpart. This difference in forecast accuracy can be offset by including inputs based on problem-specific feature engineering. While it shows that linear global methods are also able to successfully deal with heterogeneity, the required feature engineering is hard to implement for complex data generation processes, demonstrating an advantage of non-linear methods.

Mexico: Determinants Of The Real Exchange Rate, 2001.01-2021.12

Presenter: Eduardo Loría
Co-authors: Eduardo Loría

We estimate the determinants (terms of trade, tradable to non-tradable price differentials, interest rate differentials, forward exchange rate and risk premium) of the Mexican real exchange rate for the short and long run by using an ARDL model for Mexico (2001.01-2021.12). The inclusion of financial variables is the main contribution. Our results indicate no cointegrating relationship either for the entire sample, or for 2001.01-2008.12. However, for 2009.01-2021.12, there is strong empirical evidence of its existence due to the increasing international financialization process.

Temporal Aggregation Bias And Forecasts Of Effective Exchange Rates: We Know Less Then We Think!

Presenter: Martin Mccarthy
Co-authors: Martin Mccarthy;Stephen Snudden

Forecasts of a temporal average, such as monthly average effective exchange rates (EERs), necessitate the use of the disaggregated data to both construct efficient forecasts and to correctly implement hypothesis tests. However, we show that no existing study has correctly tested for predictability nor efficiently constructed a forecast for average EERs, possibly due to the absence of official data on daily and end-of-month EERs. In this paper, we construct daily real and nominal EER for all countries that match the official exchange rates from the International Monetary Fund. We then use the data to show that disaggregated forecast approaches (bottom-up, MIDAS, and PEPS) can almost halve forecast error on average across all countries, relative to existing forecast approaches. For the first time, we test forecasts against the random walk from the daily data and find predictability of monthly nominal EERs at the short-horizons for a quarter of all countries. Predictability of real EER using real-time data is more difficult, but is still found for 12 percent of countries. The results suggest that policymakers can achieve substantial forecast gains my switching from aggregated to disaggregated forecast approaches. The evidence also calls for a critical re-evaluation of existing findings that were subject to temporal aggregation bias and the need for the availability and use of the disaggregated data going forward.

Forecasting Asean Exchange Rate Using Xgboost Methods

Presenter: Brian Sloboda
Co-authors: Brian Sloboda;Rolando Santos

With the recent increase in interest rates, most ASEAN (Association of Southeast Asian Nations) countries have experienced volatility in their exchange rates. Predictions in their exchange rates have been unpredictable due to shocks in inflation and interest rates in recent years. The purpose of this paper is to assess the accuracy of exchange rate forecasts of selected ASEAN countries using the Xgboost (eXtreme Gradient Boosting) model quarterly from 1995 through 2022. Xgboost is a popular and efficient open-source implementation of the gradient-boosted trees algorithm. Gradient boosting is a supervised learning algorithm, which uses a gradient descent algorithm to minimize a loss function when adding these new models together to make the final prediction. The paper will also use other traditional time series methods, e.g., ARIMA or AR(1) models as a basis for comparison. We use both the root mean squared error (RMSE) and graphical representation approaches in order to compare the performance of forecast methods, Because Xgboost is a machine learning model, previous models from earlier studies indicate that it will perform better than traditional exchange rates models such as ARIMA or AR(1) models.

Forecasting Regional-Level Cause-Specific Mortality For China: An Optimal Immutable Reconciliation Approach

Presenter: Han Li
Co-authors: Han Li;Katja Hanewald

This paper develops a forecast reconciliation approach to model and predict mortality improvements by geographic regions and causes of death. This method ensures consistent forecasts and enhances overall forecast quality while allowing for the examination of elimination scenarios. We apply the method to deaths and population exposure data for China from 2004-2019. Our model estimates show that China's average mortality improvement rate was about 3.2% per annum in 2004-2019. Average mortality improvement rates were highest in East China and lowest in West China. Urban and rural populations showed similar average mortality improvement rates. The reconciled mortality forecasts for 2020-2029 suggest that the average mortality improvement rate will decrease to around 2.2% per annum. Our model predicts continued regional differences in mortality rates between East, Central, and West China, while the urban and rural mortality gap is predicted to narrow down. Our mortality predictions under different cause-elimination scenarios indicate that eliminating cancer mortality could lead to a 25% reduction in mortality rates on average while eliminating circulatory-related deaths could reduce overall mortality by an average of 33%.

Crossovers Between Births Falls And Deaths Spikes During The Covid-19 Pandemic: The Mexican Case

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

We aim to describe patterns of crossovers between births and deaths during the COVID-19 pandemic in the Mexican case. Likewise, it is decided to emphasize our overview of births, that is why we also measure the birth falls by mothers' ages using vital statistics taken from the Mexican Institute of Statistics (INEGI). To do that, official figures are presented to identify crossovers. To dimension the births fall in a counterfactual sense, forecasts coming from a Multivariate time series model are employed. Our results address that the negative trend of births suffers a dramatic fall, however, in some cases, there are some recoveries. It is suggested that the anticipated occurrence of demographic crossovers was a consequence of the COVID-19 pandemic. Regardless of whether the crossovers are more or less deep at the state level, the significant birth falls may trigger accelerating aging in this undeveloped country.

Boosting Mortality Models With Age And Spatial Shrinkage.

Presenter: Anastasios Panagiotelis
Co-authors: Anastasios Panagiotelis;Han Li;Li Li

Forecasting mortality rates is a problem with unique characteristics. Forecasts must be made for multiple age groups across different geographical regions so that models based on latent factors, such as the Lee Carter model have been popular. In this paper we use the Lee Carter model as a weak learner together with boosting. While boosting has seen considerable success in forecasting, most popular implementations use trees as weak learners, so our application of boosting using a domain-specific model is novel. As another original contribution we consider adding a penalty term to our loss function so that forecasts corresponding to adjacent age groups or adjacent geographical regions are shrunk closer together. Our methodologies are applied to US mortality data and are shown to outperform both the Lee Carter model as well as more sophisticated modern mortality forecasting models.

Scaling-Aware Rating Of Count Forecasts

Presenter: Malte C. Tichy
Co-authors: Malte C. Tichy

Forecasts crave a rating that reflects the forecast's quality in the context of what is possible in theory and what is reasonable to expect in practice. Granular forecasts in the regime of low count rates - as they often occur in retail, for which an intermittent demand of a handful might be observed per product, day, and location - are dominated by the inevitable statistical uncertainty of the Poisson distribution. This makes it hard to judge whether a certain metric value is dominated by Poisson noise or truly indicates a bad prediction model. To make things worse, every evaluation metric suffers from scaling: Its value is mostly defined by the predicted selling rate and the resulting rate-dependent Poisson noise, and only secondarily by the quality of the forecast. For any metric, comparing two groups of forecasted products often yields "the slow movers are performing worse than the fast movers" or vice versa - the naïve scaling trap. To distill the intrinsic quality of a forecast, we stratify predictions into buckets of approximately equal rate and evaluate metrics for each bucket separately. By comparing the achieved value per bucket to benchmarks, we obtain a scaling-aware rating of count forecasts. Our procedure avoids the naïve scaling trap, provides an immediate intuitive judgment of forecast quality, and allows to compare forecasts for different products or even industries. We illustrate the method by applying it on models for the M5 competition dataset.

Forecasting Accuracy And Inventory Performance: Evidence On Their Relationship From The M5 Competition Data

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

Although it is generally accepted that greater forecasting accuracy can contribute towards better inventory performance, this relationship may be weak, also depending on the particularities of the products being forecast, the inventory policy considered, and the underlying costs. Using the time series of the M5 competition, we empirically explore the inventory performance of various widely used forecasting techniques and state of the art approaches, including exponential smoothing, the Croston's method and some of its variants, ARIMA models, approaches based on temporal aggregation, and an implementation of the winning submission of M5. We employ a rolling simulation approach and analyze the results of various setups of the order-up-to policy, trying to better understand the relationship between decrease in forecast error and increase in business value. Based on our findings, we discuss some practical concerns and make relevant recommendations.

Does Forecast Accuracy Matter: The Impact Of Accuracy Improvement On Supply Chain Outcomes

Presenter: Jim Hoover
Co-authors: Jim Hoover

A recent topic discussed among practitioners and academics in Foresight: The International Journal of Applied Forecasting was, "Does Forecast Accuracy Matter?" The discussion centered around whether improvements in forecast accuracy metrics yielded improved outcomes in business objectives like supply chain operations and costs. In this research, improvements in the forecast relative to the actual demand is measured without regard to a specific forecasting method, and then those forecasts are applied to a popular enterprise resource planning algorithm to measure impact related to supply chain outcomes. Specifically, we measure the impact of improving forecast accuracy on order fulfillment rates, backorders, and inventory holding costs. This comparison is done utilizing various demand patterns (continuous, intermittent and lumpy demand).

Integrating Bottom-Up Machine Learning And Top-Down Statistical Approaches For Hierarchical Forecasting: A Comprehensive Framework

Presenter: Arsa Nikzad
Co-authors: Arsa Nikzad;George Monokroussos;Ryan Mitchell;I-Chen Lee

This paper proposes an integrated approach to time series forecasting that combines top-down statistical models with bottom-up machine learning techniques, specifically boosting trees (LightGBM), to address different forecasting horizons. The motivation behind this approach is to leverage bottom-up methods for short-term forecasting and top-down models for longer-term forecasting, capitalizing on the strengths of each technique. The approach incorporates a categorization scheme based on time series features such as trend, seasonality, and autocorrelation. The method is demonstrated through a case study using real-world data from Wayfair, where it has been used to forecast monthly demand for millions of products. The paper also provides insights into best practices for feature selection and model parameter tuning. By integrating top-down forecasting using statistical models with bottom-up machine learning solutions for different forecasting horizons, this paper presents a practical and effective approach to time series forecasting that highlights the potential for further research in this area.

Forecasting Demand Distributions At Wayfair: Our Approach

Presenter: Ryan Mitchell
Co-authors: Ryan Mitchell

Wayfair produces monthly demand point forecasts for millions of products at various forecast horizons. These demand forecasts are then used to determine how much inventory should be stored at Wayfair's warehouses across the US. However, these point forecasts give no indication about the uncertainty of the demand forecast and therefore do not give any information about how much safety stock Wayfair should hold. Incorrect safety stock levels are costly, leading to either missed sales or unnecessary storage costs.To solve this issue we produce product level demand forecast distributions at various forecast horizons. Specifically, we use the point forecast as the mean of the distribution, out of sample forecast errors to calculate the variance of the distribution, along with an algorithm to determine which distributional assumption to fit around the mean and variance. We use an evaluation methodology proposed by Diebold et al. (1997) to test the accuracy of the distributions by comparing the probability integral transformations from back estimated distributions against a uniform distribution via a Kolmogorov Smirnov test. Using this we find our methodology leads to forecast distribution accuracy levels of ~75% of wholesales costs in the US, and allows for product specific safety stock calculations.

Adjusting Demand Forecasts For Ongoing Experimental Results

Presenter: Tara Sullivan
Co-authors: Tara Sullivan

As a demand forecasting team at Wayfair, our goal is to provide accurate and informative forecasts at the product level based on historical sales to our suppliers and inventory managers.Separately, our broader organization is frequently implementing large scale experiments that help shape demand at the customer level. A key goal in our organization is to resolve this tension between historical accuracy and responsiveness to demand shifting initiatives. As an example, ongoing experiments at Wayfair push relevant products that are closer to our customers to a more favorable sort position. This initiative increases the likelihood that customers receive products quicker with fewer incidents. Traveling shorter distances allows us to pass ship cost savings onto our customers, and nudges suppliers towards our fulfillment network. As this initiative has been improved and refined utilizing vintages of AB test data, we have leveraged experimental results to generate forecast multipliers. This process allows us to succinctly communicate the potential impact of experimental initiatives, without fundamentally changing our forecasting processes.

Fast Autotuning Of Recurrent Neural Network Forecasting Models On Parallel And Distributed Architectures

Presenter: Taiyeong Lee
Co-authors: Taiyeong Lee;Thiago Quirino;Mahesh Joshi

Recurrent neural network (RNN) models, such as vanilla RNN, LSTM, and GRU, are widely used in a variety of areas. They are also well suited for time series forecasting because of their cyclic features. RNN models have a lot of hyperparameters, which you must tune properly to get the best-performing model. However, hyperparameter tuning usually requires substantial computing resources. SAS® Visual Forecasting supports recurrent neural network models through different time series packages that are available in the TSMODEL procedure. The parallel and distributed architecture of the SAS® Viya® platform provides a very efficient way to run multiple tasks in parallel. Further, the INSCALAR= data option of PROC TSMODEL enables easy customization and distribution of each parallel task's input. In this presentation, we show how these platform and procedure features make RNN hyperparameter tuning extremely efficient by trying many hyperparameter combinations in as little time as it takes to fit just one model. The number of combinations that you can try scales linearly with the hardware capacity that you can invest in.

Scalable, Cloud-Based Hierarchical Reconciliation Methods

Presenter: Nilesh Jakhotiya
Co-authors: Nilesh Jakhotiya;Caiqin Wang;Matt Simpson;Michele Trovero

It is a common practice in many sectors to organize time series in a hierarchical structure. Businesses often require not only good forecast accuracy for each time series in the hierarchy but also coherency across all aggregation levels—that is, the forecasts for the lower levels of data disaggregation add up to the forecasts of the upper levels of aggregation. In practical applications, the base forecasts are usually generated independently for each time series by using suitable models, and coherency is then enforced by an after-the-fact process called reconciliation. Research shows that global reconciliation of forecasts from all levels has greater accuracy than more traditional approaches, such as bottom-up, top-down, or middle-out methods. However, global reconciliation faces two major challenges. One is the scalability problem in dealing with a large collection of time series. The other is the presence of various patterns of missing values in forecast series among all aggregation levels in real-world data. This presentation compares two scalable methods of globally and locally optimal reconciliation, both of which are designed for a distributed

cloud environment that can process large-scale time series data and handle missing values. SAS® Visual Forecasting software and open source software are used to demonstrate these methods.

Applications Of Forecasting To Healthcare Data

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

Recent years have seen an increasing need to analyze the financial impact of gender differences and longer life expectancies. The healthcare field in particular gets more attention from researchers for designing treatments that take patients' gender and age into account. For example, it is important to know that, according to the Centers for Disease Control and Prevention (CDC) in the United States, about 26% of women over the age of 65 have osteoarthritis, compared to only 18% of men in the same age group. Also, the increasing cost of healthcare is a major issue in the US. Accurate forecasting of healthcare costs can provide financial benefits to patients, providers, and payers. We analyze CDC data, which includes 50,000 participants from July 2009 to June 2012. We create a cohort from age groups 46 to 75 who have chronic episodes and preventive episodes. The definition of episode conditions is from the Healthcare Incentives Improvement Institute (HCI3) Episode of Care Definitions using ICD-10 (International Classification of Diseases). We aggregate the data to monthly time periods and compute average costs. The healthcare cost time series by episodes show different patterns according to gender and age group. The patterns can be seasonal, upward trends, downward trends, etc. Different patterns derived from the time series attribute algorithm can help analysts choose appropriate time series forecasting methods to improve overall forecast accuracy. Another important choice is whether to use gender and age variables as grouping (BY) variables or independent variables. In this paper, we present the effects of gender and age group on healthcare cost forecasting by using different modeling methods for different time series patterns and segments.

Forecasting The Market Size Of Industries Using Time Series Of Patents

Presenter: Jooyoung Jeon
Co-authors: Jooyoung Jeon;Hongsun Park

Forecasting the market size of industries is an essential procedure for economic analysis in order to determine whether to invest in Research & Development (R&D) at an early stage of a business. Many studies have been conducted to investigate the relationship between R&D and patents, as patents are often the result of R&D activities.In this research, we predict the future size of the market using the number of patent applications. First, we predict the number of patents by modeling the time series of patents using AutoRegressive Integrated Moving Average (ARIMA) and Exponential Smoothing Method (ETS) based on the hierarchical structure. We use Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to derive a model with high prediction accuracy among the proposed models.Second, by analyzing the past relationship between the time series of patents and the time series of each market size using regression models, we predict the market size of each industry. We evaluate the R2 and p-value between the regression models and the market size as a dependent variable. As a result, regression models of most industrial sectors are statistically significant, indicating that patent applications can be an appropriate variable to predict the future size of the market. These regression models can serve as a reference to determine the investment on R&D by industry or as a tool to evaluate the future economic value of patents.The regression model also predicts a continuous growth pattern of the market size. The Medicine industrial sector shows the highest growth rate among the selected group.

Reconciliation Of Structured Time Series Forecasts With Graphs

Presenter: Mitchell O'hara-Wild
Co-authors: Mitchell O'hara-Wild;Rob Hyndman;George Athanasopoulos

Large collections of time series are often constrained to be coherent; for example, national forecasts should equal the sum of state forecasts. Forecast reconciliation algorithms enforce these constraints onto forecasts of these series. Hierarchical constraints are typically visualised with polytrees, where each series is represented by a node and edges connect series to their disaggregated child series. Grouped constraints are often shown as multiple disjoint polytrees, with each polytree showing a different order of disaggregation by the grouping variables. On the other hand, forecast reconciliation computation is typically done using matrix algebra, where matrices are used to encode the linear constraints.I propose using directed acyclical graphs (DAGs) to both visualise the constraints, and to facilitate forecast reconciliation computation. Using DAGs to represent the structure of a coherent collection of time series enables more flexible reconciliation structures than those possible in hierarchical and grouped designs. Graph structures can represent partial reconciliation via disjoint graphs, remove redundant aggregation with unbalanced trees, and allow sparse aggregation constraints from different levels of disaggregated series. Utilising a graph structure to describe the coherency of a time series also enables improved interfaces for analysing specific areas of a hierarchy.This talk will discuss how graphs can be used to represent a wide variety of coherent time series structures and demonstrate the advantages of using them in data exploration and forecast reconciliation.

New Product Life-Cycle Forecasting With Temporal Hierarchies

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

Predicting new product sales with life-cycle curves has traditionally been applied to data of low-frequency nature, e.g., annual or quarterly observations. However, with big data, companies now often have access to sales data at higher frequencies, e.g., weekly or daily. While this naturally provides more data points, it can introduce seasonality, a higher signal-to-noise ratio, and various irregularities. Although one can extend life-cycle curves, such as diffusion models, to capture seasonality, this can substantially increase model complexity and further complicate model parameter estimation. Furthermore, these additional high-frequency details can harm the long-term predictive performance of the life cycle curves.To address these issues, we suggest using temporal hierarchies that use optimal-suited time-series models at each aggregation level to extract model structure and subsequently combine it to increase predictive accuracy — for example, fitting a diffusion model at the quarterly level, with a long-term focus, and a seasonal exponential smoothing model at the weekly level, with a short-term focus. Combining these hierarchically results in a prediction that retains both aspects. Another benefit of using temporal hierarchies is that it allows for obtaining valuable model parameters. For example, the innovator and imitator coefficients of the Bass model can be used to draw insights into the underlying diffusion process. We demonstrate the approach's usefulness from a large computer manufacturer's dataset and provide insights on how to put these models into practice.

Applied Bayesian Neural Networks In Credit Risk

Presenter: Luis Javier Espinosa Rios
Co-authors: Luis Javier Espinosa Rios;María Rosa Nieto Delfín

The development of the financial system has been driven by the improvement of information technology. This has led financial institutions to implement high-performance and complex methodologies that contribute to risk analysis and management. The adoption of sophisticated algorithms such as Machine Learning, aims to execute these tasks. This paper explores the supervised learning algorithm in credit risk in three databases: 1) credit cards issued by a commercial bank, 2) mortgage loans and 3) LGD of financial assets. For this, the Bayesian Neural Networks (BNN) algorithm is evoked to train and estimate the probability of default. To strengthen the operation of the algorithm, the Markov Chain Monte Carlo (MCMC) method is implemented for the evaluation of the posteriori probabilities. In addition, we studied

the performance of the BNNs under different prior distribution functions, and given the importance of the activation functions, different functions are used for the hidden layer and final layer. The results obtained allow us to conclude that, for the task of classifying and estimating the probability of default, the BNNs give a robust confidence interval and allow credits to be classified correctly, making their configuration ideal and replicable. Machine learning is a vigorous field of research, where its application to complex tasks is plausible given computational progress and the mathematical assumptions of the model. The contrast between the econometric models can provide an answer to these unobservable phenomena.

