Essays on Time Series Forecasting: Applications in Macroeconomics and Finance

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
M.Sc.
Richard Schnorrenberger
aus Itapiranga

Kiel, 2024

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan:
Prof. Dr. Christian Martin

Erstberichterstattender:
Prof. Dr. Kai Carstensen

Zweitberichterstattender:
Prof. Dr. Jens Boysen-Hogrefe

Drittbegutachtung:
Prof. Dr. Matei Demetrescu

Tag der Abgabe der Arbeit:
18. März 2024

Tag der mündlichen Prüfung:
17. Juni 2024

Acknowledgements

The journey of crafting this thesis has been a challenging yet profoundly rewarding endeavor. I would like to thank everyone who has been a major part of it. First and foremost, I would like to express my gratitude to my PhD supervisor, Prof. Dr. Kai Carstensen, who provided first-class guidance for the success of this work. He made a fundamental contribution to shaping my academic trajectory and taught me a lot about the pathways of sound research in econometrics. Under his mentorship, I sharpened my capability to navigate challenging econometric landscapes with confidence.

I am indebted to the coauthors who significantly contributed to the chapters contained in this work: Prof. Dr. Kai Carstensen, Dr. Elisabeth Wieland, Dr. Jan Oliver-Menz, Prof. Dr. Günter W. Beck, M.Sc. Aishameriane Schmidt, and Prof. Dr. Guilherme Valle Moura. A special mention is reserved for Dr. Elisabeth Wieland, my supervisor during my PhD internship at the Deutsche Bundesbank. Her friendly hospitality and support seamlessly integrated me into the team. She provided me with unique opportunities to showcase my work while nurturing my passion for applied forecasting.

This thesis was written during my work as a research and teaching assistant at the Institute for Statistics and Econometrics at Kiel University. I would like to thank all my colleagues at the Institute for a pleasant time in a friendly and warm atmosphere. I am also grateful for the exceptional exchange of ideas with colleagues and students, which provided me with fresh perspectives for my research endeavors.

Above all, I am truly thankful to my wife, Karina Buss, whose unwavering belief in my potential and support emboldened my decision to pursue a PhD abroad. She has been extremely loving and supportive while having patience during the stressful times of my PhD journey.

Contents

List of Abbreviations V

List of Tables VI

List of Figures VIII

General Introduction 1

1 Harnessing Machine Learning for Real-Time Inflation Nowcasting 4

1.1 Introduction . . . 5
1.2 Data . . . 9
1.3 Methodology . . . 14
1.3.1 Nowcasting setup . . . 14
1.3.2 Machine learning methods . . . 16
1.4 Empirical results and discussion . . . 21
1.4.1 Out-of-sample results . . . 22
1.4.2 Interpreting the best-performing model . . . 26
1.4.3 A deeper assessment of key modeling features: guiding accurate inflation nowcasts . . . 28
1.5 Summary and conclusions . . . 32

2 Nowcasting Consumer Price Inflation Using High-Frequency Scanner Data 39

2.1 Introduction . . . 40
2.2 Literature review . . . 42
2.3 Data . . . 45
2.3.1 Fast-moving consumer goods (GFK:FMCG) . . . 45
2.3.2 Energy and travel services . . . 48
2.3.3 Descriptive statistics . . . 49
2.4 Nowcasting strategy . . . 51
2.4.1 Benchmark nowcasts . . . 53
2.4.2 Step 1: nowcasting item-level inflation rates . . . 53
2.4.3 Step 2: nowcasting product group-specific inflation . . . 55

2.4.4 Step 3: nowcasting headline inflation . . . 56
2.4.5 Real-time information set on non-standard policy measures, annual updates of HICP weights and imputed prices in times of crisis . . . 59
2.5 Results . . . 60
2.5.1 Results of the item-level inflation nowcasts . . . 60
2.5.2 Results of the product group-specific inflation nowcasts . . . 63
2.5.3 Results of the headline inflation nowcasts . . . 65
2.6 Robustness analysis . . . 70
2.6.1 Alternative methods of compiling COICOP-10 price indices . . . 70
2.6.2 Smoothing out volatile inflation series . . . 71
2.6.3 Alternative U-MIDAS specifications and ML hyperparameter choices 71
2.7 Conclusion . . . 72