A Large Bayesian Var Of The Colombian Economy: An Application Based On Macro Data Sets Augmented With Google Trends.

Presenter: Hector Zarate
Co-authors: Hector Zarate

Vector autoregressive, VAR, models are the benchmark for macroeconomic forecasting and structural analysis. Although VARs with more than two dozen dependent variables improve inference, it faces challenges related to having more parameters than observations and involving intensive computing due to the manipulation of large matrices. In this paper, we rely on large Bayesian models that impose a range of prior distributions on the autoregressive coefficients and the covariance parameters to conduct a forecasting and structural exercise of the Colombian economy. The data source is built on traditional public macroeconomic indicators extended with Google Trends on a quarterly frequency. Finally, the impulse response inference indicates that macroeconomic responses to some shocks seem to be suitable

Gradient Boosting Is Bayesian? Lessons From Using A Lightgbm Forecasting Model For Workers Compensation Insurance Ratemaking

Presenter: Luke Farley
Co-authors: Luke Farley;Robert Robison

Every year, Worker's Compensation Insurance rates are estimated for every job-class and state. These rates predict the medical and lost-work losses from all on the job injuries that occur in a given year. These costs are difficult to predict because historical loss data only gradually becomes available. Even 10 years after an injury, it's not uncommon for additional medical costs to crop up. However, a significant portion of the variance is explained by the report that comes out at the end of the year when injuries actually occurred.We used a LightGBM forecasting model to estimate the rate of insurance loss per exposure for thousands of job class and state combinations. The model used tweedie loss, was weighted by the total exposure of the state and job class, and enforced monotonicity on key features. A global model was chosen to enable pooling information across states and job classes, and to allow inclusion of a new data source the client had just obtained. After testing hundreds of potential features, we ultimately found the most consistently accurate model used only three features: the exposure for the state/job-class and the historical rate of insurance loss over exposure for the state/job-class and for the job-class nationally.At low exposures, the model relied more on the historical national job-class rate. As the exposure in a state/job-class increased, the predictions slowly migrated to following the historical state/job-class rate. It effectively operated as a hierarchical Bayesian model would in these circumstances, but in a fraction of the time and with much less input. It indicates that well formed gradient boosting models have the ability to pool information efficiently at large scales, opening the door for more performant global forecasting models to many domains.

Cross-Country Cross-Technology Digitalisation: A Bayesian Hierarchical Model Perspective

Presenter: Vincent Labhard
Co-authors: Vincent Labhard;Charles Hoffreumon

In this article, we present a new perspective on forecasting technology adoption, focused on the extensive margin of adoption of multiple digital technologies in multiple countries. We do this by applying a Bayesian hierarchical structure to the seminal model of technology diffusion. After motivating the new perspective and the choices of priors, we apply the resulting framework to a cross-continental data set for EU and OECD countries and different digital technologies adopted by either households/individuals or by businesses. The results illustrate that the Bayesian hierarchical structure may be used to assess and predict both the adoption process and the uncertainty surrounding the data, and is robust to the use of alternative priors. They point to heterogeneity across countries and across technologies, mostly in the timing of adoption and, although to a lesser extent, the steady-state adoption rate once technologies are fully diffused. This suggests that characteristics of countries and technologies matter for technology diffusion.https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2700~08fcb49cd5.en.pdf

Mixture Distributions In Collaborative Probabilistic Forecasting Of Disease Outbreaks

Presenter: Spencer Wadsworth
Co-authors: Spencer Wadsworth;Jarad Niemi

Collaboration among multiple teams has played a major role in probabilistic forecasting events ofinfluenza outbreaks, the COVID-19 pandemic, other disease outbreaks, and in many other fields.When collecting forecasts from individual teams, ensuring that each team's model represents forecastuncertainty according to the same format allows for direct comparison of forecasts as well asmethods of constructing multi-model ensemble forecasts. This paper outlines several common probabilisticforecast representation formats including parametric distributions, sample distributions, bindistributions, and quantiles and compares their use in the context of collaborative projects. We proposethe use of a discrete mixture distribution format in collaborative forecasting in place of otherformats. The flexibility in distribution shape, the ease for scoring and building ensemble models,and the reasonably low level of computer storage required to store such a forecast make the discretemixture distribution an attractive alternative to the other representation formats.

Weighted Kernel Scores

Presenter: Xiaochun Meng
Co-authors: Xiaochun Meng

Multivariate probabilistic forecasting is intriguing and challenging due to its inherently complex nature and computational difficulty. In some applications, directional distributions are of interest, where each directional distribution is the univariate distribution of the projection of the multivariate distribution onto a specific direction. Examples of such applications include hierarchical forecasting and financial portfolio management. Proper scoring rules are essential for accurate estimation and evaluation in probabilistic forecasting, as they encourage honest assessment. This paper focuses on kernel scores, which are less well-known and somewhat mysterious compared to other scores such as the logarithmic score. We provide a novel intuitive interpretation linking kernel scores and directional distributions, which allows for a weighting scheme to be considered that emphasizes directional distributions for certain directions. Through empirical analysis using financial and electricity smart meter data, we demonstrate that the proposed kernel scores lead to improved forecasting accuracy.

Probabilistic Forecasting With Factor Quantile Regression: Application To Electricity Trading

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

This paper presents a novel approach for constructing probabilistic forecasts, which combines both the Quantile Regression Averaging (QRA) method and the Principal Component Analysis (PCA) averaging scheme. The performance of the approach is evaluated on datasets from two European energy markets - the German EPEX SPOT and the Polish Power Exchange (TGE). The results indicate that newly proposed solutions yield results, which are more accurate than the literature benchmarks. Additionally, empirical evidence indicates that the proposed method outperforms its competitors in terms of the empirical coverage and the Christoffersen test. In addition, the economic value of the probabilistic forecast is evaluated on the basis of financial metrics. We test the performance of forecasting models taking into account a day-ahead market trading strategy that utilizes probabilistic price predictions and an energy storage system. The results indicate that profits of up to 10 EUR per 1 MWh transaction can be obtained when predictions are generated using the novel approach.

Moteef: Metrics Based On Optimal Transport For Explaining Epidemic Forecasts

Presenter: Lanyin Zhang
Co-authors: Lanyin Zhang;Srini Venkatramanan

Probabilistic forecasts of epidemic trajectories obtained from multi-model ensembles have been very influential for policy making during the COVID-19 pandemic as well as seasonal influenza. While several standardized evaluation metrics exist (for example, Weighted Interval Score) for quantifying the performance of such methods, limited attention has been given to interpreting the change in forecasts over time as well as the level of disagreement across the component models during different periods. In this work, we develop metrics to investigate forecast updates and diversity based on Wasserstein Distance (WD), a distance metric between probability distributions from the mathematical field of optimal transport. We use data from the ongoing CDC-coordinated Flu Forecasting exercise for the 2022-23 influenza season in the United States, which contains 1-4 week ahead hospitalization forecasts submitted by 28 modeling teams over 24 weeks for 50 states and national level in the United States. We quantify per-model velocity (average change over the forecast horizons) as well as ensemble diversity (average level of disagreement across the component models). For the latter, we employ a WD-based barycenter of the different models as well as network-based clustering approaches. Through the velocity metric, we note that the ensemble balances between the most-sensitive and least-sensitive models. We also note that the overall velocity of the ensemble changes based on the current ground truth, capturing the underlying uncertainty. Finally, through the ensemble diversity metric, we note that the overall disagreement among component models increases at times of high uncertainty, as expected. We believe that our work provides a robust metric to qualitatively and quantitatively explain the behavior of epidemic forecast ensembles. Such methods can also be adapted across domains where there are multiple probabilistic forecasts (either over time or across models/users).

Sparse Solutions For A Complex World: Determining Optimal Variable Transformations For High-Dimensional Macroeconomic Forecasting

Presenter: Tim Reinicke
Co-authors: Tim Reinicke;Maurizio Daniele;Philipp Kronenberg

Determining the most informative economic variables and their optimal transformations remains an open challenge in macroeconomic forecasting, particularly in high-dimensional data-rich environments. This

study investigates the potential of incorporating sparsity and diverse data transformations to enhance forecast accuracy within the context of the US FRED-MD data set. By conducting an extensive out-of-sample forecast exercise, we assess the performance of various dynamic factor models incorporating distinct sparsity implementations for forecasting key macroeconomic indicators. We compare standard dynamic factor models with specifications utilizing a variable pre-selection, and sparse dynamic factor models that incorporate sparsity directly in the estimation process. We evaluate the performance of these models with machine-learning models such as lasso, elastic-net, and random forests. Our findings reveal that sparsity and the inclusion of diverse data transformations enhance forecast accuracy. Furthermore, sparsity facilitates the identification of key driving factors, thereby enhancing interpretability.

Maximally Machine-Learnable Portfolios

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

When it comes to stock returns, any form of predictability can bolster risk-adjusted profitability. We develop a collaborative machine learning algorithm that optimizes portfolio weights so that the resulting synthetic security is maximally predictable. Precisely, we introduce MACE, a multivariate extension of Alternating Conditional Expectations that achieves the aforementioned goal by wielding a Random Forest on one side of the equation, and a constrained Ridge Regression on the other. There are three key improvements with respect to Lo and MacKinlay's original maximally predictable portfolio approach. First, it accommodates for any (nonlinear) forecasting algorithm and predictor set. Second, it handles large portfolios. Lastly, the predictive algorithm is a black box, dissimulating the successful strategy to other market participants. We conduct exercises at the daily and monthly frequency and report significant increases in predictability and profitability using very little conditioning information. Interestingly, predictability is found in bad as well as good times, and MACE successfully navigates the debacle of 2022.

Reservoir Computing For Macroeconomic Forecasting With Mixed Frequency Data

Presenter: Giovanni Ballarin
Co-authors: Giovanni Ballarin;Petros Dellaportas;Lyudmila Grigoryeva;Marcel Hirt;Sophie Van Huellen;Juan-Pablo Ortega

Macroeconomic forecasting has recently started embracing techniques that can deal with large-scale datasets and series with unequal release periods. The aim is to exploit the information contained in heterogeneous data sampled at different frequencies to improve forecasting exercises. Currently, MIxed-DAta Sampling (MIDAS) and Dynamic Factor Models (DFM) are the two main state-of-the-art approaches that allow modeling series with non-homogeneous frequencies. We introduce a new framework called the Multi-Frequency Echo State Network (MFESN), which originates from a relatively novel machine learning paradigm called reservoir computing (RC). Echo State Networks are recurrent neural networks with random weights and trainable readout. They are formulated as nonlinear state-space systems with random state coefficients where only the observation map is subject to estimation. This feature makes the estimation of MFESNs considerably more efficient than DFMs. In addition, the MFESN modeling framework allows to incorporate many series, as opposed to MIDAS models, which are prone to the curse of dimensionality. Our discussion encompasses hyperparameter tuning, penalization, and nonlinear multistep forecast computation. In passing, a new DFM aggregation scheme with Almon exponential structure is also presented, bridging MIDAS and dynamic factor models. All methods are compared in extensive multistep forecasting exercises targeting US GDP growth. We find that our ESN models achieve comparable or better performance than MIDAS and DFMs at a much lower computational cost.

Deep Learning With Non-Linear Factor Models: Adaptability And Avoidance Of Curse Of Dimensionality

Presenter: Maurizio Daniele
Co-authors: Maurizio Daniele;Mehmet Caner

In this paper, we connect deep learning literature with non-linear factor models and show that deep learning estimation makes a substantial improvement in the non-linear additive factor model literature. We provide bounds on the expected risk and show that these upper bounds are uniform over a set of multiple response variables by extending Schmidt-Hieber (2020) theorems. We show that our risk bound does not depend on the number of factors.In order to construct a covariance matrix estimator for asset returns, we develop a novel data-dependent estimator of the error covariance matrix in deep neural networks. The estimator refers to a flexible adaptive thresholding technique which is robust to outliers in the innovations. We prove that the estimator is consistent in spectral norm. Then using that result, we show consistency and rate of convergence of covariance matrix and precision matrix estimator for asset returns. The rate of convergence in both results do not depend on the number of factors, hence ours is a new result in the factor model literature due to the fact that number of factors are impediment to better estimation and prediction. Except from the precision matrix result, all our results are obtained even with number of assets are larger than the time span, and both quantities are growing.Various Monte Carlo simulations confirm our large sample findings and reveal superior accuracies of the DNN-FM in estimating the true underlying functional form which connects the factors and observable variables, as well as the covariance and precision matrix compared to competing approaches. Moreover, in an out-of-sample portfolio forecasting application it outperforms in most of the cases alternative portfolio strategies in terms of out-of-sample portfolio standard deviation and Sharpe ratio.

The Impact Of Supply Chain Disruptions On Business Expectations During The Pandemic

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

Utilizing the Federal Reserve Bank of Atlanta's Business Inflation Expectations (BIE) survey, which has been continuously collecting subjective probability distributionsover own-firm future unit costs since October 2011–we build a measure of firms' aggregated marginal cost expectations (a key determinant in price-setting behavior).We document two facts about firms' marginal cost expectations and risk during the COVID-19 pandemic. First, in the early months of the pandemic, firms, on net, sawCOVID-19 largely as a demand shock and lowered their one-year ahead expectations. However, as the pandemic wore on, firms' one-year ahead unit cost expectations rose sharply alongside their views on supply chain and operating capacity disruptions. Second, entering into the pandemic, the balance of unit cost risks were weighted to the downside, as more weight was assigned to the lowest two bins in the five-bin distribution. By December 2022, however, upside cost risks had sharply outweighed the potential for perceived downside risks over the year ahead. We find that both positive demand shocks (e.g. large order backlogs) and negative supply shocks (e.g. long supplier delivery times and labor shortages) have contributed to elevated short-term unit cost expectations and risk.

Global Inflation Persistence And Inflation Forecasting During The Covid-19 Pandemic

Presenter: María José Orraca Corona
Co-authors: María José Orraca Corona;Karla Neri Hernández;Lenin Arango-Castillo

Inflation persistence is understood as the speed at which inflation returns to its long-term levels after a shock. Evidence suggests a positive correlation between inflation level and its persistence, which may be concerning from a monetary policy perspective as it may delay the convergence of inflation to its target. On the other hand, the literature argues that a more persistent time series is easier to forecast, meaning that periods of high inflation persistence would mean more precise forecasts, so they have higher credibility. Since the second half of 2021, the world has undergone a period of persistently elevated (and rising) inflation. At the same time, forecasting errors have been high. This paper studies the increase in inflation persistence during the COVID-19 pandemic across countries and evaluates if this phenomenon has improved forecasts. To do so, we estimate the evolution of headline and core inflation persistence for a sample of emerging and advanced economies using rolling windows of different sizes between 2002 and 2022. First, we document that inflation persistence has increased across countries. By comparing the evolution of persistence across countries and through descriptive exercises, we argue that the increase in persistence is a global phenomenon plausibly driven by external factors associated with the accumulation of subsequent shocks of considerable magnitude. However, we also document that this increase has been accompanied by an increase in inflation expectations and wages, which could further pressure persistence upwards. Second, we analyze if a higher persistence implies simpler models achieve a more accurate forecast. Our results indicate that higher inflation persistence has not implied more accurate forecasting during the pandemic. This result could be a feature specific to this period because it has been very atypical and uncertain since the global economy has been subject to several unexpected shocks, which may have counteracted the positive impact of higher persistence on forecasting performance. Our results suggest that the accuracy of forecasts depends on the properties of the time series and the economic context where these forecasts are performed.

Inflation Nowcasting In Persistently High Inflation Environments

Presenter: Richard Schnorrenberger
Co-authors: Richard Schnorrenberger;Aishameriane Schmidt;Guilherme Valle Moura

Recent episodes in global markets have shown that inflationary waves can unfold extremely fast while leading to considerable macroeconomic uncertainty as spiraling inflation expectations become a real threat. In this scenario, real-time inflation nowcasts are of utmost importance for a more timely reaction of monetary policy. To that end, we investigate the predictive content of high-frequency macro-financial indicators to nowcast monthly inflation in an environment characterized by persistently high inflation rates, namely the Brazilian economy of the past decades. Using machine learning methods within a Mixed-Data Sampling (MIDAS) structure, we identify two key elements that produce accurate inflation nowcasts on a weekly basis. First, a combination of shrinkage methods with timely price indicators and daily experts' forecasts delivers considerable gains in rising inflation periods compared to the median of market expectations. In particular, shrinkage-based models can better nowcast the inflation surge following the Covid-19 pandemic. Second, a direct forecasting approach that adjusts model specification based on the forecast horizon and taking into account the real-time flow of data releases leads to higher quality nowcasts.

Performance Of The Michigan Survey Of Consumer Inflation Expectations Under Asymmetric Loss: Comparison Of The Pre-Pandemic And Pandemic Periods

Presenter: Yasemin Ulu
Co-authors: Yasemin Ulu;Matthew Higgins

We examine the forecast rationality of consumer inflation expectations from the Michigan Survey under asymmetric loss functions using the methodology proposed by Elliott, Komunjer and Timmermann (2005). We utilize a rolling sample window to allow the value of the asymmetry parameter to change over time. We find that during the pandemic under predictions were more highly penalized, as indicated by asymmetry parameter values close to one, compared to the pre-pandemic period. The forecasts of inflation during the pandemic, however, were still systematically below the realized values. This shows that even

with the overprediction bias, consumers failed to anticipate the high levels of inflation associated with the pandemic. We also apply a Fluctuation Rationality Test proposed by Rossi and Sekhposyan (2016) to account for possible instabilities due to structural breaks. We still find evidence of asymmetric loss.

Climate Risks And State-Level Stock-Market Realized Volatility

Presenter: Oguzhan Cepni
Co-authors: Oguzhan Cepni;Matteo Bonato;Rangan Gupta;Christian Pierdzioch

We analyze the predictive value of climate risks for state-level realized stock-market volatility, computed, along with other realized moments, based on high-frequency intra-day U.S. data (September, 2011 to October, 2021). A model-based bagging algorithm recovers that climate risks have predictive value for realized volatility at intermediate and long (one and two months) forecast horizons. This finding also holds for upside ("good") and downside ("bad") realized volatility. The benefits of using climate risks for predicting state-level realized stock-market volatility depend on the shape and (as-)symmetry of a forecaster's loss function.

Problems With Machine Learning, High-Dimensional Data And Forecasting Stock Returns

Presenter: Erik Mekelburg
Co-authors: Erik Mekelburg;Jack Strauss

Using a multi-level ensemble model design, we forecast international market returns with a novel high-dimensional data set of aggregated cross-sectional predictors. We first use LASSO, Ridge Regression, Elastic-Net, Multi-layer Neural Net and Random Forest methods to forecast U.S stock returns using over 350 predictors. We encounter substantial problems of model uncertainty and parameter instability that contribute to stark and sudden forecast failures due to model overfitting and structural change. Several machine learning forecasts lead to very large negative out-of-sample R2. Allowing the window size to vary does not effectively reduce the forecast failures. We then introduce ensemble procedures that weight the machine learning forecasts by recent performance and shrink the forecasts to the historical average. Both procedures however do not effectively boost predictability. The failures occur quickly, so an exponential-weighted moving average procedure or Dynamic Bayesian Model Average methods do not quickly enough identify better performing models. Shrinkage to the mean does not sufficiently improve the very large forecast failures. Results additionally show popular methods to induce sparsity such as Principal Components and Partial Least Squares does not adequate summarize the data since variance contribution changes dramatically over the time horizon, making model training inconsistent as the weighting of the individual components change dramatically over-time. Nonlinear machine learning methods further do not consistently extract useful information when there is model uncertainty, parameter instability and dense data.Our contribution is to show a real-time winsorization of the forecasts effectively reduces extreme forecasts and significantly improves predictability. We provide evidence that it is important to focus on sources of forecast failure including parameter instability, model uncertainty, and nonlinearities and shed light on the sparsity/density debate in the stock return forecasting dialogue. To highlight the robustness of our approach, we forecast three international stock market markets (UK, DE, CA) as well as apply the Goyal-Welch dataset. The approach is further applied to an interest rate forecasting problem using a high dimensional macroeconomic dataset from the St Louis Federal Reserve.

Forecasting Expected Shortfall With Multiple Quantiles

Presenter: Pit Goetz
Co-authors: Pit Goetz;Joerg Laitenberger

The Expected Shortfall (ES) of a distribution of returns at a confidence level ,which is defined as the expectation of the losses beyond the corresponding Value-at-Risk (VaR), can be approximated as the sum of appropriately chosen quantilesof the distribution of the losses beyond . In this paper we evaluate empiricallythe performance of ES forecasting models based on the forecast of multiple VaRs.We compare models with equal weights and with weights optimized according to asuitable scoring function, and we explore the number of quantile levels required toobtain reasonable results. The VaR forecasts are performed using quantile regressions(CAViAR). In an empirical study we apply the models, employing 2,3 and 5VaR forecasts, to different datasets, including index, stock, forex and crypto dataover different time periods with and without market turmoils. The forecasts are evaluatedusing adequate ES–backtests and we compare the performance of the forecaststo standard models from the literature. We observe that the inclusion of a singleadditional quantile with a level above the considered ES level can lead to significantimprovements in forecast performance and model applicability. Our results indicatethat ES can be very safely forecasted with models using two or three VaR forecastsand using simple equal weights, when the forecasted VaRs used in the constructionof the ES are themselves reliable.

Line Components In Trading Volume And Realized Volatility Time Series: Effects On The Estimation Of The Long-Range Dependence Parameter

Presenter: Lenin Arango-Castillo
Co-authors: Lenin Arango-Castillo;Glen Takahara

In this paper, under the Gaussian assumption and using spectral analysis methods, we analyze trading volume and realized volatility time series based on data available at the Center for Research and Security Prices (CRSP). Our goal is to detect the presence of line components and measure their effects on estimating the Long-Range Dependence (LRD) parameter, H. The LRD parameter is an essential input in quantities of interest in finance, and Gaussian LRD modeling continues to be used in the economics and finance literature. First, we investigate the presence of periodicities in daily time series of trading volume and realized volatility. We detect and remove periodicities (line components) using spectral analysis methods. Periodicities are of interest because (i) it has been shown that they play an important role in forecasting volatility, and those models taking periodicities into account have better forecasting performance, and (ii) they might adversely affect estimators of the LRD parameter. Second, we investigate the effects of periodicities on several estimators of H for the trading volume and realized volatility time series. Third, within the class of Gaussian LRD processes, we estimate the LRD parameter based on a test statistic that determines the type of process: either a fractional Gaussian noise (fGn) or fractionally integrated noise with Gaussian innovations (GFI). Our results indicate, in instances where the removal of line components has a pronounced effect on local estimators of H (Geweke and Porter-Hudak; local Whittle; and Hou and Perron), that the removal of line components brings these estimators closer to the parametric Whittle estimator designed from the model specification (fGn or GFI), chosen by the test.

Forecasting Migrant Encounters At The Southwest Border: Leveraging Quantitative And Qualitative Insights To Maximize Forecasting Accuracy

Presenter: Justin Schon
Co-authors: Justin Schon;Douglas Baals;Nadwa Mossaad

How many migrants will attempt to cross the United States Southwest land border with Mexico in the next 6 months? The Office of Immigration Statistics (OIS) in the Department of Homeland Security (DHS) seeks to answer this question with a mixed methods approach that provides new forecasts each month. First, OIS estimates separate Bayesian Structural Time Series (BSTS) models for 33 country-family type groups (top 10 countries plus all other, each broken down into Single Adults, Family Units, and Unaccompanied Children). Then, predictions from the BSTS models are adjusted based on insights into

legal and policy dynamics by subject matter experts from across DHS components – Customs and Border Patrol (CBP), Immigration and Customs Enforcement (ICE), Office of Policy, and Federal Emergency Management Agency (FEMA). This mixed methods approach maximizes forecasting accuracy during periods that follow normal historical patterns and during periods with anticipated migration surges or sharp reductions. The final predictions are rigorously validated against actual migrant encounter numbers to evaluate forecasting accuracy. Overall, OIS migrant encounter predictions have an average mean absolute percent error (MAPE) under 10% one month ahead – with the greatest accuracy for Single Adults and Mexican nationals – with growing error further into the future, an unprecedented accuracy given the high uncertainty around these migration flows.

Bigquery's Arima plus With External Regressors

Presenter: Haoming Chen
Co-authors: Haoming Chen;Xi Cheng;Honglin Zheng

This talk introduces the ARIMA with external regressors forecasting model built on BigQuery, Google Cloud's fully-managed, serverless data warehouse. The model is designed to provide accurate and scalable predictions for a wide range of use cases. The talk covers the key features of the new BigQuery forecasting model, including its ability to handle large datasets, integration with other Google Cloud services such as Data Studio. The talk also provides a detailed overview of the model's architecture and how it can be implemented in a production environment. In addition, the talk walks through some critical user journeys including creating a model, doing forecasting and evaluating the result, with examples of how the new forecasting model has been used in real-world scenarios.

Boosted Learning On Level Imbalance Data Through Hierarchical Data Augmentation

Presenter: Weijie Shen
Co-authors: Weijie Shen;Steve Thomas;Casey Lichtendahl;Haoyun Wu

In real world applications, the hierarchical and/or group structure of time series data can be deep and complicated, which often forms a pyramid structure in terms of series counts by levels: with a large amount of series on the low level that are less important, but only a handful of top level series that are much more essential to business. This level imbalance presents a big challenge for deep learning methods which relies on uniformity and representability of the empirical data. Scaling and reweighting might not be sufficient to boost overall performance. In this talk, we present a data augmentation framework based on the hierarchical structure to improve the learning during training, especially on higher level series. Results show not only improved performance on higher level series but also on lower level series due to regularization from the augmented series.

Fast And Accurate Deep Forecasting On Vertex With Probabilistic Loss Functions

Presenter: Rajat Sen
Co-authors: Rajat Sen;Abhimanyu Das;Weihao Kong;Andrew Leach

In the pursuit of fast and scalable long horizon forecasting, we propose two improvements built into Vertex AI Forecasting. First, we introduce TiDE (Time-series Dense Encoder) which is a simple multi-layer perceptron based encoder decoder architecture that is up to 10x faster in training time and inference latency. At the same time, TiDE can provide up to 40% improvement in forecasting accuracy on popular traffic forecasting benchmarks over transformer based methods. Second, we advocate the use of carefully designed

probabilistic loss functions that minimize the maximum likelihood over a family of mixture distributions. These loss functions are better suited to handle different data modalities like sparse count data, and data with non sub-Gaussian tails, leading to 6% improvement on the M5 forecasting competition metric. During inference, we can adapt the model to the target metric of interest by predicting the appropriate statistic from the learnt predictive distribution. Finally, both these improvements are complimentary in nature and can be built into the same model.

Combining Econometric And Delphi Forecasts Of Tourism Demand In Time Of A Major Crisis

Presenter: Hanyuan Zhang
Co-authors: Hanyuan Zhang;Shanshan (Vera) Lin;Haiyan Song

Compared to the extensive research of forecasting combination in other fields, little research has comprehensively examined the effect of combining judgmental and econometric forecasting methods in the tourism literature. Prior research shows that forecast combinations outperform single models because they can reduce the risk of forecasting failure. The judgmental forecasting approaches are useful for overcoming the limitations of statistical methods, especially in times of crises. To take full advantage of the strengths of these two methods, this study aims to combine econometric models with Delphi forecasting methods in predicting quarterly outbound tourist arrivals from mainland China to Hong Kong and Macao SARs over 2023Q1-2027Q4. Both times-series models such as exponential smoothing with state space models, and ARIMAX) and econometric models (ARDL-ECM, and TVP) will be used to generate the statistical forecasts. A two-round Delphi survey will be carried out to collect experts' forecasts from tourism academics and practitioners. Both equal and unequal weighting schemes will be adopted to combine the statistical and judgmental forecasts. The MAPE and RMSPE error measures will be used to evaluate the forecasting accuracy among different combined forecasts and individual forecasts. It is expected that the combined forecasts will improve the forecasting accuracy relative to its constituent forecasts by considering the impacts of a major crisis. This study contributes to the tourism demand forecasting literature by proposing a novel and effective combined forecasting method post pandemic.

Judgment And Simple Rules– The Effect Of Information Quality And Decision-Autonomy On Forecasting Performance

Presenter: Sebastian Achter
Co-authors: Sebastian Achter;Kai Mertens;Matthias Meyer

Our study explores the effect of information quality and decision-autonomy among other work environment factors on forecasting performance.A vast body of literature is concerned with enhancing forecasting accuracy by improving statistical and judgmental forecasting techniques. However, the effects on forecasting and organizational performance are still ambiguous. While many advancements of forecasting methods are proposed, forecasting practices with simple methods are still largely present.Research has only recently begun to approach forecasting practices in organizations as a whole to understand the ability to cope with demand uncertainty. This call is substantiated by the claim that distinct disciplines and industry sectors, such as supply chain forecasting, deal with subject-specific issues beyond methodical questions. Especially the role of human judgment and intervention is challenging to explore outside their organizational context, where the forecasting practice is often part of a larger decision-making problem. We conducted a multi-method single case study in a semiconductor supply chain company, examining the forecasting practice of demand planners within the broader decision-making context of the S&OP process. We collected data through employee surveys, interviews, and formal process analysis. We focus on understanding how information quality as input for creating forecasts and work environment factors, defined by task difficulty, decision-making autonomy, role ambiguity, controllability, and self-efficacy, affect forecasting performance.Our results suggest that information quality and decision-autonomy are positively

associated with forecast performance. Specifically, we find that low information quality resulting from market volatility and product characteristics can be offset by high decision autonomy. We further observe that decision-autonomy of individual planners is expressed in a practice consisting of simple forecasting rules and the ability of exception handling embedded in an advanced planning tool. Under these conditions, the judgment of demand planners can result in high forecasting performance, although information quality is low. Overall, our results stress the vital role of human intervention and judgment in its organizational context.

Relative Performance Of Scenarios For Forecasting The Success Of Megaprojects

Presenter: Konstantia Litsiou
Co-authors: Konstantia Litsiou;Yiannis Polychronakis;Stylianos Sapountzis;Konstantinos Nikolopoulos

Megaprojects impact people's lives and attract several stakeholders' interest. They deliver benefits to communities for social good. In this research, we examine the role of scenarios in forecasting the success of megaprojects. Participants are asked to generate their forecasts (scenarios) in a narrative form from stakeholders' perspectives. To what extent the socio-economic benefits are expected to be realised is the key point for this study.

Transferability Of Neural Forecast Methods

Presenter: Kin G. Olivares
Co-authors: Kin G. Olivares;Cristian Challu;Federico Garza;Max Mergenthaler

Transfer learning refers to pre-training a flexible model on a large dataset and using it later on another dataset with little to no training. It is one of the most outstanding achievements in Machine Learning and has many practical applications, yet for time series forecasting, the technique is under-explored. Studying the transferability of neural forecast methods constitutes an exciting line of research, as it would allow neural forecasting methods' to produce lightning-fast predictions under a fraction of the computational cost, solving one of the most significant limitations of ML methods: the tradeoff between accuracy and speed. We propose systematically evaluating neural forecasting architectures to compare their accuracy on a broad range of transfer learning forecasting tasks and carefully study transferability's enabling conditions.

Analysis Of Consumer Preferences For New Electric Vehicle Technologies: Can Future Vehicle Steering System Steer Consumer's Purchase Intention?

Presenter: Woojae Kim
Co-authors: Woojae Kim;Youngsang Cho

As the mainstream of vehicle industry is changing from internal combustion engine vehicle (ICEV) to electric vehicles (EVs), consumer preference for new vehicle technology of EVs needs to be analyzed in multi-perspective. Therefore, this study proposes new framework for consumer preference analysis, which is two-stage choice experiment, and we performed an empirical analysis on the four-wheel independent steering system (4-WISS). In the first stage, we analyzed consumer preference of availability of 4-WISS in EV. In the second stage, we investigated consumer preference of the specific technology of 4-WISS in EV. Stated preference data was collected through survey for the choice experiment and we used mixed logit model to analyze consumers' preference and willingness-to-pay quantitatively. The results show that consumers are willing to pay KRW 19,045,098.6 (USD 13,513.9) to use 4-WISS in high-end EV. Furthermore, consumers who are female, older and who have more knowledge of autonomous vehicle prefer 4-WISS more. We found that consumers prefer the parallel movement among 4-WISS modes, display touch for a tool for selecting 4-WISS, steering wheel for operating 4-WISS mode and gauge for displaying information. Through the

results, we discuss about the new business model in vehicle industry and automakers are able to establish effective strategy for the initial market for EV with 4-WISS.

Forecasting The Likelihood Of Ai Dominance And The Governance Of Agi

Presenter: Claudio Antonini
Co-authors: Claudio Antonini

The complexity of the Human-AI future relationship stems from its many branches and levels. In this work, firstly, we present a general framework to capture many of the parameters used to describe the interaction and, secondly, discuss approaches to govern Artificial General Intelligence (AGI).Firstly, we combine several of the Human-AI manifestations and activities in an equation with 5 factors (similar in spirit to Frank Drake's equation to calculate the number of transmitting civilizations in the Milky Way) to estimate the likelihood of AI dominance. We indicate sources that help quantify values or ranges for each of these factors, and evaluate how previous methodologies approached the issue.Secondly, we propose a framework for governing AGI based on the lessons learned from the governance of nuclear activities in the past century.Nuclear technologies (initially for military and later for civilian applications) were rather successfully and actively managed by reducing the number of institutions and companies involved in their development, controlling its diffusion through secrecy and military installations, and heavy regulation. Public fear and pressure on elected officials, and the inherent danger, difficulty, and cost of working with radioactive material, also created natural barriers to its propagation.Even with these measures in place, incidents and accidents occurred. However, one important lesson was learnt: large-scale accidents are unlikely to be caused by reduced-likelihood events, for example, an airplane hitting a nuclear facility. Rather, they happen by a series of small mistakes, such as design errors or misinterpreting instructions that can lead to a chain of events resulting in severe problems.Governing AGI will be more challenging due to the fact that many are working to develop it, it is not a secret technology, publishing is not always peer-reviewed, it can propagate at the speed of light, and companies are competing to capture market share, creating an AI arms race while observing and ignoring surprising model behavior that is well beyond the intended design. Additionally, AGI has the potential to actively look for new things to do, finding patterns that are opaque to humans - its most distinguishing and aggravating feature when compared to nuclear technologies.

Capturing International Influences In U.s. Monetary Policy Through A Nlp Approach

Presenter: Laurent Ferrara
Co-authors: Laurent Ferrara;Nicolas De Roux

The U.S. Federal Reserve has a statutory dual domestic mandate of price stability and full employment. In this paper, we question about the role of the international environment in shaping U.S. monetary policy decisions. In this respect, we apply natural language processing (NLP) techniques to minutes of the Federal Open Market Committee (FOMC) and construct indexes of the attention paid by U.S. monetary policymakers to the international economic and financial situation, as well as sentiment indexes. By integrating those text-based indicators into a Taylor rule, we derive various quantitative measures of the external influences on Fed decisions. Our results show that when there is a focus on international topics within the FOMC, the Fed's monetary policy generally tends to be more accommodative than expected bya Taylor rule. This result is robust to various alternatives including a time-varying neutral interest rate or a shadow central bank interest rate.

Financial Stress And Economic Activity: Evidence From A New Worldwide Index

Presenter: Davide Furceri
Co-authors: Davide Furceri

This paper uses text analysis to construct a continuous financial stress index (FSI) for 110 countries over each quarter during the period 1967-2018. The new indicator has a larger country and time coverage and higher frequency than similar measures focusing on advanced economies. It relies on a computer algorithm along with human expert oversite and is thus easy to maintain and update. And it complements existing binary chronologies in that it can assess the severity of financial crises. We use the indicator to assess the impact of financial stress on the economy using both country- and firm-level data. Our main findings are fourfold: i) consistent with existing literature we show an economically significant and persistent relationship between financial stress and output; ii) the effect is larger in emerging markets and developing economies where there is also evidence of a non-linear relationship absent in advanced economies; iii) we deal with simultaneous causality by constructing a novel instrument using information from the text analysis, and show that, while there is clear evidence that financial distress harms economic activities, OLS estimates tend to overestimate the magnitude of this effect; (iv) we confirm the presence of an exogenous effect of financial distress through a diff-in-diff exercise by showing that its effects are larger for firms that are more likely to depend on external finance.

What's The Role Of Perceived Oil Price Shocks In Inflation Expectations?

Presenter: Xuguang Simon Sheng
Co-authors: Xuguang Simon Sheng;Zidong An;Xinye Zheng

We identify the perceived oil price shock as well as perceived global demand and supply shocks using sign restrictions in a factor-augmented vector autoregression model that includes forecasts for crude oil price growth, real GDP growth, and inflation across 84 economies. The perceived oil price shock explains only 10% of the fluctuations, on average, in global inflation expectations from January 2012 to December 2022, and accounts for an even smaller fraction during the COVID-19 pandemic. Allowing for oil price noise shock – reflecting exogenous shifts in agents' optimism and pessimism – does not materially change the limited pass-through of the perceived oil price shock to inflation expectations. In contrast, perceived global supply and demand shocks dominate, especially since the onset of the pandemic. Over the first eight months, professional forecasters viewed the pandemic, on net, as a negative demand shock and lowered their short-term inflation expectations. In early 2021, professionals quickly switched their views and sharply increased their inflation expectations amid burgeoning and persistent supply chain disruptions and labor constraints.

Electricity Consumption Patterns And Price Volatility In Turkey: A Wavelet Analysis

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

For two decades, the now decentralized Turkish electricity market has been undergoing liberalization, similar to other electricity markets worldwide, and has essentially adopted the European market structure. Hourly electricity prices are determined by maximizing market surplus. These prices fluctuate wildly, owing to the nature of electricity as a commodity. Hourly electricity consumption data, on the other hand, exhibit a strong periodic pattern resulting from consumers' daily routines. These may be disturbed on "non-regular" days, such as weekends and holidays, making the pattern of electricity consumption less regular.Using wavelet analysis, this study uncovers a link between electricity consumption and its price

volatility in Turkey. In particular, we hypothesize that a low level of regularity of consumption translates into elevated price volatility. The empirical basis for this study consists of hourly electricity consumption and price data for Turkey during 2017-2022. We perform a wavelet analysis of the consumption data in order to capture the strength, or wavelet power, of periodic patterns across time, and investigate the effect of holidays and weekends on wavelet power. These wavelet power series, quantifying consumption regularity, serve as input to regression models explaining intraday price volatility. With this novel approach, we can show, for example, that an interruption of daily routines (on "non-regular" days) leads to higher electricity price volatility.

On The Benefits Of Synthetic Data And Transfer Learning For Solar Power Forecasting

Presenter: Joris Depoortere
Co-authors: Joris Depoortere

Accurate solar power forecasting is an increasingly important field given our changing power infrastructure. Deep learning models are employed more and more in the field of time series forecasting with increasing levels of success. But with it comes a need for high quality data, and lots of it, which is often not available in solar power forecasting.To solve the issue of limited data availability we often look at global models. These models are trained on a similar problem (the source domain), one where a lot of data is available, and are then applied to the field in question (the target domain). These models are typically deep neural networks, which can be directly applied or finetuned to the target domain. Another method to handle a lack of data is the use of synthetic data to train models. Algorithms exist to 'predict' the solar power output, given weather conditions at the time and meta-data of hypothetical solar panels used. In this manner, completely artificial solar power generation data can be produced to further train deep learning models for locations with only limited data. PVGIS, a tool made by the European Commission, does exactly that.This paper explores these two methods for improving solar power forecasts. Synthetic data generated from PVGIS is used as source domain to train deep learning architectures, which are then apply to actual solar power data from different datasets in Belgium, Germany and the Netherlands. The results show that using transfer learning with synthetic data can improve forecasting performance, having a lower mean squared error, compared to target made (benchmark) models, when limited to no data is available in the target domain. However, this effect decreases over time as more data becomes available, and seasonal patterns have a major impact on the results. When used correctly, these models can contribute to improving solar power forecasting in locations where historical data is limited.