3 Bond Portfolio Optimization in Turbulent Times: a DNS-WSV Approach 86
3.1 Introduction . . . 87
3.2 Yield Curve Modeling . . . 89
3.2.1 The baseline dynamic Nelson-Siegel model . . . 89
3.2.2 The DNS with Wishart stochastic volatility . . . 91
3.3 Bayesian Posterior Analysis . . . 93
3.3.1 MCMC algorithm . . . 93
3.3.2 Parameter restrictions and prior choices . . . 94
3.3.3 Gibbs sampling . . . 95
3.4 Empirics . . . 96
3.4.1 Data . . . 96
3.4.2 Estimation results . . . 98
3.4.3 Forecasting exercise . . . 100
3.4.4 Bond portfolio optimization . . . 104
3.4.5 Value-at-risk forecasting and backtesting . . . 108
3.5 Concluding remarks . . . 111

List of Abbreviations

AR autoregression

BART Bayesian additive regression trees

BCB Brazilian Central Bank

COICOP-10 classification of individual consumption by purpose at the item-level

COVID-19 coronavirus disease 2019

CPI consumer price index

CUMSFE cumulative sum of squared forecast error

DGP data generating process

DM Diebold-Mariano

DNS dynamic Nelson-Siegel

ENet elastic net

ESS effective sample size

EWMA exponentially weighted moving average

GDP gross domestic product

GFK:FMCG household panel of the market research company GfK containing daily purchases of fast-moving consumer goods

HICP harmonised index of consumer prices

IPCA Brazilian broad national CPI

LASSO least absolute shrinkage and selection operator

LLF local linear forest

LPDS log-predictive density scores

MCMC Monte Carlo Markov chain

MH Metropolis-Hastings

MIDAS mixed-data sampling

ML machine learning

MSV multivariate stochastic volatility

NEIG non-energy industrial goods

OLS ordinary least-squares

RF random forest

RMSE root mean squared error

RW random walk

SD-AR seasonal dummy autoregression

sg-LASSO sparse-group LASSO

SPF survey of professional forecasters

SSM state-space model

SV stochastic volatility

TPD time-product dummy

TVP time-varying parameters

UK United Kingdom

U-MIDAS unrestricted MIDAS

US United States of America

VaR value-at-risk

VAR vector autoregression

VAT value-added tax

WOB Weekly Oil Bulletin

WSV Wishart stochastic volatility

List of Tables

1.1 Real-time database of Brazilian macro-financial indicators 13
1.2 Summary of the ML competing models 17
1.3 RMSE: ML methods relative to the SPF benchmark 23

2.1 Illustrative example of the household scanner data (GFK:FMCG) 46
2.2 Mapping between high-frequency price data and the German HICP 50
2.3 RMSE for FMCG product-level inflation: U-MIDAS relative to the SD-AR benchmark 62
2.4 RMSE for FMCG product-group inflation: shrinkage methods relative to the SD-AR benchmark 65
2.5 RMSE of headline inflation and its components: bottom-up U-MIDAS and direct ML approaches relative to the bottom-up SD-AR benchmark 66
2.6 Data sources of high-frequency price series and official inflation in Germany 74
2.7 Summary statistics of GFK:FMCG 76
2.8 RMSE for FMCG product-level inflation: OLS-match relative to the SD-AR benchmark 82
2.9 RMSE of headline inflation and its components: bottom-up OLS match approach relative to the benchmark approach 83

3.1 RMSE and LPDS: competing models relative to the RW benchmark 103
3.2 Performance of optimal bond portfolios with daily rebalancing 107
3.3 Value-at-Risk forecasting results for optimal portfolios. 110
3.4 Descriptive statistics of US Treasury yields 114
3.5 MCMC posterior estimates of Wishart parameters in DNS-WSV 119

List of Figures

1.1 Time series of Brazilian price indicators, 2001 – 2022 10
1.2 Release calendar of Brazilian price indicators in December 2022 11
1.3 CUMSFE: shrinkage methods versus the SPF benchmark 25
1.4 Variable relevance via coefficient estimates using LASSO 27
1.5 Absolute RMSE: alternative versus the baseline specification 31
1.6 CUMSFE: tree-based methods versus the SPF benchmark 35
1.7 Fluctuation test: ML competing models versus the SPF benchmark 36
1.8 Heatmap of coefficient estimates using LASSO 37
1.9 Heatmap of coefficient estimates using LASSO on the SPF nowcasting errors 38