Airu-Wrf: A Probabilistic Spatio-Temporal Wind Forecasting Model And Its Application To The U.s. North Atlantic Offshore Wind Energy Areas

Presenter: Ahmed Aziz
Co-authors: Ahmed Aziz Ezzat;Feng Ye;Travis Miles;Joseph Brodie

The reliable integration of wind energy into modern-day electricity systems largely relies on accurate short-term wind forecasts. We propose a spatio-temporal model called AIRU-WRF (short for the AI-powered Rutgers University Weather Research & Forecasting), which fuses numerical weather predictions (NWPs) with local observations in order to make hub-height wind speed forecasts that are short-term (minutes to hours ahead), and of high resolution, both spatially (turbine-level) and temporally (minute-level). In contrast to purely data-driven methods, we undertake a "physics-guided" probabilistic machine learning approach which captures salient physical features of the local wind field without the need to explicitly solve for those physics. Tested on real-world data from the U.S. North Atlantic where several offshore wind projects are in-development, AIRU-WRF achieves notable improvements, in terms of both wind speed and power, relative to various forecasting benchmarks including physics-based, hybrid, statistical, and deep learning methods.

Evaluating Density Forecasts Using Kernel Scores In A Risk Management Context

Presenter: Jie Cheng
Co-authors: Jie Cheng

Scoring rules are commonly applied to assess the accuracy of density forecasts in both univariate and multivariate settings. In a financial risk management context, we are mostly interested in a particular region of the density: the (left) tail of a portfolio's return distribution. The dependence structure between returns on different assets (associated with a given portfolio) is usually time-varying and asymmetric. In this paper, we conduct a simulation study to compare the discrimination ability between the well-established scores and their threshold-weighted versions with selected regions. This facilitates a comprehensive comparison of the performance of scoring rules in different settings. Our empirical applications also confirm the importance of weighted-threshold scores for accurate estimates of Value-at-risk and related measures of downside risk.

Forecasting Daily Commodity Price Volatility Based On U.s. Macroeconomic Announcement Surprises Using A Midas-Garch Model

Presenter: Matthew Higgins
Co-authors: Matthew Higgins;Yasemin Ulu

Asset market volatility is driven by information flows. Return volatility is often modeled using GARCH and stochastic volatility models in which the high frequency information flow is represented by random innovations which are not structurally identified. In this paper, we estimate models for commodity price volatility that incorporate observable exogenous information flows using a MIDAS-GARCH model. The paper focuses on energy commodities, including crude oil, natural gas and heating oil. We also consider other durable commodities such as metals. The observable exogenous information we use is the information provided by surprises in macroeconomic announcements such as Federal Reserve policy announcements and the release of monthly unemployment and inflation data. Surprises in these announcements are measured using changes in the futures price of financial assets between the days surrounding the announcement date. We also consider the effects of the EIA's natural gas and crude oil storage announcements on volatility. We estimate both univariate and multivariate MIDAS-GARCH models. We find the low frequency macro surprises are important for conditioning daily volatility. We conduct an out of sample volatility forecasting experiment to assess the usefulness of the low frequency announcement information. We compare our surprise based MIDAS-GARCH model forecasts to volatility forecasts from a conventional GARCH(1,1) model and a Markov switching GARCH(1,1) model. To rank the volatility forecasting models, we use loss functions based on appropriate information measures, MZ regressions and the accuracy of prediction intervals. We find the MIDAS-GARCH model to have better predictive power over the other models considered.

Roc And Prc Approaches To Evaluate Recession Forecasts

Presenter: Kajal Lahiri
Co-authors: Kajal Lahiri;Cheng Yang

We use the Receiving Operating Characteristics (ROC) and Precision-Recall (PRC) approaches to evaluate the enduring power of interest rate spread to forecast peak turning points, with more focus on PRC. We evaluate the predictive power as measured by different functionals of the PRC curve e.g., precision, recall, the Area Under Precision-Recall curve (AUPRC), F0.5, F1 and F2. Based on daily U. S. data from January 2, 1962, we find that the optimal threshold stays around spread inversion if F0.5 or F1 metrices are used but has increases to about 0.9% if the more recall/hit rate weighted F2 is used. Further, by

reformulating the PRC curve, we can compare it directly with ROC. The results analytically show why ROC approach tends to give a more favorable predictive evaluation particularly when the event is rare or uncommon.

Optimizing Food Pantry Stocking Strategies: A Client-Preference Based Approach Using Multidimensional Knapsack Problem

Presenter: Benjamin F. Morrow Jr., Ph.D.
Co-authors: Benjamin Morrow;Lauren Davis

This study proposes a novel framework for demand forecasting in food pantries, using a multidimensional bounded multiple knapsack problem (MBMKP) formulation to inform an optimal stocking strategy based on client preferences, item limits, and limits on poundage. The MBMKP is a powerful tool that has yet to be explored in the literature for food pantry demand forecasting, making this study unique and innovative. Survey-based data is collected to explore client food preferences and determine demographic factors associated with those choices. Non-Parametric Analysis of Variance (ANOVA) and binary logistic regression (BLR) are applied to classify responses by client preferences and identify significant predictors. These results serve as inputs to the MBMKP formulation, which yields an optimal stocking strategy for the pantry. The study concludes that the MBMKP formulation, based on client preferences, is an effective tool for demand forecasting and stock optimization in food pantries. It allows pantry managers to consider the preferences of their clients, as well as storage constraints and item limits, to create an optimal stocking strategy that reduces food waste and improves client satisfaction. Overall, this study provides a comprehensive decision support system for food pantry managers, addressing the deficiency in literature examining food bank operations. The proposed framework can aid traditional pantries in transitioning to client-choice distribution and can be applied to different types of food pantries, ensuring that each pantry is treated uniquely for planning purposes.

Novel Forecasting Approaches For Food Bank Operations

Presenter: Ekram Adan
Co-authors: Ekram Adan;Seongtae Kim

There is growing interest in hierarchical time series forecasting, which allows forecasting individual time series within a hierarchical structure. We investigate the effectiveness of various hierarchical forecasting approaches in analyzing food donations at the Second Harvest Food Bank of Metrolina (SHFBM). Specifically, we model weekly time series food donation patterns using bottom-up and optimal combination approaches across different geographical regions and product categories. Our findings indicate that the optimal combination approach, which involves individually forecasting each series at each hierarchy level and then combining and reconciling these forecasts using a regression model, outperforms the independent and bottom-up approaches in terms of mean squared error and mean absolute percentage error achieving 10-20% higher accuracy, both in simulation and in real-world data. Our study shows the advantages of applying hierarchical forecasting approaches to time series data with hierarchical structures. Hierarchical forecasting can be useful in many application areas including forecasting sales or demand for multiple products, locations, or divisions.

Predicting The Food Security Status Of College Students Using Machine Learning

Presenter: Lauren Davis
Co-authors: Lauren Davis;Clinton Griffin;Shona Morgan

Food insecurity occurs when food intake is disrupted because of a lack of money or other resources. In the United States, approximately 13.5 million households were food insecure in 2021. More recently, in the open literature, food security status of college students has received more attention. In this work, we investigate the food secure status of college students at an historically black college and university. We present several machine learning models to predict if a student is food insecure, based on demographic and socio-economic factors.

Forecasting Turnaround Times For Air Cargo Operations

Presenter: Sandria Weisshuhn
Co-authors: Sandria Weisshuhn;Anne Lange;Sarah Van Der Auweraer

Forecasting process times of air cargo operations is a challenging endeavor. All-cargo airlines typically operate on tight schedules to maximize flight hours and minimize ground hours, also known as turnaround times. They also operate on circular routes, only returning to the home base after multiple stops scheduled in close succession. Tightly scheduled round trips increase the risk of introducing and propagating delays through the carrier's network. As a result, originally scheduled turnaround times become an unreliable predictor of actual turnaround times. Other potential drivers of actual turnaround times, such as routes, airports, aircraft, cargo, and weather can be highly inter-dependent and difficult to control or adjust. Despite these difficulties, it is valuable to invest in actionable and accurate forecasts of turnaround times. Even if they lead to only small improvements in operational efficiency, these can still have a huge impact on the bottom line in a marginal business, such as air cargo. Motivated by the goal to develop a forecasting tool for turnaround times, we collaborate with a leading all-cargo airline. We base our analysis on a large data set including 72.000 turnarounds between 2019–2021. We first develop a better understanding of how turnaround times depend on various operational, cargo-related, and external variables. We then use the identified drivers to predict upcoming turnaround times with information available sufficiently early before the aircraft arrives at the turnaround station, such that the company can still make operational adjustments. Methodologically, we draw on econometric modeling and machine learning, combining the strengths of both to produce explainable and accurate forecasts. We aim to contribute to the literature on air cargo operations and on-time performance in aviation. We further seek to inform the forecasting literature regarding how econometric and machine learning models can complement each other and how to balance the trade-off between more prediction accuracy closer to the turnaround but less time to act on the prediction. Our work also suggests important implications for improvements in ground operations and caters to both cost reduction and sustainability objectives of all-cargo airlines.

Empowering Demand Planners With Explainable Ai For Supply Chain Demand Forecasting

Presenter: Rijk Van Der Meulen
Co-authors: Rijk Van Der Meulen

Demand forecasting plays a critical role in supply chain planning and management, and while machine learning methods have become the new standard, they can be difficult to understand and interpret. As a result, compared to classical statistical forecasting, users are less likely to accept outcomes of advanced models - thus companies don't fully benefit from these models. Explainable AI techniques aim to provide transparency and help decision makers understand how the model arrived at its predictions. In this talk, we will present how we as EyeOn, a supply chain consulting firm, are utilizing explainable AI in practice to help customers understand the machine learning forecast. The talk will focus on the application and benefits of explainable AI for demand planners in making better-informed decisions. Translating explainability results like Shapley values to something that planners can use can be challenging. This will be the focus of the talk, where we will present best practices for making explainability results actionable for demand planners and other supply chain practitioners.

Effective Supply Chain Demand Forecasting In Practice

Presenter: Naveen Raja
Co-authors: Naveen Raja

Within the supply chain planning domain, high quality demand forecasting is the backbone of effective decision-making - enabling businesses to manage inventory, improve customer service, and reduce costs. Classical statistical models and more recent machine learning techniques fulfil an essential role in this process. However, for these models to reach their full potential and actually add business value they need to be embedded in a robust and integral process. In this talk, we will explore the key challenges associated with setting up and maintaining an effective supply chain demand forecasting process, including data management, technology infrastructure, data visualization, and stakeholder engagement. Drawing on real-world examples of EyeOn's forecasting as-a-service offering, we will provide insights and best practices for addressing these challenges and building a sustainable forecasting process that delivers ongoing value to the business.

Timely Humor: Data And Forecasting Comedy

Presenter: Evan Wimpey
Co-authors: Evan Wimpey

Evan is a predictive analytics comedian (and we're not just talking about his coding skills). From his earliest days he's always been eager to make people laugh, and maybe one day he finally will. The good news is, there's lots to laugh about in predictive analytics and forecasting! Evan has been using generative language models to come up with the funniest jokes ever told at a forecasting symposium (note the baseline, here).Evan has won awards for his statistics jokes, and performs data comedy regularly for his clients (if his models don't perform, it helps relieve the tension)* This is not meant to be learning content, but would be appropriate for an intermission, opening, reception, etc. Evan has performed privately for clients and is scheduled to perform at Machine Learning Week in June 2023.

Building Combined Forecasts With Machine Learning Using Sas Viya And Python

Presenter: Ari Zitin
Co-authors: Ari Zitin

Machine learning approaches to improving forecast accuracy are especially effective when working with data comprised of large collections of similar time series, many exogenous variables and external information. However, traditional time series approaches can still regularly outperform machine learning methods. One way to take advantage of both methods is to build a combined forecast model that includes both traditional and machine learning methods. Simple averaging of different models can improve forecast performance on a single time series, but with a collection of similar time series we can take advantage of cross-learning opportunities by using machine learning a second time to calculate a weighted average of individual forecast methods. Following the Feature-based Forecast Model Averaging approach from the M4 competition (Montero-Manso, Pablo et. al, 2020) we use a custom gradient boosting model to find the best weights for a combined forecast made up of both traditional forecasting approaches and machine learning based forecasts.A key feature of this method is the use of SAS Viya to efficiently extract features from a collection of time series and generate a collection of different forecasts for each time series. Extracted features and forecast errors are then moved to Python where a custom gradient boosting model is trained to predict the best combination weights for each time series in the collection. Although each time series is forecast using the same collection of traditional and machine learning methods, the combination weights will be specific to each series. We extract features and generate forecasts using SAS Viya, while we use

Python to build the combined forecast machine learning model. This approach is intentionally flexible, so forecasts generated using different tools, languages, or interfaces can be combined in Python.References: Montero-Manso, Pablo, et al. "FFORMA: Feature-based forecast model averaging." International Journal of Forecasting 36.1 (2020): 86-92.

Energy Forecasting For The Future

Presenter: Jennifer Whaley
Co-authors: Jennifer Whaley

The integration of distributed energy resources, compliance to climate goals, and optimal upgrades to infrastructure are adding complexities to the traditional load forecasting process. Integrating the increasing number of distributed energy resources located within the distribution system—rooftop solar photovoltaic, electric vehicles, demand response—requires the analytics of load forecasting to know when, how much, and where these resources are contributing on the power grid, while also balancing the amount of traditional and renewable generating resources to meet demand. Forecasting the load on assets along the distribution system provides insights on which components are overloaded beyond their manufactured design limits and identifies where the grid needs to be hardened as electric power providers are committed to delivering safe, reliable power. In this session we will illustrate the value of a flexible, scalable, and automated solution to meet the requirements of increasing complexity in the load forecasting discipline.

Developing And Refining Custom Forecast Models

Presenter: Chip Wells
Co-authors: Chip Wells

In a large-scale forecasting scenario, analysts don't have time to manually scrutinize each series, specify models, choose a champion for each series, and so on. Automated tools are needed to accomplish the identify, estimate, and forecast steps for each of the many series an analyst is responsible for. In this context the role of the forecasting analyst changes. Analysts mainly spend their time in two ways; monitoring the system and improving results for problematic or high values series. This talk presents a large-scale forecasting toolbox for refining system generated models and for creating new, custom models. The focus is on how analysts interact with the forecasting system to efficiently leverage their expertise in improving the precision of system generated forecasts.

Evolution Of Topics In Time Series Forecasting: Insights From Social Media And Research Articles

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

Topic modeling has been a critical research topic over the past two decades, and understanding the evolution of time series forecasting topics is essential to forecasters. For example, de Gooijer and Hyndman's article in a 2006 issue of the International Journal of Forecasting (IJF) highlighted the key papers about time series forecasting methods published in journals that were managed by the International Institute of Forecasters from 1982 to 2005. In this paper, we apply dynamic topic modeling to social media data from Twitter and IJF research paper data in order to gain insights into the evolution of time series forecasting topics. Dynamic topic modeling enables us to analyze how topics change over time and how different topics are related to each other at different time points. Through our analysis, we identify some key events that have influenced the time series forecasting community and have shaped trends in new time series forecasting methods. Overall, our study provides valuable insights into the evolution of time series forecasting topics, which can inform forecasting researchers and practitioners in their work and determine research directions.

Improving The Forecast Accuracy Of Protected Data Using Time Series Features

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

Existing data privacy methods degrade forecast accuracy to unusable levels. To overcome this problem, we investigate the similarity between time series features that are predictive of forecast accuracy. We develop a matrix-based privacy method called k-nearest time series + (k-nTS+) swapping tailored to maintain forecast accuracy. We apply our privacy method to a forecasting competition data set where the identities of the time series are hidden but an adversary seeks to identify them. Using only six time series features, we find that k-nTS+ swapping maintains forecast accuracy and preserves the distribution of time series features much better than competitor methods at similar privacy levels. The k-nTS+ protected time series are also more representative of the original data, potentially leading to increased trust between data owners and forecasters. By preserving the original time series features and improving the privacy of sensitive data, our k-nTS+ method could produce data for multiple applications and increase data availability for theforecasting community.

Beyond Causality: Using Forecasting Analytics For Business, Industrial And Policy Problems

Presenter: Vasileios Bougioukos
Co-authors: Vasileios Bougioukos;Konstantinos Nikolopoulos;Julian Williams;Ujjal Mukherjee;Alexandros Kalantzopoulos

The current doctrine in the operations management literature emphasizes the sensitive nature of the relationship between forecasting and causal inference. While forecasting is considered to be a distinct problem from establishing causality, the overarching argument is that we need both in order to understand the underlying assumptions of our problems, and subsequently optimize data-driven decision-making. Be that as it may, we cannot ignore the benefits of being able to forecast the next state of a system, even if we are not sure why. We argue that big data, computational power, and (human and artificial) intelligence are instrumental in improving forecasting performance, and we test the power of the driving and constraining forces in a novel conceptual framework. To that end, we employ a combination of a deep survey, our hypotheses, data for testing (FRED, WHO, Kaggle, Twitter, Google Trends, Amazon reviews) and methods which will be made available publicly through a dedicated website, and pseudo-code for in freeware platforms (R/Python). Emphasis is given in replicability, reproducibility, transparency, and data availability issues. We finally document the current state-of-the-art analytical techniques in forecasting research and practice fathoming on the interface between data, techniques, andcomputing, and we provide a series of illustrative successful examples from business, industrial, and policy problems.

A Time Series Of Networks. Is Everything Ok? Are There Anomalies?

Presenter: Sevvandi Kandanaarachchi
Co-authors: Sevvandi Kandanaarachchi;Rob Hyndman

Consider how bills get voted in the Parliament/Congress. Members belonging to different parties may vote differently. As time passes the voting patterns can change. These bill voting patterns can be denoted as a network. At each time stamp, a different network emerges. The collection of networks indexed by time is a time series – of networks. We study these network time series. What are the features of the networks? How do the features change over time? Are there any anomalous networks? We investigate these questions using real world networks. We use graph theoretic features to transform the network to a feature space and model their evolution using time series methods. Then we find anomalous networks using time series residuals. Our results coincide with noteworthy, historical events.

Forecasting Spoofing Manipulation

Presenter: Tatiana Franus
Co-authors: Tatiana Franus;Richard Payne;Malvina Marchese

Spoofing manipulation activity on financial markets could lead to losses for retail traders. Significant resources have been invested in automated surveillance systems to detect manipulative behaviours. This paper introduces a data-driven approach to forecast the market state when spoofing manipulation is likely to appear. We present a novel Real-Time Spoofing Probability (RTSP) measure based on several machine learning (ML) algorithms. We train a ML method on spoofing cases detected on Moscow Exchange and develop a new forecasting framework. Our contribution is three-fold. First, we identify a hundred predictors of the market state and use the lasso and the elastic net as variable selection methods. Secondly, we match spoofing orders with all orders and trades, allowing us to track the manipulative order's lifetime and order book price level. We choose ML methodology as spoofing events are not equally spaced within an extensive dataset. Chosen models (Random Forest, XGBoost, and Decision Trees) predict correctly above 70% of spoofing events for balanced data. We use cross-validation to avoid overfitting and keeping the analysis out-of-sample. We also use different cross-validation checks for imbalanced data. Finally, we introduce the RTSP measure that forecasts intraday manipulative activity. To test our measure on out-of-sample data, we train ML algorithms on five previous trading days and run a rolling cross-validation forecast for the next 10-, 30-, and 60-minutes. We endogenize the RTSP measures as a simple average of ML outcomes. Consequently, RTSP shows a probability in real time that the market states is preferable for the spoofer to place his order. Moreover, we compare the forecasting performance of RTSP to the performance of other ML algorithms and show that the designed methodology performs better in the given environment. The model is self-training, so no adjustments are needed to detect other manipulative practices. This approach reduces the importance of model selection through forecast combination. An empirical evaluation of the proposed framework demonstrates significant forecasting accuracy for short-term forecasts in a high-frequency environment. We employ the model to discuss market regulation, particularly financial market surveillance.

Anomaly Detection In Image Time Series (Its) Using Explainable Ai (Xai)

Presenter: Priyanga Dilini Talagala
Co-authors: Priyanga Dilini Talagala

Analysis of Image time series (ITS) has become increasingly important as a tool for monitoring and understanding complex systems and phenomena, such as climate change, urbanization, and land-use changes. By capturing images at regular intervals, it is possible to track and analyze changes and trends over time, which can help inform decision-making and policy development. The existing deep learning-based anomaly detection methods solely focus on the classification task, leading to a lack of explainability in the model and reducing human trust in the system. We proposed a framework for early detection of anomalies in image time series using explainable AI, which defines an anomaly as an observation that is very unlikely given the forecast distribution for the corresponding time period. Our previous study focused on developing the entire framework for detecting anomalies in image time series. In this work, our attention is on further improving the explainable AI component of the proposed framework. Our proposed framework consists of a deep learning-based computer vision module that derives a feature space from the ITS data, a deep learning-based time series forecasting module that generates a probability distribution for each time period, and an extreme value theory-based boundary prediction process that helps identify anomalies. To address the lack of transparency and interpretability in the deep learning-based anomaly detection component, we integrate an explainable AI (XAI) module that provides post-hoc, model-agnostic, and local explanations, increasing the trustworthiness of the prediction of anomalous behaviours. Our experiments with satellite image time series (SITS) show that the proposed algorithm can work well even in the presence of noisy image time series data and that the XAI module improves the transparency and interpretability of the deep learning-based anomaly detection framework. Overall, our proposed framework provides a powerful tool for

detecting the progression of unusual behaviours in ITS with increased trustworthiness and interpretability and has a wide range of applications in environmental monitoring, disaster management, and climate change analysis.

Forecasting Via Reduced Big Data Matrices Under Lag-Sparsity Of Relevant Information

Presenter: Livio Fenga
Co-authors: Livio Fenga

Big Data matrices are becoming more and more a common place when secondary data are employed. However, in many instances, the exploitation of a dataset containing a great number of variables has many drawbacks, the most important one being associated with the extraction of only the relevant portion of the whole information set (usually a small subsample of the original dataset). Such a task becomes even more complicated in time domain, given the "physiological" inclusion in the design matrix of the lagged versions of all the available variables. But this is not the whole story. In fact, lag sparsity of the design matrix is very likely to happen in such a setup, which should always be considered. The present paper aims at developing a method designed to enhance the accuracy of the predictions of time series affected by high degrees of uncertainty generated by lack of parsimony in the original design matrix. Usually, Auto Regressive Distributed Lag (ARDL) models can efficiently accommodate a "small" number of independent lagged variable, how small depending on both the number of independent variables being investigated and the number of significant lags needed to be accounted for in the final model. The proposed dimension reduction procedure focuses on ARDL models, which are envisioned to work in conjunction with a LASSO (Least Absolute Shrinkage and Selection Operator) method for the efficient suppression of the irrelevant information (that is the greatest part of the data matrix) not only by constraining the variables space but also the lag-space. While the former is based on a pure LASSO method, the latter relies on a novel criterion, which is aimed at discovering lag-consistency in the variables extracted by the LASSO procedure in the first place. The rationale is as follows: noisy and otherwise corrupted signals might generate many false relevant lags. Essentially, false positives are those lags which play no role in explaining the variability of the time series of interest and, if included in the final model, can generate many types of unwanted phenomena, e.g., degradation of the forecasting performances or noise.

The Reconciled Output Gap: A State-Space Framework To Model Revisions

Presenter: Nina Mühlebach
Co-authors: Nina Mühlebach;Marc Anderes;Florian Eckert

Output gap estimates are often subject to strong revisions after the initial release. We use a state-space framework to estimate a 'true' output gap that is reconciled from ten well-established output gap methods. The use of multiple vintages allows us to model revisions and decompose them into news and noise components. We show that this leads to an intuitive and economically meaningful output gap estimate. In a comprehensive real-time study for the United States, we review the performance of all output gap methods and compare them to a median model and our reconciled gap. We provide evidence that our output gap features the lowest root mean squared revisions in real-time over the entire sample-span, while also being reliable in times of crisis. In contrast to benchmark methods like the Congressional Budget Office estimate, the reconciled gap is characterized by symmetry both in real-time and ex-post, thus advocating its use for fiscal planning. Finally, our results point to the importance of incorporating multiple releases and methods to build a reliable output gap.

Integrating Forecasting And Optimization

Presenter: Benedikt Sonnleitner
Co-authors: Benedikt Sonnleitner;Nikolaos Kourentzes

Traditionally forecasting and optimization are separate disciplines even though both are required to support decision making under uncertainty, where forecasts serve as input into optimization. Forecasting literature mostly aims to provide an accurate forecast, including accurate uncertainty estimates. Optimization provides for many instances provable near-optimal decisions, assuming the forecast and its uncertainty estimate is correct. To integrate the two disciplines we evaluate whether we can solve a single optimization problem in which the parameters of a linear regression are estimated along with the actual decisions jointly, instead of treating the problems sequentially as usual. This leads to a dependence between forecast parameter estimation and decisions made, introducing a trade off in the solution space which we analyse. For our investigation we use a simple network flow problem with uncertain demand on the edges. We consider both the univariate case and the inclusion of covariates.

Modelling The Reserve Demand To Facilitate Central Bank Operations

Presenter: Romain Michel Veyrune
Co-authors: Romain Michel Veyrune;Zhuohui Chen;Nikolaos Kourentzes

Central banks exercise monetary policy with the aim of keeping short-term rates close to their policy rates. Central banks can control short-term rates because they issue the most liquidity asset in a financial system, the banks' accounts in their book, which is traded at those interest rates. Short-term interest rates represent the marginal cost of funding of financial intermediaries and controlling them allow the central bank to influence the interest rate and the yields of the various financial products. In addition, short-term rate volatility is undesirable because it adds an unnecessary cost to the economy, that is the liquidity premium. While the central bank can decide the reserves that are available for the banks, it also needs to know how much reserves banks want. The aim of this work is to offer a modelling methodology for estimating the demand for reserve that itself is influenced by various macro and market structure variables. The estimate would help central bank to identify "stable points" on the demand for reserves that is the levels of reserves for which short-term rate volatility is minimal compared to other points on the demand curve. In pursuit of its interest rate objective, the central bank should endeavor to keep reserves available on the stable points of the demand curve. We provide both parametric and non-parametric approaches, with particular focus on capturing the modelling uncertainty and, therefore, facilitating scenario analysis. The proposed methodology is evaluated using data from the European Central Bank and the US Federal Reserve System.

Structural Changes In Asset Correlations And Macroeconomic Fundamentals

Presenter: Malvina Marchese
Co-authors: Malvina Marchese;Ioannis Kyriakou

This article proposes a novel approach to modelling and forecasting structural changes in asset returns correlations and their relationship to macroeconomic fundamentals. We introduce a new correlation component model, the Regime-switching DCC-MIDAS, that incorporates breaks of different type in the conditional and unconditional correlations. Breaks in the secular component are associated with low-frequency macroeconomic fundamentals via a Smooth Transition mixed-data sampling regression, while short-run correlations are characterized by abrupt regime switches linked to market constraints. Following a discussion of estimation, inference and simulation-based evaluations, the model is applied to the prediction of future energy returns. The results indicate that the Regime-switching DCC-MIDAS is a very useful specification especially in periods of intense market instability, such as the recent pandemic crisis.

Credit Market Sentiment And Stock Returns

Presenter: Gergely Ganics
Co-authors: Gergely Ganics;Chaoyi Chen;Zhou Ren

Credit market sentiment is characterized by mean reversion and has been shown to possess significant predictive power for future economic activity through a dynamic mechanism known as "diagnostic expectations." Typically, this relationship has been analyzed using linear regressions with two predictors of future changes in credit spreads: the level of credit spreads and the share of high-yield bonds issued. However, it remains unclear whether the reversal of prior sentiment holds significant predictive power for changes in credit spreads under both high and low sentiment conditions. In this study, we find that diagnostic expectations exhibit asymmetry by examining quantile regressions. We then propose an entropy-based measure of positive sentiment to capture the asymmetric reversion. This measure not only performs well in cross-sectional asset pricing regressions but also delivers meaningful out-of-sample gains when predicting stock returns. Our findings highlight the importance of considering asymmetric reversion in credit market sentiment when forecasting economic activity and asset prices.

Forecasting Natural Gas Consumption Using Mixed Frequency Dynamic Factor Models And Machine Learning Methods

Presenter: María-Pilar Zazpe-Qintana
Co-authors: María-Pilar Zazpe-Qintana;Roberto Morales-Arsenal

Natural gas is the main heating fuel in the European Union (EU). Regulation (EU) 2022/1369 establishes that EU states must reduce their national gas demand by 15% between August 1 and March 31, 2023, as part of the measures adopted to address the energy crisis stemming from the war in Ukraine. In this context, it is crucial to have an effective forecasting system. This paper analyzes the results of three methods to obtain point, interval and density forecasts: 1) A mixed frequency dynamic factor models (MF-DFM) to forecast natural gas consumption in the European Union using a large data set of economic and temperature variables and weather conditions (Bańbura and Modugno, 2014). A mixed frequency observations was used (weekly, monthly, quarterly and annual data). 2) Artificial neural networks (ANN), and deep learning (DL) was used to make comparisons and 3) Point forecast combination approach (Bates and Granger, 1969) between both types of models. The results indicate that, in terms of the MAPE, the MF-DFM presented more accuracy focused on long and medium term. Machine and deep learning methods were more suitable for short term forecasting but a simple (arithmetic mean) point forecast combination between both approaches outperformed the individual methods in different forecasting horizons.

Forecasting Natural Gas Demand Using Hierarchical Frameworks

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

Natural gas consumption data can be hierarchically organized to improve demand forecasting. Gas distribution companies must provide enough gas to meet the demand expected at the most aggregate level of the hierarchy (total consumption of the system). Individual consumers burn gas as needed at the least level of aggregation (personal-use gas). Forecasts at each of these levels are relied on for economical and sustainable operation of the pipeline network, however, it is often the case that consumption data at one or both aggregate levels of interest are not available. This work evaluates the effectiveness of applying grouped time series structures of varying complexity to the problem of gas demand forecasting. We present our framework by selecting hierarchical structures representing gas consumption of 100,000 customers located in the western United States. Individuals are grouped into hierarchies using temporal and spatial information, as well as historical consumption trends. All series within the hierarchy are modeled

simultaneously using multi-parameter linear regression. Novel coherence constraints are determined to manage overlapping and inconsistently spaced cycle-billing observations and enforced throughout hierarchies using an optimal-combination reconciliation technique. Effectiveness of this forecasting framework is shown in a temporal disaggregation application, with daily consumption at the lowest level of aggregation being the target hierarchy. This framework for forecasting natural gas demand produces a 12.4% MAPE evaluated over a three-year period and is the most accurate disaggregation result as shown in the case studies included in this work.

Hourly Natural Gas Forecasting In The U.s.

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

There is a world-wide movement to green energy. However, while this transition is being made, there is still a need for fossil fuels. Natural gas (primarily methane) is one of the cleaner fossil fuels as it does not produce particulate matter and produces less carbon dioxide than coal. Natural gas is primarily used for space heating, cooking, and electricity generation. It is in space heating that accurate forecasts of natural gas are most needed. On extreme cold days without an accurate forecast, there may not be enough natural gas to heat homes. Unlike electricity, when natural gas is shut off to a home, it cannot be simply turned on. Rather, a technician must come and restart the gas supply. Thus, especially on those coldest of days, accurate natural gas forecasts are needed.Our method for forecasting natural gas starts with a data cleaning process. After anomaly detection and imputation, the data is detrended such that the whole data set has similar statistics to the final year used in the training process. Next, a feature selection process reduces a set of approximately 1200 features down to 120 – 170 features. This trimming process builds a linear model and keeps only those features that have statistical significance. The final preprocessing step is to weight the data by giving more importance to the most recent year and the coldest days. Our method is a direct forecasting method, that is models are built to forecast each 168 hourly horizons. The model is a stacked ensembler of linear, tree, and neural network regression.We evaluate our models using weighted mean absolute percent error (wMAPE), which is unitless and allows comparison of forecasting accuracy across all operating areas. Weighted MAPE weights the error towards the largest consumption period, which is the winter. The models are trained on approximately 15 years of data and tested on the next year.

Forecast Combination: Between Equal And Individual Weights

Presenter: Lars Averkamp
Co-authors: Lars Averkamp;Uta Pigorsch

For over 50 years, forecast combination provides a great value for forecasting problems. Since the initial work by Bates and Granger (1969), in which they propose to combine forecasts based on an error-variance-minimization approach, many other methods have been developed ranging from simple combination schemes to more complicated approaches. Nevertheless, oftentimes, equal-weighted forecasts still outperform more sophisticated and theoretically superior methods – a phenomena that is commonly referred to as the "forecast combination puzzle'' and that may be attributed to the error occurring in the estimation of the forecast combination weights, see Smith and Wallis (2009). Different methods have been developed to somehow constrain or shrink weights in order to reduce the error-variance of the combined forecast, such as the well-known LASSO. We contribute to this literature by proposing a novel method that nests the equal-weights forecast, the error-variance-minimization approach of Bates and Granger (1969) and its variant with non-negativity constraints, but importantly, also allows for solutions transitioning between those three methods. This is achieved by imposing additional constraints involving two hyperparameters, that shrink a subset of weights to a positive equal-weights solution and / or a subset of weights to a negative equal-weights solution while others are allowed to deviate from those equal-weights solutions as the hyperparameters change. In doing so, we harness the benefits from both individual and equal-weights

solutions. To assess the potential of our method we compare its out-of-sample forecasting performance to other commonly used forecast combination methods in terms of a simulation study based upon different error-covariance matrices, and an empirical application. For both we find that our method oftentimes outperforms the benchmarks.

Another Look At Forecast Trimming For Combinations: Robustness, Accuracy And Diversity

Presenter: Xiaoqian Wang
Co-authors: Xiaoqian Wang;Yanfei Kang;Feng Li

Forecast combination is widely recognized as a preferred strategy over forecast selection due to its ability to mitigate the uncertainty associated with identifying a single "best" forecast. Nonetheless, sophisticated combinations are often empirically dominated by simple averaging, which is commonly attributed to the weight estimation error. The issue becomes more problematic when dealing with a forecast pool containing a large number of individual forecasts. In this paper, we propose a new forecast trimming algorithm to identify an optimal subset from the original forecast pool for forecast combination tasks. In contrast to existing approaches, our proposed algorithm simultaneously takes into account the robustness, accuracy and diversity issues of the forecast pool, rather than isolating each one of these issues. We also develop five forecast trimming algorithms as benchmarks, including one trimming-free algorithm and several trimming algorithms that isolate each one of the three key issues. Experimental results show that our algorithm achieves superior forecasting performance in general in terms of both point forecasts and prediction intervals. Nevertheless, we argue that diversity does not always have to be addressed in forecast trimming. Based on the results, we offer some practical guidelines on the selection of forecast trimming algorithms for a target series.

Combining Forecasts Based On Prediction Intervals

Presenter: Devon Barrow
Co-authors: Devon Barrow

Prediction intervals provide a range within which we expect our forecasts to lie given a pre-specified probability thus providing a measure of the uncertainty in the forecasts. Prediction intervals derived based on bootstrapping overcome the need to assume that model residuals (errors) are normally distributed and only require that the residuals are uncorrelated with constant variance. Simultaneously bootstrapping based on residuals has been shown to improve forecasting accuracy by combining the resulting forecasts generated based on the bootstrapped data. This approach referred to as bagging for Bootstrap and Aggregating is generally applied to a single model and equal weights are used to combine the forecasts. In this research we explore the use of bootstrapping as applied to prediction intervals for combining forecasts. We propose a procedure for generating forecast combination weights based on the coverage achieved by a given bootstrapped prediction interval across a set of candidate forecasts. This coverage is based on the width of the prediction interval of each model, and the probability of candidate forecasts falling within that interval. We evaluate our procedure on simulated data and several well-known datasets including the M3-Competition and a dataset of 229 weekly demand series from a leading household and personal care UK manufacturer. Findings suggest that this new forecasting combination approach improves on predictive accuracy while naturally leading to predictive intervals for the combined forecasts which are narrower.

Composite Ai Using Forecasting And Optimization

Presenter: Jay Laramore
Co-authors: Jay Laramore

Optimization is a natural extension to the forecasting process. Once we answer, "What will happen?", those forecasts can then be embedded into an optimization model to determine, "What should we do?". By integrating forecasts into an optimization model, analysts are able to provide practical guidance and decision making under uncertainty. The model results can be used to streamline business operations by moving from reactive to proactive decision making. This session focuses on industry examples and applied strategies for building composite AI solutions using forecasting and optimization.

Forecast Explainability Part 1: Improve Decision-Making And Forecast Adoption

Presenter: Jessica Curtis
Co-authors: Jessica Curtis

Often the biggest challenge with forecasting is adoption – incorporating forecasts into decision-making. Some companies report that planners still adjust 80-90% of forecasts, despite consistently strong accuracy results. This is frequently due to a lack of trust in statistical forecasts.Further, organizations are incorporating new data sources into forecasts and using both time series and machine learning techniques to capture rapidly changing demand patterns. To be valuable, statistical forecasts must be explainable and transparent, enabling an understanding of the key drivers of demand to make more informed, data-driven decisions.Establishing trust and gaining forecast adoption involves explaining the purpose of statistical forecasts and the analytical models themselves. This leads to a clear understanding of forecast models while bringing visibility to how forecasts change when causal inputs or historical trends shift. Join this talk to learn: • Common challenges with forecast explainability, including how to develop mutually exclusive buckets of causal impacts and how to combat the accuracy vs. explainability dilemma. • Practical methodologies for explainability across time series and machine learning models. • Recommendations on meaningfully illustrating and communicating forecast explainability results to planners, executives and other consumers of statistical forecasts.

Forecast Explainability Part 2: Insights From An Implementation For A Cpg Company

Presenter: Sofie Michiels
Co-authors: Sofie Michiels;Nitzi Roehl

As forecast practitioners, we strive to generate the most accurate forecast possible, thereby often sacrificing explainability of the forecast model. A lack of understanding into what contributes to the forecast results (or a so-called "black-box" approach) leads to lack of trust and reduced forecast adoption. In this presentation, we will share a case study of an implementation of an explainable forecast at a Consumer Package Goods (CPG) company. We will discuss how we generated a shipment forecast through a "glass-box" approach that provided the business with visibility into the key business factors driving the forecast results: points of distribution, price, promotions, media expenditure, and category competition. Since the main variable influencing the shipments at a CPG company is the end-user consumption, we focused our efforts on providing an explainable consumption forecast. To that end, we leveraged proprietary and syndicated data from multiple sources. We divided the forecast into a base and a promotional component and modeled these using time series and machine learning methods. You will learn how this approach allowed us to further decompose the historical and forecasted consumption values into components that are meaningful to the business. This transparency into the underlying assumptions that make up the forecast boosts trust with the business and reduces the need to override the forecast. It also provides the business with the ability to test alternative assumptions on key business drivers and evaluate the impact on the overall forecast and its components.