2.1 German HICP inflation and high-frequency scanner counterparts 51
2.2 Timeline of survey-based market expectations and German HICP data releases 59
2.3 Predictive gains of GFK:FMCG in relation to the fit with official counterparts 63
2.4 Contribution of HICP components to squared headline forecast errors . . . 67
2.5 Cumulative sum of the squared forecast error differentials: bottom-up U-MIDAS and direct ML versus Bloomberg survey-based expectations 69
2.6 German HICP inflation and high-frequency counterparts: energy and package holidays 76
2.7 Inflation series of selected COICOP-10 items and weekly scanner-based counterparts 77
2.8 Absolute forecast errors over time: tracking the nowcasting performance of GFK:FMCG product groups 83
2.9 Cumulative sum of the squared forecast error differentials: models versus Consensus market expectations 84
2.10 RMSE change of implementing the best compiled COICOP-10 price indices 84
2.11 RMSE change of a moving average smoother as a function of the volatility level 85
2.12 RMSE change of additional high-frequency distributed lags 85

3.1 Daily US Treasury yields at constant maturities 87
3.2 Daily US term structure of interest rates at constant maturities 97

3.3 Performance of optimal bond portfolios under alternative investment scenarios . . . 108
3.4 Sample mean and standard deviation of US Treasury yields . . . 113
3.5 Posterior estimates of factor volatilities in DNS-WSV . . . 120
3.6 Posterior estimates of TVP parameters in DNS-WSV . . . 120
3.7 Accumulated log-predictive density scores . . . 121
3.8 Yields predictability in response to crisis shocks . . . 122
3.9 Sample statistics and autocorrelation test of predictive Pearson residuals . 123
3.10 Cumulative compounded excess returns under alternative investment scenarios . . . 124

General Introduction

In the face of significant economic and financial uncertainties marked by events like the Great Recession and the COVID-19 pandemic, the demand for reliable and timely forecasting tools has become increasingly urgent. In particular, the inflationary shock induced by the COVID-19 crisis and Russia's invasion of Ukraine highlighted the challenges posed to anticipating in real-time the current state of the inflation process, which can quickly escalate under such turbulent conditions. After more than two decades of low and stable inflation in advanced economies, the immense inflationary wave that unfolded in 2021 caught many economists, practitioners, and central banks off guard.

Simultaneously, the pronounced volatility spikes and structural instabilities observed in bond markets during the Great Recession of 2008 and the COVID-19 crisis have underscored the need for more plausible density forecasts of the yield curve. The challenge of mitigating forecasting uncertainty in times of financial distress became evident and brought to the forefront its importance for effective interest rate risk management and strategic bond portfolio allocation.

This thesis is composed of three independent articles, each contributing distinctively towards achieving two primary objectives. The first two articles aim to enhance the precision of inflation nowcasts in real-time by harnessing the power of machine learning (ML) methods and capitalizing on the granular and high-frequency aspects of scanner price data. The third article aims to refine density forecasts of the yield curve within volatile bond markets, with a focus on portfolio allocation and risk management. Hence, the essays are unified by their reliance on state-of-the-art econometric and ML methods better capable of adapting to swift economic changes.

Chapter 1, titled "**Harnessing Machine Learning for Real-Time Inflation Nowcasting**", introduces a mixed-frequency framework designed for using ML methods to elevate the accuracy of inflation nowcasts. Additionally, we construct a novel real-time database of macro-financial indicators from the Brazilian economy – characterized by its persistently high inflation rates – to evaluate the effectiveness of machine learning in a real-time nowcasting experiment. Our study lays down clear directives for constructing weekly inflation nowcasts, emphasizing the importance of variable selection and the integration of accurate timely signals from price indicators and informed judgment entailed in survey-based market expectations.

We show that a well-designed mixed-frequency approach, particularly when coupled with linear shrinkage methods like the LASSO, delivers major nowcasting gains during the onset of the COVID-19 crisis where professional forecasters underestimated the rapidly evolving inflationary environment. Moreover, we demonstrate that predictive accuracy substantially increases when the model specification is free of ragged edges and guided

by the real-time data release of price indicators. We also provide evidence that the most recent high-frequency signal alone suffices for reliably updating inflation nowcasts.

This paper is joint work with Aishameriane Schmidt and Guilherme Valle Moura. I extensively contributed to all fronts of this article, starting with the conception of our research idea and constructing the real-time database from the Brazilian macroeconomy. From there, I played a major part in designing the mixed-frequency ML framework and computationally implementing the shrinkage-based models. I also intensively contributed to the co-writing of the paper. A current version is published as a De Nederlandsche Bank Working Paper (see Schnorrenberger et al., 2024).