Forecasting At Scale In The Cloud

Presenter: Rajesh Selukar
Co-authors: Rajesh Selukar

In recent years, cloud computing has emerged as a go-to solution for organizations seeking cost-effective and flexible computing options. However, the current economic landscape has raised questions about the cost-efficiency of this approach, urging companies to seek new ways to reduce costs while maintaining high levels of operational performance. Large enterprises that engage in time-series forecasting face a unique challenge to achieve better forecasting accuracy at a reasonable cost due to the massive number of items that need to be forecasted. In order to optimize the forecasting process, these organizations must navigate a complex web of decisions regarding their choice of forecast models, programming languages, and compute infrastructure to ensure efficient computations. These decisions are also dependent on the availability of tools, skills, and hardware in the organization. In this presentation, we highlight the significance of addressing these challenges and propose a benchmark of state-of-the-art forecasting implementations and approaches. We demonstrate how popular open-source and proprietary algorithms can be integrated and scaled in a distributed setting using SAS® Visual Forecasting software. This research aims to provide organizations with valuable insights to optimize their forecasting processes and maintain their competitive edge in today's fast-paced business landscape.

Discrete Forecast Reconciliation

Presenter: Bohan Zhang
Co-authors: Bohan Zhang;Anastasios Panagiotelis;Yanfei Kang;Feng Li

While forecast reconciliation has seen great success for real valued data, the method has not yet been comprehensively extended to the discrete case. This paper defines and develops a formal discrete forecast reconciliation framework based on optimising scoring rules that produces coherent joint probabilistic forecasts for count hierarchical time series. Two discrete reconciliation algorithms are proposed and compared to generalisations of the top-down and bottom-up approaches to count data. Two simulation experiments and two empirical examples are conducted to validate that the proposed reconciliation algorithms improve forecast accuracy. The empirical applications are to forecast criminal offences in Washington D.C. and the exceedance of thresholds in age-specific mortality rates in Australia. Compared to the top-down and bottom-up approaches, the proposed framework shows superior performance in both simulations and empirical studies.

Exploring Robust Covariance Estimators On Forecast Reconciliation

Presenter: Maurício Lila
Co-authors: Maurício Lila;Fernando Oliveira;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 based on regression models, the weighting system can be compromised when the covariance structures are not well estimated, causing distortions to the reconciliation process. This work introduces the concept of hierarchical forecast reconciliation using robust estimation of the covariance structure. 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 and data on Australian tourism. We compare the performance of our strategy with that from traditional regression-based state-of-the-art methods in HTS. Overall, robust covariance estimates 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.

Hierarchical Forecasting At Scale

Presenter: Olivier Sprangers
Co-authors: Olivier Sprangers;Sebastian Schelter;Maarten De Rijke

Generating coherent forecasts across millions of products at an ecommerce company is computationally demanding using existing state-of-the-art hierarchical forecasting techniques, as these techniques are typically applied as a post-processing step that requires a matrix inversion that scales with the number of products or product groups. For one of the largest ecommerce companies in the Netherlands, we investigated the use of learning a coherent forecast for millions of products with a single bottom-level forecast model by using a loss function that directly optimizes the hierarchical product structure. We implemented this using sparse linear algebra, such that the number of operations in our loss function scales linearly rather than quadratically with the number of products and levels in the hierarchal structure. We chose a gradient boosting model for our forecasts, as this is also the baseline forecasting model in the company and these models have shown strong performance in recent forecasting competitions with similar data characteristics (e.g. M5 competition). Unfortunately, results were negative: our hierarchical loss resulted in a worse forecasting performance as measured by RMSE of about 1% at the product level, as compared to the the baseline model that used a tweedie loss function. Subsequent verification on the public M5 dataset yielded neutral results: our hierarchical loss function performed on par with the baseline loss function. Despite the negative results, the benefit of our sparse hierarchical loss function is that it provides practitioners a method of producing bottom-level forecasts that are coherent to any chosen (weighted) hierarchy. In addition, removing the need for a post-processing step as used in traditional hierachical forecasting techniques reduces the computational cost of the prediction phase in the forecasting pipeline.

High-dimensional covariance estimation with a diagonal target

Presenter: Mingmei Xiao
Co-authors: Mingmei Xiao;Sakai Ando

Covariance estimation is important for areas like finance, macroeconomic forecasting, GMM estimation etc. The sample covariance matrix is an unbiased estimator but is ill-conditioned when its dimension is close to sample size and non-invertible when its dimension is larger than sample size. Current literature deals with this problem by either using a few factors to capture the structure of the matrix or shrinking the sample covariance matrix towards a target matrix with fewer parameters. Most of the papers including the Ledoit and Wolf (2003) and Chen et al. (2010) propose a shrinkage target of common variances. This doesn't apply when there are large differences between the true variances of the variables. We propose a new estimator using a shrinkage target of a diagonal matrix of the sample variances and derive the optimal shrinkage weight. Simulation shows good performance when compared with competing methods. The method is then applied to forecast reconciliation and inflation nowcasting and shows competitive performance.

Deepplan – Forecasting As A Means To An End, I.e. Personalizing Demand Planning

Presenter: Devavrat Shah
Co-authors: Devavrat Shah

Demand planning is a complicated, collaborative process wherein the understanding of future customer demand helps in deciding inventory, production, revenue, and services across the organization. While accurate forecasting is an important step towards it, the end decisions are about deciding between various plausible tradeoffs. Typically, such tradeoffs are evaluated in an ad-hoc manner at the best, and there is no one-size-that-fits-all solution in terms of which tradeoff to choose to operationalize. In this talk, we'll be discussing recent technological advances that enable evaluation of feasible tradeoffs between multiple objectives using historical data with the help of scenario analysis so that teams working collaboratively can make such decisions that are right for them at that time. Join us to learn about:• Scenario analysis and how it can help evaluate feasible tradeoffs • Role of forecast accuracy in demand planning through scenario analysis • End-to-end tools that incorporate data, enable accurate forecasts, perform scenario analysis and operationalize demand planning across organization based on the chosen tradeoffReading:https://ikigailabs.medium.com/we-know-how-much-money-you-lost-in-sales-this-year-3d1b1157c94ehttps://www.ikigailabs.io/reinforcement-learninghttps://www.ikigailabs.io/blog/three-technologies-shaping-demand-planning/subsection%7BIntegrating Machine Learning And Optimization With Applications In Public Health And Sustainability}Presenter: Kai Wang
Co-authors: Kai Wang

This talk summarizes the importance of integrating optimization in both offline and online learning with applications in public health and environmental sustainability. Existing machine learning approaches primarily focus on training predictive models separately from optimization, which leads to a mismatch in predictive performance and decision quality in the downstream optimization tasks. This talk covers my work on decision-focused learning to integrate feedback from optimization to train predictive models, to avoid this mismatch. My work provides the first decision-focused learning algorithm for sequential decision problems and it significantly reduces the computation cost to enable applications in large-scale public health problems. My decision-focused learning algorithm is currently deployed in a maternal and child health program used by 100,000 beneficiaries in India to effectively schedule limited health workers to improve mothers' engagement with health information.

Explainable And Robust Electricity Forecasting For Decision-Making

Presenter: Liang Sun
Co-authors: Liang Sun

Electricity forecasting, including electric load and renewable energy forecasting, is critical for scheduling and planning future electric loads and improving the reliability and safety of the power grid. However, decision-making in the electricity system is high-risk, as the prediction results can influence both future assessments and the balance of subsequent power dispatch. Also experts still play an essential role in understanding the forecasting models/results and making further decisions. To narrow the trust gap between algorithms and experts, a more explainable forecast is desired. On the other hand, the diverse data for electricity forecasting from various sources is often affected by outliers, noises, and missing values due to sensing, acquisition, and record errors. This makes accurate forecasting challenging.In this talk, I will introduce a robust and explainable electricity forecasting framework. Specifically, a set of decomposition-based algorithms are introduced to handle different types of noises and outliers. To make accurate and interpretable forecasts in extreme conditions, such as record-breaking high temperatures in the summer, where training samples are limited, we develop an interactive generalized additive model (GAM). The GAM is interpretable and can incorporate specific domain knowledge from the electric power industry for improved performance. Furthermore, to provide explainability for black-box forecasting algorithms, we propose feature attribution explanations and dependency plots as model-agnostic explanations. This framework has already been deployed in both power grids and wind farms.

Decision-Aware Learning For Global Health Supply Chains

Presenter: Vahid Rostami
Co-authors: Vahid Rostami;Hamsa Bastani

The combination of machine learning (for prediction) and optimization (for decision-making) is increasingly used in practice. However, a key challenge is the need to align the loss function used to train the machine learning model with the decision loss associated with the downstream optimization problem. Traditional solutions have limited flexibility in the model architecture and/or scale poorly to large datasets. We propose a principled decision-aware learning algorithm that uses a novel Taylor expansion of the optimal decision loss to derive the machine learning loss. Importantly, our approach only requires a simple re-weighting of the training data, allowing it to flexibly and scalably be incorporated into complex modern data science pipelines, yet producing sizable efficiency gains. We apply our framework to optimize the distribution of essential medicines in collaboration with policymakers at the Sierra Leone National Medical Supplies Agency; highly uncertain demand and limited budgets currently result in excessive unmet demand. We leverage random forests with meta-learning to learn complex cross-correlations across facilities, and apply our decision-aware learning approach to align the prediction loss with the objective of minimizing unmet demand. Out-of-sample results demonstrate that our end-to-end approach significantly reduces unmet demand across 1000+ health facilities throughout Sierra Leone. Joint work with O. Bastani, T.-H. Chung and V. Rostami.

Demand Forecast As An Input To Network Capacity Planning

Presenter: Arpita Mukherjee
Co-authors: Arpita Mukherjee

For a company like Meta it is extremely crucial to build scalable systems that ensure reliable network infrastructure across the vast array of products and services. Building the right network capacity is crucial and it is guided by accurate forecasts of network traffic across different regions. The backbone capacity planning framework of Meta is “hose based”, i.e. the demand forecast is done for total ingress/egress traffic of each data center region. The final capacity plan is done for the actual (physical) network space though. Thus the demand forecast output goes through multiple transformations before it finally gets ingested in the optimization engine to determine the capacities. Since capacity addition is always incremental and there is no notion of capacity deletion, it imposes some consistency requirement on the demand forecast signals. In other words if the demand forecasts for a specific target quarter are too different from one release to another that might lead to a completely different capacity plan, requiring more capacity addition than what can actually be delivered. In this talk I will provide an end to end overview of the capacity planning framework and highlight the key challenges we face in this process.

An Exploration Of Trained Ensemble Models For Epidemic Forecasting Based On Interval Scores

Presenter: Nibir Chandra Mandal
Co-authors: Nibir Chandra Mandal;Aniruddha Adiga;Gursharn Kaur;Srinivasan Venkatramanan;Bryan Lewis;Madhav Marathe

An ensemble of multi-model forecasts has been shown to perform better than its constituent models in forecasting epidemics such as COVID-19 [1], influenza [2], and other diseases. Typically, in epidemic forecasting, model forecasts are probabilistic in nature, and several formats have been considered for their representation. In COVID-19 forecasting efforts, the forecasts for multiple locations and horizons (1-4 weeks ahead) are represented as predictive quantiles [1]. In this work, we investigate the utility of a trained linear ensemble model that assigns weights to individual models based on their historical performance.

The historical performance of model forecasts is determined through weighted interval scores (WIS), a proper scoring rule designed specifically for forecasts in a quantile format [3]. In this ensembling approach, the value of a specific quantile of the ensemble model forecast is the weighted sum of the values of the corresponding quantiles of the individual models. We consider forecasts from five different models, belonging to statistical, deep learning, and compartmental classes of models, as input to the ensemble. For comparison with other ensembling models, we included forecasts from the popular Bayesian averaging model. The performance is evaluated on forecasts of weekly COVID-19 cases. We observe that overall, the proposed ensemble model has a stable performance across forecast weeks. Ranking the models by their WIS scores, we observe that the proposed model is one of the top models in terms of the median rank (computed across different locations and forecast weeks). In addition, we observe a lower variability in its ranking across locations and weeks. The 95% and 50% coverage estimates of our ensemble model forecasts are also close to the nominal coverage. [1] Cramer, Estee Y., et al. "Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States." Proceedings of the National Academy of Sciences 119.15 (2022): e2113561119.[2] Reich, Nicholas G., et al. "A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States." Proceedings of the National Academy of Sciences (2019): 3146-3154.[3] Bracher, Johannes, et al. "Evaluating epidemic forecasts in an interval format." PLoS computational biology (2021).

The Effect Of Missing Value Imputation On Forecasting Population Abundance Depends On Both Model Structure And The Threshold Of Missing Values

Presenter: Vihanga Gunadasa
Co-authors: Vihanga Gunadasa;Glenda Wardle;Aaron Greenville

Missing values are common obstacles in time series data modelling and forecasting. However, the imputation of missing values can affect the model performance and will depend on how the model structure and assumptions are informed by the level of complexity of the system being modelled. To address this, we explore how the performance of a model varies with respect to 1) the number of imputed missing values, and 2) the number of parameters to be estimated. We use a Bayesian Multivariate State-Space Model framework on a long-term abundance data set for a mammal species exhibiting asynchronous population dynamics across three sites in arid central, Australia, and with contrasting levels of missing values (0%, 20% & 40%). To explore how the ecological assumptions of this study system are best included in the model we use a three-way factorial combination between 1) density dependence of abundances, 2) predator-prey interactions, and 3) the spatial independence of abundances, which can be understood by estimating the process error and observation error variances separately. The missing values were imputed through multiple imputation using chained equations (MICE). To avoid overfitting, forecasts were obtained through time series cross validation based on a rolling forecast origin. Forecast accuracy was evaluated using Root Mean Squared Error (RMSE) at each sequence of training and testing data and averaged across the sequences to obtain a comparative measure among each model combination. The results indicate that the RMSE of forecasts were lowest for models with less than 20% of missing values, equal error variances for process and observations, and included a predator-prey interaction. However, in the case where missing values were greater than 20%, forecasts were more accurate when observation error variance was allowed to vary across populations. Regardless of the number of missing values, the model combination that allowed the process error variance to vary by population had a higher RMSE, depicting the increased uncertainty of the forecasts. Insights from this data science study will be applied to forecast multi-species population dynamics and have implications for the many ecological monitoring studies with missing values.

Assessing Quality Of Statistical Scenarios In A Stochastic Optimization In An Energy Management System

Presenter: Evgenii Genov
Co-authors: Evgenii Genov;Julian Ruddick

Scenario generation and stochastic optimization are useful tools for energy management systems (EMS) in buildings, as they allow for the incorporation of uncertainty in energy demand and supply. While most research on combined stochastic forecasting and optimization problems has focused on wind power, these methods have shown promise in microgrid energy management, particularly in problems requiring collaborative agent involvement and high degrees of uncertainty. Forecasting, including scenario generation, plays an important role in such problems enabling the optimization algorithm to make more informed decisions. The degree of impact that scenarios can have on decision-making benefits is believed to be strongly associated with the quality of the scenarios. Traditional methods in scenario generation rely on evaluation of predictive marginal distributions and temporal correlation within the forecast horizon. Using theoretical analysis and running a case study, we investigate the scenario generation and evaluation methods in the context of their downstream utility in a multi-agent battery scheduling problem. The utility is considered in terms of energy cost, carbon emissions and grid stability. Some discrepancy is detected between theoretical and practical performances of scenario sets. We look closer into the causes of such discrepancies. Finally, we propose an approach that leverages the prior information about the EMS performance. Our research builds upon previous work in the field, with a focus on validating scenario generation methods in an online environment. Overall, our results suggest that incorporating advanced scenario generation methods can significantly enhance the capabilities of stochastic optimization in energy management applications.

Nowcasting The State Of The Economy: An Application Of Linear Combinations Of Dynamic Common Factors To The Colombian Economy

Presenter: Mario E. Arrieta-Prieto
Co-authors: Mario E. Arrieta-Prieto;Fabio H. Nieto

The main goal of this work is to propose a general methodology to create a coincident index for the economy of a given country or region based on linear combinations of dynamic common factors, and to validate it on simulated scenarios and on a case study for Colombia. The methodology proposed can effectively handle both stationary and nonstationary macroeconomic variables as input and provides tools to obtain point estimates and confidence regions, and to test hypotheses for the linear-combination coefficients. This work highlights how promising this new proposal is in terms of its contribution with respect to its antecedents in the literature, by showcasing situations in which a linear combination of dynamic factors can enhance the accuracy of the nowcast and address potential problems and limitations of considering only one dynamic factor as index. The application of this work to the Colombian economy is based on the macroeconomic analysis conducted by previous researchers. This work, however, considers and analyzes a much larger set of candidate indices via linear combinations of factors, providing consistent results; which strengthens and validates their previous findings.

Nowcasting Recession Risk In The Us And The Euro Area

Presenter: Francesco Furno
Co-authors: Francesco Furno;Domenico Giannone

This paper presents timely coincident recession risk indicators for the United States (US) and the Euro Area (EA) at a monthly frequency. Our indicators are constructed by estimating a parsimonious Bayesian logit based on two predictors which summarize financial conditions and real economic activity. We select the Composite Indicator of Systemic Stress (CISS) to measure financial conditions, and the US PMIs and the EA Economic Sentiment Index (ESI) to summarize real economic activity. These predictors are available immediately after the month of reference concludes. Back-testing the indicator over the periods 1980-2021 for the US and 1985-2021 for the Euro Area reveals a 96% and a 92% in-sample accuracy and 95% and 88% pseudo-out-of-sample, respectively. Our indicators are more accurate than popular indicators

such as the Sahm-Rule - especially at determining when the economy leaves a recession - and complement spread-based indicators which are good at forecasting, instead of nowcasting recessions.

Nowcasting World Trade With Machine Learning: A Three-Step Approach

Presenter: Baptiste Meunier
Co-authors: Baptiste Meunier;Menzie Chinn;Sebastian Stumpner

We nowcast world trade using machine learning, distinguishing between tree-based methods (random forest, gradient boosting) and their regression-based counterparts (macroeconomic random forest, linear gradient boosting). While much less used in the literature, the latter are found to outperform not only the tree-based techniques, but also more "traditional" linear and non-linear techniques (OLS, Markov-switching, quantile regression). They do so significantly and consistently across different horizons and real-time datasets. To further improve performances when forecasting with machine learning, we propose a flexible three-step approach composed of (step 1) pre-selection, (step 2) factor extraction and (step 3) machine learning regression. We find that both pre-selection and factor extraction significantly improve the accuracy of machine-learning-based predictions. This three-step approach also outperforms workhorse benchmarks, such as a PCA-OLS model, an elastic net, or a dynamic factor model. Finally, on top of high accuracy, the approach is flexible and can be extended seamlessly beyond world trade.

Nowcasting Inflation In Brazil Using Web Data

Presenter: J. Renato Leripio
Co-authors: J. Renato Leripio

Inflation expectations are essential inputs both for monetary policy and market participants, since they have a direct effect on the yield curve and stocks. Although the focus is mostly on the longer horizons, surprises in monthly releases can and do shift expectations accordingly. This is especially true when they occur in the so-called core items, which are associated with more persistent movements. However, the presence of some very volatile items with a high share in the consumption basket makes short-term inflation particularly hard to forecast. Furthermore, special events such as out-of-season sales pose an additional challenge that statistical models fail to capture. In this work, we explore price data for goods and services available on websites to improve the monthly forecast (or nowcast) of inflation in Brazil. Preliminary results are very encouraging and show that these data can greatly improve forecasting performance, with enormous potential to become a mandatory input in the near future.

Early Career Forecasters' Panel: How To Work Efficiently In An International Environment?

Presenter: Nikolaos Kourentzes; Fotios Petropoulos; Romain Michel Veyrune; Elaine Deschamps
Co-authors: NA

There are many opportunities to participate in forecasting projects with research groups and businesses around the world. At an early stage of your career, this way of working might sometimes seem overwhelming and chaotic. How do you efficiently manage such a project? How do you handle different time zones? How do you communicate effectively? Our guests will share their best practices and tips on communicating with your colleagues and managing your work efficiently. We will cover topics such as proper communication within a group, handling time zones while maintaining your work-life balance, efficient project management, and many more. Our guests will also answer your questions and discuss your potential doubts. With our guest from forecasting community we will discuss most crucial topics adn tips for starting your career.It would be a great opportunity to meet and netwrok with fellow forecasters as well!Hope to see you at the session.ECF team.

Watch your tone!: Forecasting Base Metal Prices with financial report tone

Presenter: Nicolas Magner
Co-authors: Nicolas Magner;Tiago Alvez;Nicolas Hardy;Jaime Lavin;Maria Jose Quinteros

We find evidence that information extracted from financial reports of U.S. corporations can be used to predict the future prices of some base metals. We concentrated on two variables in particular. First, we develop a novel measure of Knightian uncertainty based on uncertainty disagreement of U.S. firms. We employ a network method to calculate the minimum spanning tree length on uncertainty tone extracted from text mining of 837,342 financial reports. Second, we expand our test using the tone of financial reports from U.S. firms belonging to the mining industry. The results show that periods of increased uncertainty disagreement predict an increase in the price of the futures of based metals commodities. Additionally, a more positive tone in the financial reports forecasts a decrease in future prices. Our findings align with the present-value theoretical framework, as predictions made using a model that incorporates the uncertainty disagreement and the mining industry tone exceed those made without it in terms of Mean Squared Prediction Error. These results support the intuition that, in the case of mining companies, the commodities they produce are the main variables influencing their financial performance and, as a result, the price of based metal futures, so they are likely to report information capable of driving expectations about future commodity prices. Our findings support the hypothesis that the firm discloses information strategically to influence investor expectations for future performance. We contribute to the literature that proposes that the level of expectation dispersion predicts market behavior. Our findings also contribute to a better understanding of financial stability and assist practitioners working on delegated portfolio and risk management tasks.

Forecasting Corporate Treasury Operational Account Balances With Supervised Machine Learning Methods

Presenter: Jie Wu
Co-authors: Jie Wu;Jie Wu;Chun Lei He;Tamara Silvera;Diana Chan;Eugene Wen

Cash forecasting is critical for treasury departments as it is a key factor for financial planning and liquidity management. Traditionally it largely relies on spreadsheet-type manual processes which are prone to errors, time consuming and the accuracy is dependent on individual analysts' experience. In this work we tackle the problem by using supervised ML methods to provide automated and more accurate short term cash balance forecasts, for cash management optimization– to plan for upcoming cash gaps or make investments with surplus cash.In this prototype work, all balance and transaction data from four operational accounts at two business locations were extracted from Treasury Management System. Univariate time series were constructed and fit into Prophet in its original form as well as in cumulative sum space. Multiple features were extracted to construct multivariate time series: temporal features based on time, statistical features from transactions according to transaction types, source codes, amounts, credit/debit etc., rolling window time series features, along with features generated from automatic time series feature engineering algorithm. The multivariate feature sets were then fit into Xgboost regressor for training and forecasting. A comparison of performance is made between the models for each of the accounts. Our approach has shown promising results for producing forecasts of 30 days of cash balances.

"Fueling Predictability:" Can Commodity-Equities Forecast Fuel Prices?

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

In this paper we show that several MSCI stock indices have a remarkable ability to predict the returns of oil prices (WTI and Brent) and of three additional oil-related products: gasoline, propane and

heating oil. The theoretical underpinnings of our findings rely on the present-value theory for stock price determination and on the strong co-movement displayed by some industrial commodity prices. Interestingly, this predictive ability is stronger than that embedded in commodity-currencies. We find substantial evidence of predictability both in-sample and out-of-sample. One distinctive feature of our paper is a focus on MSPE differences at both the population and sample level. While several papers in the literature have recently found predictability for commodity returns at the population level, they typically tend to show poor gains in MSPE at the sample level. We address this failure with a simple approach based on the covariance between the target variable and our forecasts. With our approach we find substantial evidence of predictability at the sample level as well, in sharp contrast with the weak results reported by the traditional Giacomini and White (2006)/ Diebold and Mariano(1995)-West(1996) test when evaluating MSPE differences.

Enhancing Corporate Earnings Forecasts: A Generalized Additive Model With Approximate Nearest Neighbors Algorithm

Presenter: Levent Bulut
Co-authors: Levent Bulut;Sami Keskes

When forecasting corporate earnings, combining financial analysts' consensus forecast data with accounting data in a regression-based parsimonious model provides more accurate forecasts with less bias. However, traditional linear regression methods may not capture the effects of recent stock price movements or adequately incorporate financial analysts' past performance on other stocks. To address these limitations, we propose using a generalized additive model to capture any non-linearities in the data, and an approximate nearest neighbors (ANN) algorithm to account for analysts' performance on other stocks. We used earnings per share (EPS) forecast data from Institutional Brokers' Estimate System (I/B/E/S) provided by Wharton Research Data Services (WRDS), as well as accounting data from Compustat.

Alternative routine vaccination forecasts and demand plans - A Mozambique case study

Presenter: Laila Akhlaghi
Co-authors: Laila Akhlaghi;Kikelomo Lambo;Dauda Majanbu;Johnna Sundberg

JSI will describe the typical method used by country immunization programs for forecasting vaccines, primarily using demographic data; and the challenges this method can create. Alternate methods were applied to determine several different forecasts for routine immunizations in Mozambique showing that the demographic forecast primarily provided lower figures. Results and actions taken in Mozambique will be shared.Clinton Health Access Initiative (CHAI) will describe the state-specific method of forecasting vaccines in Nigeria with a focus on multi-level scenario forecasting within the set parameters of target population, coverage, and wastage rates. CHAI will describe how the resulting forecasted figures are used to engage the relevant stakeholders to ensure the availability and timely release of funds toward the procurement of life-saving vaccines in Nigeria. VillageReach will share insights on forecasting methods used by private providers in urban poor communities in Nigeria as a means to improve access to immunization. Forecasting in these environments is complicated by migrants and internally displaced persons i.e. they live in poor urban communities outside of formal urban centers, receive private health care, and are often not taken into account in census data and health system planning. The lack of data makes it difficult to forecast their needs accurately. Despite these challenges, private health providers in these environments still need to forecast demand and acquire supply. Dauda will share lessons learned from how practices in the private sector can influence forecasting in public health systems.Macro-Eyes will describe the challenges facing countries in forecasting vaccine demand at the health facility level, and how DHIS2 data can be used to overcome these challenges. Especially, data on monthly patient demand combined with supply chain data can be used with machine learning methods to improve over heuristic-based allocation methods, such

as the 3-month rolling average, even in the case of messy data with lots of stock-outs. Results from Sierra Leone will be shared.

A Multi-Level Approach To Vaccine Forecasting And Demand In Nigeria: Implications For Micro-Level Immunization Sustainable Financing

Presenter: Kike Lambo
Co-authors: Kikelomo Lambo

This article discusses a multi-level approach to vaccine forecasting and demand in Nigeria and its implications for micro-level immunization sustainable financing. Vaccination is a crucial public health intervention that has been demonstrated and documented to prevent the spread of infectious diseases and save lives, but vaccine availability and access remain a challenge in many low- and middle-income countries, including Nigeria. The multi-level approach to vaccine forecasting involves using data from different levels of the health system to inform vaccine supply and demand. This approach considers the supply and demand of vaccines at the national, state, and health facility levels. Demand for vaccines in Nigeria is influenced by several factors, including population size, disease burden, and vaccine acceptance. Micro-level immunization sustainable financing is critical for ensuring that health facilities have the resources they need to provide immunization services consistently. A multi-level approach to vaccine forecasting and demand can help improve vaccine availability and reduce stock-outs, which, in turn, can help increase vaccine coverage and prevent the spread of infectious diseases. However, implementing this approach requires sustained funding for data collection and analysis, as well as coordination among stakeholders. By implementing these recommendations, Nigeria can make progress towards achieving universal immunization and improving public health outcomes.

Forecasting Childhood Health Vaccination Demand With Machine Learning

Presenter: Johnna Sundberg
Co-authors: Johnna Sundberg

Forecasting health commodity demand at the facility level is challenging for many LMICs, but imperative to ensure access to critical health services. Often, there exists limited historical data, data quality issues, and frequent stock-outs that may be unreported. Together, these make forecasting using traditional statistical methods challenging. Instead, many countries rely on heuristics such as the 3-month rolling average to forecast at the facility level or the previous month's consumption. In this talk, Macro-Eyes will present a machine-learning alternative to both these heuristic and traditional statistical forecasting methods and share results from using these methods for childhood vaccine forecasting in Sierra Leone. First, we will describe how the machine learning algorithm adjusts for both reported and unreported stock-outs. Next, we will present a machine learning method that simultaneously learns from all products and facilities in order to generate more robust forecasts. We will conclude by sharing our results from using these methods in Sierra Leone and comparing our results to traditional and heuristic-based forecasting methods.

Vaccine Supply Forecasting By Private Health Care Providers Serving Urban Poor Communities: Lessons From The Private Sector

Presenter: Dauda Majanbu
Co-authors: Dauda Majanbu

The Gavi 5.0 strategy is focused on addressing equity in immunization and reaching under-served populations, especially zero-dose children—defined as children who don't receive a single dose of diphtheria, tetanus, and pertussis-containing vaccine (DTP1). Almost 50% of the zero-dose children live in urban areas,

remote communities, and populations in conflict settings. Urban poor communities are often challenged by a lack of infrastructure, poor housing, poor water and sanitation, and limited access to health services. One of the challenges in meeting the needs of under-served populations, especially the urban poor, is the lack of data which inhibits accurate forecasting for supply chain. Many internally displaced and migrants also live in poor urban communities on the periphery of formal urban centers and may not volunteer information about themselves or actively seek public services, relying on private healthcare. Further, they are often not counted in census data and included in health systems planning or supply forecasting. Private health providers often operate in such environments forecasting demand and acquiring supply to meet the needs. In Nigeria, 75% of the 60 million urban population live in slums. Considering the vital role of private service providers in improving access to immunization, this assessment in Nigeria explored forecasting methods used by private providers working in urban poor communities.

Timenet: A Large-Scale Benchmark Dataset For Time Series Forecasting

Presenter: Max Mergenthaler
Co-authors: Federico Garza

The availability of large-scale benchmark datasets has greatly facilitated the development of deep learning algorithms in fields such as image recognition and natural language processing. Such a dataset could widely improve the time series field. This paper introduces TimeNet, a comprehensive, large-scale dataset of over 10 million unique time series designed to serve as a reference for benchmarking time series forecasting models and algorithms. TimeNet will be freely available for non-commercial research purposes.

To construct TimeNet, we collate over 10 million time series from Wikipedia traffic data, offering a diverse and realistic set of time series characterized by varying degrees of seasonality, trends, and noise. We present a thorough analysis of TimeNet's statistical properties and provide statistical baselines including accuracy and computational complexity. These experiments aim to address longstanding questions in the field of time series forecasting, such as the effectiveness of global versus local models and the relationship between deep learning model performance and dataset size.

By encouraging further research and innovation in this area, we anticipate that TimeNet will lead to improvements in the accuracy and efficiency of time series forecasting models across a wide range of applications.

Polynomial Time Approximately Optimal Correlated Mechanism Design Using Constrained Proper Scoring Rules

Presenter: Michael Albert
Co-authors: NA

Fundamentally, mechanism design under settings of correlated valuations can be viewed as a peer prediction problem. Specifically, when a bidder reveals their evaluation for an item, it is an implicit prediction over the valuations of all other bidders. However, the implementation of these mechanisms requires a set of highly constrained proper scoring rules in order for the mechanism to achieve good performance. The computation of these proper scoring rules has traditionally been viewed as impractical due to a) the need to know the full distribution of valuations precisely and b) the exponential increase in computation time and space as the number of bidders increases. This is primarily due to the exponential size of the distribution in the number of bidders.In this work, we assume oracle sample access to the underlying correlated distribution, and we use the samples from the distribution to construct a nearly optimal mechanism in polynomial time in the number of bidders. We do this by a novel dimensionality reduction of the exponential size distribution to a polynomial size for each individual bidder, which allows us to construct the optimal constrained scoring rule for each bidder independently. This sidesteps the exponential blowup in the naive formulation of the problem. We also demonstrate the practical implementation of this mechanism design technique using data from a Google server cluster on resource requests.This dimensionality reduction

technique is applicable to the construction of constrained proper scoring rules where the prediction space is highly multi-dimensional, but the submitted information is low dimensional. We also propose ways in which reasonable assumptions about the underlying correlation structure can lead to even more efficient dimensionality reduction techniques.

Forecast Augmentation from multiple participants vertically (downstream and upstream) and horizontally (multiple participants at the same level) in a supply chain

Presenter: Salman Yousaf
Co-authors: Andrew Brooks

Torqata Data & Analytics, LLC is a Charlotte-based startup focused on helping the automotive aftermarket industry drive better decisions. As part of this, Torqata has taken on several cutting-edge research projects, including the creation of global time-series forecasting models that utilize data from multiple participants vertically (downstream & upstream) and horizontally (multiple participants at the same level) in a supply chain. We show that data sharing can increase demand forecasting accuracy for any individual participant when forecasting from a global model. This is driven both by the ability to create new features from the shared data that otherwise wouldn't be possible based on a single participant's data, but also by providing the models with larger samples of data to learn from. We walk through the theory, existing literature, research methodology, aggregated results, future direction, and implications for other industries. The talk will be presented via PowerPoint, and focus on a specific automotive aftermarket study.

Forecast Accuracy: How Much Is Too Much? Finding The Sweet Spot With Ceiling Analysis And Sensitivity Analysis

Presenter: Johann Robette
Co-authors: NA

Have you ever heard the saying "the sky's the limit"? Well, that's not always the case when it comes to Forecast Accuracy (FA) improvement. While there's always room for improvement, not all improvement leads to better decisions and added value.

So, how do we determine what improvements will actually make a difference?

That's where Ceiling Analysis and Sensitivity Analysis come in.

We'll delve into the power of these methods to determine the maximum potential profits and the quickest route to achieving them. With this information, it becomes easier to know where it makes sense to invest resources to improve FA and where it's not worth it.

In this presentation, you'll get a better understanding of these methods through a generic approach and results from an experiment conducted in the retail industry (M5 based). You'll see how Ceiling Analysis and Sensitivity Analysis can guide you to make wise decisions on where to allocate your resources for optimal results and where to avoid unnecessary spending.

Come along and discover a fresh perspective on improving Forecast Accuracy efficiently.

Enhanced Forecasting With Lstvar-Ann Hybrid Model: Application In Monetary Policy And Inflation Forecasting

Presenter: Michal Chojnowski
Co-authors: Michal Chojnowski

The author proposes and investigates the features of a new LSTVAR-ANN hybrid model. The neural network in the LSTVAR-ANN model enables to capture of complex relations between a vast dataset (here: Google Trends) and the transition variable (here: customer confidence).The author has chosen monetary policy and inflation forecasting as a topic of the analysis. The model assumes the existence of two economic regimes, which can be identified with a transition variable. Google Trends phrases were carefully preselected to identify the most influential drivers for customer confidence. The LSTVAR-ANN helps to select Google Trends, aggregate them and extract from them unobservable factors influencing customer confidence.Additionally, the model allows calculating an impulse response function conditional on customer confidence value. Hence the author can estimate the impact of monetary policy changes on inflation depending on the current environment. Moreover, LSTVAR-ANN's forecasting capabilities were checked compared to a benchmark of five "classical" time-series models. The results show room for improvement in forecast accuracy, which is expected from neural networks in macroeconomic forecasting literature.The author believes that in uncertain times LSTVAR-ANN model is a valuable tool for macroeconomic forecasters.

Machine Learning For New Product Forecasting

Presenter: Mohsen Hamoudia
Co-authors: Mohsen Hamoudia;Mohsen Hamoudia;Lawrence Vanston

Forecasting the demand for new products is crucial given the level of investment required for a launch. It's also increasingly challenging and risky in an environment of vigorous economic competition, evolving customer expectations, and the emergence of new technologies and innovations. Given the high failure rate of new launches (70-80 percent for consumer-packaged goods), the accuracy of demand forecasts is a top priority for decision-makers. Underpredicting demand leads to a loss of potential sales; overpredicting it leads to costly excess inventory. Forecasting new product demand has traditionally been done using a variety of techniques: judgmental methods, market research like surveys of buyer's intentions, market testing, expert opinion methods like the Delphi method, diffusion models like the Bass model, and statistical modeling through a combination of time series and/or multivariate techniques. More recently, Machine Learning has been added to the mix. The selection depends somewhat on whether the new product is: (a) new to the world, (b) new to the firm, (c) an addition to existing product lines, or (d) an improvement or revision to existing products. A common forecasting method is to use the demand for existing products to forecast the demand for new ones. This method is simple if the new product is a variation of an existing product. If, on the other hand, the product is completely new, we may have lots of data on similar products but what that data means for the forecast is unclear. In such situations, numerous studies have shown that Machine Learning can provide the answer and produce a better forecast.Many papers and case studies are available on forecasting new products with historical data. When it comes to new products with little or no history, the literature is very limited. In this paper, we will review the main Machine Learning methods for predicting new product demand. Based on some recent case studies, we also will assess if these methods have improved forecast accuracy.

Prospects Of Ai For Long-Term Forecasting

Presenter: Lawrence Vanston
Co-authors: Lawrence Vanston

The rate of progress in AI, specifically machine leaning (ML), has been notable for both forecasting and other applications. For example, prior to the M4 forecasting competition, statistical methods proved superior to ML, in the M4 competition, ML proved its value in hybrids with traditional statistical methods, but by the M5 competition, ML owned the leaderboard. Further, ML has begun to displace, or at least supplement, statistical methods in both research and practice.However, most of this experience has been

for short-term forecasting where there is lots of historical data and reasonable continuity. What about long-term forecasting where there is little or no historical data and the future may be substantially different than the past? And where the resulting decisions may be truly strategic and, thus, demanding documentation of the logic behind the forecast? We begin addressing these questions by outlining some of the differences between short-term and long-term forecasting. These include long-term forecasting's reliance on analogies, technological principles (e.g. adoption, performance, and learning curves), relationship among trends (e.g. price-demand), external research, subject area knowledge, analysis of drivers and constraints, expert opinion, and judgment. While some of these aspects of long-term forecasting (analogies, for example) creep into current AI-based short-term forecasts, especially when using global models, we suspect that others do not.In this paper, we review the experience to date with AI for long-term, or at least longer-term, forecasting. We also discuss, and open for discussion, the prospects for AI to incorporate or improve on the knowledge and wisdom that humans have developed to address complex problems with lots of moving parts and profound uncertainty.

Master Data Management: A Fundamental Element Of Supply Chain Management In An Mro Environment

Presenter: Stephen Spulick
Co-authors: Stephen Spulick

Master data management (MDM) is a critical aspect of supply chain management, as it involves the collection, storage, and maintenance of the key data that drives business processes. This includes data on products, customers, suppliers, and other key stakeholders. Effective MDM allows organizations to gain a single, consistent view of their data, enabling them to make better informed decisions, streamline operations, and improve efficiency.One of the key benefits of MDM in the supply chain is improved data quality. Poor quality data can lead to a range of issues, such as incorrect forecasting, inefficient planning, and missed opportunities. MDM helps to ensure that data is accurate, consistent, and up to date, enabling organizations to make more reliable and informed decisions.Another important aspect of MDM in the supply chain is the ability to gain greater visibility into the supply chain. With accurate and consistent data, organizations can track the movement of goods and materials, identify bottlenecks, and optimize logistics. This can help to reduce costs, improve customer satisfaction, and increase competitiveness.MDM can also help organizations to manage risk in the supply chain. For example, by tracking supplier performance and identifying potential issues, organizations can proactively address potential disruptions before they occur. This can help to minimize the impact of supply chain disruptions on the business, and ensure that operations continue to run smoothly.In addition, MDM can support the development of new products and services, as well as the expansion into new markets. By having a clear understanding of customer and market needs, organizations can more effectively design and deliver products and services that meet these needs.Overall, the importance of MDM in the supply chain cannot be overstated. It is a key foundation for effective supply chain management, and is essential for organizations looking to improve efficiency, reduce costs, and increase competitiveness.