Chapter 2, titled "**Nowcasting Consumer Price Inflation Using High-Frequency Scanner Data: Evidence from Germany**", shows how scanner price data combined ML methods can substantially enhance the nowcasting of consumer price inflation in Germany. Our study exploits the granular nature of household scanner data to address different hierarchy levels of inflation, from individual products to headline inflation. Hence, by exploiting the virtue of granular scanner data, we provide a deeper understanding of the disaggregate dynamics underlying overall inflation, whilst furnishing policymakers with valuable real-time information about policy-relevant price developments.

Specifically, at the item-level, such as butter and coffee beans, we construct a large set of weekly scanner-based price indices that closely match their official counterparts and proved valuable for anticipating current inflation dynamics as soon as the first week of a month. For nowcasting policy-relevant product groups such as (un-)processed food, we resort to the machine learning toolkit to exploit the large set of scanner-based price indices. In the final step, we target headline inflation by also adding high-frequency information on energy and travel services. Moreover, we use a bottom-up approach based on the economic structure inherent in the construction of official price indices, which allows us to better organize and condense the information carried by granular scanner data. This approach yields highly competitive nowcasts that outperform not only a competitive time series benchmark model but also survey-based market expectations.

This paper is joint work with Günter W. Beck, Kai Carstensen, Jan-Oliver Menz and Elisabeth Wieland. I contributed significantly to the implementation and estimation of the nowcasting models as well as to the co-writing of the paper. A current version is published as a Deutsche Bundesbank Discussion Paper (see Beck et al., 2023).

Chapter 3 is a single-authored paper entitled "**Bond Portfolio Optimization in Turbulent Times: a Dynamic Nelson-Siegel Approach with Wishart Stochastic Volatility**" and shifts the focus to yield curve modeling in periods of financial distress. Notably, it delves into the application of yield curve models augmented with Wishart stochastic volatility (WSV) for bond portfolio analysis. This essay highlights the importance of Wishart modeling strategies in mitigating forecasting uncertainty within volatile

bond markets, as evidenced by a forecasting experiment utilizing US Treasury yields data during recent financial crises. Furthermore, it evaluates the economic value of yield curve density forecasts derived from WSV structures in the context of portfolio allocation and Value-at-Risk (VaR) forecasting. The findings underscore the economic merits of Wishart strategies in terms of portfolio performance due to higher excess returns and Sharpe ratios compared to constant volatility models, especially within aggressive investment scenarios. Finally, WSV strategies fare best overall in terms of VaR forecasting, leading to better estimates for interest rate risk.

Overall, this dissertation contributes to the broader field of economic forecasting by pushing the boundaries of how novel methods from the forecaster's toolkit and non-traditional high-frequency datasets can be leveraged to construct more reliable and timely projections. It offers fresh insights into the efficacy of these methodologies in real-time forecasting, particularly during periods of turmoil such as those experienced in the aftermath of the pandemic. Ultimately, the research questions hereby investigated support better-informed decisions by practitioners and enable a timelier and more targeted response by policymakers. Potential avenues stemming from my research are the economic implications of improved inflation and yield curve forecasts on monetary policy decisions, thereby extending the thesis's contributions to a more theoretical aspect of economic forecasting.

Chapter 1

Harnessing Machine Learning for Real-Time Inflation Nowcasting

Abstract

We investigate the predictive ability of machine learning methods to produce weekly inflation nowcasts using high-frequency macro-financial indicators and a survey of professional forecasters. Within an unrestricted mixed-frequency ML framework, we provide clear guidelines to improve inflation nowcasts upon forecasts made by specialists. First, we find that variable selection performed via the LASSO is fundamental for crafting an effective ML model for inflation nowcasting. Second, we underscore the relevance of timely data on price indicators and SPF expectations to better discipline our model-based nowcasts, especially during the inflationary surge following the COVID-19 crisis. Third, we show that predictive accuracy substantially increases when the model specification is free of ragged edges and guided by the real-time data release of price indicators. Finally, incorporating the most recent high-frequency signal is already sufficient for real-time updates of the nowcast, eliminating the need to account for lagged high-frequency information.

Keywords: inflation nowcasting, machine learning, mixed-frequency data, survey of professional forecasters.

JEL classification: E31, E37, C53, C55.