Multiple Horizons, Partial Observability And Scale: The Challenges Of Predicting Materialization Rates In One Of The World's Largest Shipping Networks

Presenter: Marco Zugno
Co-authors: Marco Zugno;Lasse Petersen;Julija Tastu

Just like in other industries, knowing at what rate existing bookings will materialize is critical within shipping and logistics. Predictions of booking materialization rates can be employed by a shipping company in a range of commercial and operational decision-making processes.In essence, predicting the materialization rate is a binary classification problem (either cargo materialized or not) conditional on a set of features of the booking. However, this forecasting problem has two characteristics that deviate from the standard

"textbook" classifier formulation.The first characteristic is the multiplicity of horizons. Since bookings are received continuously, from different locations and with different time-to-departures, decisions must be constantly re-evaluated as new information becomes available. This implies that materialization rates must be updated throughout a booking lifetime, rather than just estimated at booking date.The second feature of the problem is the partial observability of actual outcomes. Indeed, the carrier can intervene and alter features of bookings, which potentially affect customer satisfaction, e.g., if they result in a delayed arrival of the cargo at destination. In such cases, the "pure" (i.e., without carrier intervention) customer materialization behavior, which is the modelled response, is not observed.We address the multiple-horizon issue by using a landmarking setup, where each valid observation enters the training set multiple times with different horizons to a time-of-interest (departure), in a binary classification model. To keep results unbiased under the presence of unobserved data, we apply an Inverse Probability of Censoring Weight (IPCW) approach to weigh valid observations based on estimated probabilities of carrier intervention.We employ Gradient Boosted Trees to solve the resulting large-scale, IPC-weighted classification problem. Results show decreased bias compared to naïve approaches that disregard unobserved data, as well as positive skill score with respect to simpler benchmarks.

Publishing In Foresight

Presenter: Michael Gilliland
Co-authors: Michael Gilliland

This session is for anyone interested in publishing in Foresight.Foresight: The International Journal of Applied Forecasting is a quarterly publication by the IIF that is oriented toward forecasting practitioners. Foresight publishes several types of content including articles, commentaries, book reviews, interviews, tutorials and minitutorials, and opinion-editorial pieces. All content, including articles based on academic research, focus on the practical "takeaways" that forecasters can put to use in their daily roles. All submissions are reviewed by the editorial staff for topic relevance and are edited for clarity of presentation.This session will describe Foresight's mission and scope, the manuscript submission and editorial review process (including timelines), and the kinds of topics most relevant to our readers. Audience questions will be addressed by Foresight's EiC and members of the editorial staff.

Is There A Hype Bias? The Perception Of Ai In Demand Forecasting

Presenter: Anna Sroginis
Co-authors: Anna Sroginis;Tove Helldin;Nikolaos Kourentzes

Many of the forecasting methods available to companies have a proven track record of high accuracy. Nonetheless, they are frequently adjusted by human experts to incorporate additional information. This information is often unstructured and difficult to systematically add to statistical models. Extensive research has shown that the performance of judgemental adjustments of model outputs is inconsistent. Although there are cases where they add value to predictions (for example, incorporating special events), on average adjustments, tend to harm accuracy. Furthermore, users often have a limited understanding of statistical methods and have been found to be averse to model errors, resulting in experts increasingly relying on their judgement due to lack of trust and understanding of the models available to them. Recently, we have observed the rapid rise of artificial intelligence (AI) in predictive tasks, often touted as working out of the box, and being able to handle unstructured rich contextual information. Some empirical evidence supports such claims, adding to the general hype of AI. Nonetheless, AI methods are typically black box and mathematically intractable. Explainable AI (XAI) methodologies attempt to increase the transparency of the models, however there is limited evidence of the success of XAI tools in business environments. It is arguably reasonable to assume that the average user has limited understanding of the inner workings of AI models. We hypothesise that users can exhibit a "hype bias" where they will be less averse to AI implementations, compared to conventional modelling (primarily statistical) approaches, even though

they are complex and opaque, due to hype. We investigate this hypothesis with a laboratory experiment. Simulating characteristics of well-designed forecasting support systems available in the market, we design a set of controlled experiments in which users are asked to forecast future demand. We aim to analyse human behaviours by providing both statistical and AI forecasts with or without model details, exploring for cognitive biases towards more complex methods simply due to their hyped performance. The results help understand how users might use algorithm advice with different complexity and perceived value, and furthermore guide users and software houses on how to balance statistical and AI models to support predictive tasks.

Discerning The Relationship Between Demand Realization, Task Information, Performance Information, And Supply Chain Forecasting

Presenter: Niles Perera
Co-authors: Niles Perera;Dilina Kosgoda;Shari De Baets

Organizations employ forecasting systems including applications of artificial intelligence, machine learning and big data analytics to derive demand forecasts. Even though forecasters have access to automated systems, they tend to manually adjust system-generated forecasts. This can be due to many reasons, ranging from persistent biases to the rational view that they may improve forecast accuracy by incorporating real-time (contextual) information that they assume to have not been considered in the system. Forecasters may be influenced by provision of relevant information in the latter case. Forecasters may receive guidance pertaining to forecast accuracy in previous periods as well as information regarding demand realization when determining their present forecast adjustments. It remains unknown whether forecasters can effectively use this information. We studied the effect of providing performance information, demand realization and task information for forecasters using a computerized laboratory experiment, to explore the effect of these types of information on forecast adjustment behavior and subsequent effects on accuracy. Our laboratory experiment consisted of 40 rounds, including 10 practice rounds to familiarize participants with the task. We recruited 173 students studying operations and supply chain management at the undergraduate level from a Sri Lankan university, who were randomly assigned to four conditions. Participants in all four conditions were asked to forecast demand for a particular period. They were free to either use the system forecast or adjust it. Participants in condition 1, condition 2 and condition 3 had access to performance information, demand realization and task information respectively. Analysis of the results reveal that participants who had access to previous performance information or demand realization were able to improve the accuracy of their final forecast (after adjustments). However, we found no significant effect of task information. Our analysis indicates that the Mean Absolute Percentage Error (MAPE) when participants had performance information is lower than the other tested scenarios. Participants with access to demand realization and task information had the 2nd and 3rd lowest MAPE respectively. Based on these preliminary results, we suggest that providing performance information and demand realization in decision/forecast support system could play a vital role in improving forecast accuracy.

Forecast Value Added In Demand Planning: Questions For Meta-Analysis Of Empirical Evidence

Presenter: Shari De Baets
Co-authors: Shari De Baets;Robert Fildes;Paul Goodwin

'Forecast value added' (FVA) is a term commonly used to measure the improved accuracy achieved by judgmentally modifying a set of forecasts produced by statistical methods or algorithms. Assessing the factors that prompt such adjustments, and when they are likely to improve accuracy, is important in company demand forecasting and planning but has not been studied sufficiently. The published research has taken various individualistic approaches, both in the questions examined and the data analysis and modelling. In this paper we have collected the publicly available data from these studies, six in total,

to analyse them using a common framework. Questions include when do demand planners adjust their statistical forecasts, do adjustments improve accuracy and reduce any bias, does the size of the adjustment signal a more substantive and useful piece of information gathered by the demand planner, and are improvements are consistent across companies? These questions are important in practice since the costs of error are substantial, while the process of adjustment is expensive and time consuming, but they are also theoretically interesting raising the question of why consistencies across companies arise and the circumstances when one organization is more effective than another. The key question is how organizations can improve on their current forecasting processes to achieve greater 'forecast value added'.

The Predictive Power Of The Yield Curve, Factors And Foreign Interest Rates For Economic Activity Across Countries

Presenter: Menzie Chinn
Co-authors: Menzie Chinn;Laurent Ferrara

We re-examine the evidence for the term spread, augmented by additional domestic and international factors, as a predictor for future economic activity in developed economies, including the US (following Kucko and Chinn, Int'l Finance 2015), and four emerging market economies. We believe a re-examination is warranted given advances in reducing the dimensionality of sets of macroeconomic variables using factor models, and integration of international financial markets. Economic activity is measured using a recession indicator, and industrial production growth. We examine the sensitivity of the results to the selection of countries, and time periods. We find that the predictive power of the yield curve in developed economies has deteriorated in recent years, while never particularly high in non-developed economies. Estimated factors, financial conditions indices, and foreign interest rates are also assessed in terms of predictive power.

Nowcasting Consumer Price Inflation Using High-Frequency Scanner Data: Evidence From Germany

Presenter: Elisabeth Wieland
Co-authors: Elisabeth Wieland;Jan-Oliver Menz;Günter Beck;Kai Carstensen;Richard Schnorrenberger

Using a granular set of high-frequency scanner data, we explore their potential to nowcast monthly inflation in Germany. First, we compute a large set of price indices at the product level and show that these scanner-based indicators map the respective official series well, with a correlation coefficient of 0.8 for German food inflation. Second, we combine our weekly scanner-based price indicators with monthly official price indices within a Mixed-Data Sampling (MIDAS) nowcasting approach. We find that high-frequency information in the first week of a given month already contains valuable predictive content about that month's inflation rate. Overall, for major food and beverage subcomponents, the average predictive gains of the MIDAS nowcasting strategy yield 30-50% compared to the AR(p) benchmark.

Improving Output Gap Estimation – A Bottom-Up Approach

Presenter: Sina Streicher
Co-authors: Sina Streicher;Alexander Rathke

We propose a multidimensional Bayesian state-space model to identify potential output and the output gap consistent with the dynamics of the underlying production sectors of the economy and those of inflation and the labor market. Our model connects output to employment and unemployment via Okun's law and captures inflation dynamics via a Phillips curve relationship. The structure of the model is inspired both by Jarocinski and Lenza (2018) and Hasenzagl et al. (2022). Its primary innovation is the integration of consistent trends and cycles of sub-sector output and employment. In addition, our approach allows us

to decompose the business and employment cycles and respective long-term trend growth into their driving factors. Finally, tracking the cycles of individual sectors allows policy actions to be targeted at specific industries, thereby increasing their efficiency and reducing the chance of pro-cyclical outcomes.Applying our model to the Swiss economy reveals substantial divergence among the considered production sectors. The business cycle and the growth potential of the Swiss economy are most clearly influenced by the sectors that are most dependent on the global economy – manufacturing and financial and other economic services. While activities in trade, transport and hospitality are responsible for the slow decline in trend growth, manufacturing is counteracting this development. A comparison to established estimation approaches shows that our enriched information set can help paint a more comprehensive picture of the business cycle. Hasenzagl, T., Pellegrino, F., Reichlin, L., and Ricco, G. (2022). A model of the Fed's view on inflation. Review of Economics and Statistics, 104(4):686–704.Jarocinski, M. and Lenza, M. (2018). An inflation-predicting measure of the output gap in the euro area. Journal of Money, Credit and Banking, 50(6):1189–1224.

The Power of Many: The Procrustes Approach to Proxy-SVAR Identification with Multiple Instruments

Presenter: Srecko Zimic
Co-authors: Srecko Zimic;Skander Garchi Casal

This paper proposes a novel methodology to identify structural VARs using multiple proxy instruments. The proposed methodology does not impose any restrictionson the total number of proxy variables used and allows for the incorporation of plausibly exogenous and/or weak instruments. In a nutshell, our method identifies theshocks maximizing their contribution to the instruments' forecast error variance decomposition. Monte Carlo experiments suggest that a complete set of instruments(meaning one instrument per shock) can correct for biases generated from instruments' cross-correlations. Additionally, the presence of exogenous instruments can substantially improve the identification of other shocks. We apply our method to a small scale VAR characterizing classical NKM-model and identify demand, supply and monetary drivers in US data. In the second application we identify separately credit supply and monetary policy shocks for the European monetary union.

An Empirical Evaluation Of Some Long-Horizon Macroeconomic Forecasts

Presenter: Kurt Lunsford
Co-authors: Kurt Lunsford;Kenneth West

We use long-run annual cross-country data to evaluate pseudo out-of-sample forecasts of five variables for horizons up to 50 years. The variables we forecast are real per capita GDP growth, CPI inflation, labor productivity growth, and long- and short-term nominal interest rates. Our models for forecasting include simple time series models and frequency domain methods recently developed in Müller and Watson (2016). We focus on coverage of 68% forecast intervals (that is, coverage of 68% confidence intervals for forecasts). For GDP growth, CPI inflation and labor productivity growth, median coverage across countries is roughly 68% for several models, but with considerable dispersion around that median. For these three series, a reasonable model choice is a frequency domain model that does not require the user to take a stand on the order of integration of the data. For interest rates, forecast intervals for most models and samples include markedly fewer than 68% of realized values. For interest rates, a reasonable model choice is a driftless random walk.

Dynamics Of Kimchi-Premium And Its Determinants

Presenter: Bonsoo Koo
Co-authors: Bonsoo Koo;Yangzhuoran Yang;Myung Seo

With the rapid growth of cryptocurrency as an investment vehicle, it is of interest to understand Kimchi premium, the persistent non-zero premium that exists between the US and Korean crypto-markets. Not only does the premium represent a violation of the one price law but it may also reflect the bubble aspect of crypto-markets or crypto-market segmentation. Contrary to the literature which relies on linear modelling, we employ threshold cointegration with multiple regimes to study the nonlinear dynamics of Kimchi premium and its determinants. The premium has been found to be mean-reverting outside of a certain range of past values of the premium and a unit root inside the range. In the long run, Kimchi premium is estimated to have a steady state level of 1.02% for Bitcoin. We also find the trading fee, as a form of market friction, is positively correlated with the positive threshold of the unit root range, which implies arbitrage trading has an impact on the dynamic of Kimchi premium.

Mitigating The Choice Of The Duration In Ddms Models Through A Parametric Link

Presenter: Fernando Mendes
Co-authors: Fernando Mendes;Douglas Turatti;Guilherme Pumi

One of the most important hyper-parameters in duration-dependent Markov-switching (DDMS) models is the duration of the hidden states. Because there is currently no procedure for estimating this duration or testing whether a given duration is appropriate for a given data set, an ad hoc duration choice must be heuristically justified. This is typically a difficult task and is likely the most delicate point of the modeling procedure, allowing for criticism and ultimately hindering the use of DDMS models. In this paper, we propose and examine a methodology that mitigates the choice of duration in DDMS models when forecasting is the goal. The idea is to use a parametric link instead of the usual fixed link when calculating transition probabilities. As a result, the model becomes more flexible and any potentially incorrect duration choice (i.e., misspecification) is compensated by the parameter in the link, yielding a likelihood and transition probabilities very close to the true ones while, at the same time, improving forecasting accuracy under misspecification. Two Monte Carlo simulations, based on classical applications of DDMS models, are employed to evaluate the methodology. The results demonstrate that using the Aranda-Ordaz link function leads to more precise forecasts and transition probabilities, not only in cases of duration misspecification but also when the model is correctly specified. Furthermore, an empirical study is conducted to forecast the volatility of the S&P 500, which illustrates the effectiveness of the proposed methodology. The results indicate that the Aranda-Ordaz DDMS outperforms fixed logit transitions in terms of forecast precision for most duration parameters. Moreover, the Aranda-Ordaz link function improves the forecasting performance to such an extent that it is equivalent to MS-Garch models. Notably, such an improvement in forecasting accuracy is not observed when using fixed link functions.

An Unpleasant Feature Of The Historical Mean Forecast

Presenter: Pablo Pincheira
Co-authors: Pablo Pincheira;Nicolás Hardy;Andrea Bentancor

The historical mean forecast has been widely used as a benchmark in the finance literature. Yet, our empirical results show an unpleasant feature of this popular benchmark: it is negatively correlated with the target variable. We show this empirical regularity using a big database including one period returns of 12 floating exchange rates, 15 industrial commodities and 7 equity indices in daily, monthly and quarterly frequencies. Our analysis reveals that this negative correlation is a spurious artifact stemming from the combination of finite samples with the overlapping nature of traditional out-of-sample analysis making use of rolling or expanding windows. In particular, our Monte Carlo simulations using i.i.d. observations, show a clear downward bias in the distribution of sample correlations of the historical mean with these independent and simulated target variables. We propose an adjustment in the calculation of correlations

that works well when the historical mean is used to predict variables displaying low persistence, such as short term financial returns. Our adjustment also works well in tests evaluating the null hypothesis of zero covariance between the forecast and the target variable. Without the adjustment we found statistically significant evidence of negative correlation of the historical mean with the predicted financial returns in 62% of our exercises. With the aid our adjustment, this percentage sharply reduces to 2%.

State-Level Spatio-Temporal Forecasting Of Covid-19 Hospitalization With Uncertainty

Presenter: Jingyuan Chou
Co-authors: Jingyuan Chou;Adiga Aniruddha;Jiangzhuo Chen;Madhav Marathe

We often observe a lag in the timing and pattern of epidemics spread across different regions. These spatio-temporal spread patterns and dependencies in epidemics spread across regions are governed by multiple factors such as human mobility, social behaviors, environmental factors, etc. Understanding and incorporating these patterns into a forecasting model is important but often difficult due to insufficient data on the governing factors. In addition, the dependencies are often nonlinear and time-varying. Under these circumstances, one approach often considered is extracting the patterns in the epidemiological signals of interest, for example, COVID-19 cases, deaths, or hospitalizations time series, and incorporating them into the model. In this work, we develop a deep neural network framework to model the patterns in spatio-temporal data. Specifically, we consider graph neural networks (GNNs) which provide a natural framework to connect regions and extract the nonlinear spatial dependencies in the data. A key requirement for the GNN is the specification of the graph structure. Multiple graph structures can be derived based on adjacency between regions, mobility, or correlation between signals across regions. In addition, we employ a Long Short-term Memory (LSTM) network to capture the temporal patterns in the time series data. The resulting network is a multivariate forecasting model that is applied to forecasting 1-4 weeks ahead COVID-19 hospitalizations across US states. As specified by the US COVID-19 ForecastHub, we provide probabilistic forecasts in terms of predictive quantiles [1]. As a case study, we selected five states, Virginia, Georgia, North Carolina, South Carolina, and Tennessee, and considered the adjacency-based graph in our model. We observed that our method was able to provide accurate forecasts. We compared the forecast performance with the COVID-hub Ensemble model forecasts. COVID-hub Ensemble is a median-based ensemble model that combines forecasts from dozens of teams [1]. Our model outperformed COVID-hub Ensemble across most states and horizons. We continue exploring various graph structures and extend the model to forecast for all states and counties. In addition, we are exploring techniques to integrate heterogeneous datasets that guide graph construction.[1] Cramer, Estee Y., et al. “The united states covid-19 forecast hub dataset.” Scientific data 9.1 (2022): 462

Infectious Disease Forecast Evaluation Based On Realized Social Utility

Presenter: Aaron Gerding
Co-authors: Aaron Gerding;Evan Ray;Nicholas Reich

The COVID-19 pandemic has led to rapid methodological and practical innovation in the elicitation and evaluation of forecasts of infectious disease burdens, a primary goal being to help public health workers make informed decisions about how to manage these burdens in a transparent manner. An implicit expectation and motivation in these efforts has been that refined evaluation metrics and corresponding elicitation procedures will steer the epidemiological modeling and forecasting community toward practices which demonstrably yield added utility for society through the decisions they support. Explicit descriptions or quantifications of such utility are, however, elusive. And moreover there has only been limited discussion of how predominant evaluation metrics, such as quantile-based proper scoring rules for probabilistic incidence forecasts, might be indicative of the success policies tethered to those forecasts will have with the type of multivariate stochastic optimization tasks public health policy makers constantly confront during

an infectious disease emergency.Here we pursue one possible tether between multivariate forecasts and policy: the allocation of limited medical resources in response to COVID-19 hospitalizations in various regions so as to minimize expected unmet need. We formulate a standardized allocation algorithm following techniques developed in operations research to solve constrained newsvendor problems with multivariate demands. We then score forecasts according to how much unmet need their associated allocations would have allowed. This scheme is implemented with respect to a hypothetical limited resource for quantile forecasts of COVID-19 hospitalizations in the US at the state level recorded in the COVID-19 Forecast Hub. The forecast skill ranking given by this allocation scoring rule can vary significantly from the ranking given by the weighted interval score now used by the CDC, especially during surges in hospitalization such as in late 2021 as the Omicron wave began. We see this as strong evidence that the allocation scoring rule detects forecast value that is missed by traditional accuracy measures and that the general strategy of designing scoring rules directly linked to policy performance is a promising research direction for epidemic forecast evaluation, ranking, and combination.