This study is coauthored by Aishameriane Schmidt and Guilherme Valle Moura.
It is published as the DNB Working Paper No 806.

1.1 Introduction

The inflationary shock that reverberated through global markets following the COVID-19 pandemic highlighted the importance of accurate and timely inflation nowcasts for better-informed monetary policy, business pricing strategies, and portfolio allocation decisions. While official price statistics are only measured at the monthly frequency and released with a significant delay, high-frequency (e.g., weekly or daily) and quickly released data have become particularly useful for anticipating the current state of the inflation process.[1] The relevance of updating inflation nowcasts in a timely fashion extends beyond disruptive environments, such as those witnessed in the aftermath of the pandemic, offering a means to anticipate swift inflationary shocks, as well as inflationary trends that may escalate or dwindle. Moreover, short- and medium-term inflation forecasts highly benefit from taking high-quality nowcasts as a jumping-off point (see, e.g., Faust and Wright, 2013; Krüger et al., 2017).

Machine learning (ML) methods have recently enjoyed great popularity in inflation forecasting under a data-rich environment (Garcia et al., 2017; Medeiros et al., 2021; Joseph et al., 2021; Hauzenberger et al., 2023; Araujo and Gaglianone, 2023; Barkan et al., 2022), exhibiting substantial improvements upon well-established benchmarks (e.g., Atkeson and Ohanian, 2001; Stock and Watson, 2007). However, there remains insufficient guidance on key modeling choices when using ML methods to construct inflation nowcasts in a real-time setup, especially during high inflation periods. The pandemic, in particular, posed challenges to nowcasting frameworks that struggle to anticipate rapidly evolving inflation dynamics not often seen in past data. Furthermore, in a nowcasting setting, the dimensionality challenge is amplified by the presence of high-frequency lags from numerous predictors, which may easily lead to overfitting.

This paper provides clear guidance for inflation nowcasting by evaluating a battery of easy-to-implement ML methods within a mixed-data sampling (MIDAS) approach. We contribute to the nowcasting literature by thoroughly investigating key modeling practices in an environment characterized by persistently high inflation, namely the Brazilian economy of the past decades. Moreover, we assess the predictive value of selected macro-financial predictors for inflation, including informed judgment entailed in a timely survey of professional forecasters (SPF). Specifically, we show that a well-designed unrestricted MIDAS (U-MIDAS) approach (Foroni et al., 2015) combined with linear shrinkage methods, especially the LASSO, produce inflation nowcasts that significantly improve upon SPF expectations. These predictive gains are particularly large at the onset of the COVID-19 crisis, whereas meaningful off-model information from SPF helps to discipline

[1] Giannone et al. (2008) and Bańbura et al. (2013), e.g., provide a comprehensive review of how the rapidly increasing availability of high-frequency data proves invaluable in obtaining early estimates of the current economic landscape while official statistics on key macroeconomic variables are yet to be released.

our model-based nowcasts. The unrestricted mixed-frequency ML structure also facilitates model interpretation and allows us to exploit potential nonlinear dynamics in the data, hereby assessed via tree-based methods.

In the broader field of macroeconomic nowcasting, research has typically focused on the holy grail of constructing high-frequency estimates of GDP, influenced by the success of Giannone et al. (2008). Dynamic factor models and mixed-frequency Bayesian VARs have emerged as popular tools amongst practitioners and policymakers (see, among others, Schorfheide and Song, 2015; McCracken et al., 2015; Carriero et al., 2015; Hindrayanto et al., 2016; Dahlhaus et al., 2017; Cimadomo et al., 2022; Cascaldi-Garcia et al., 2023; Huber et al., 2023). An early use of these econometric frameworks to exploit high-frequency data for inflation nowcasting is presented in Modugno (2013) and Knotek and Zaman (2017). Large-scale factor models, however, are not designed to capture fast-moving inflation dynamics at very short horizons and suffer from the ragged-edge problem that considerably worsens the forecasting properties of the model (see Marcellino and Schumacher, 2010). In addition, Knotek and Zaman (2017) show that inflation nowcasting may benefit from choosing a small number of highly informative predictors in contrast to extracting common factors from a large dataset.

Andreou et al. (2013), Monteforte and Moretti (2013), Breitung and Roling (2015) and Knotek II and Zaman (2023) consider MIDAS regressions with leads to eliminate ragged edges and effectively exploit more daily information of financial markets that are highly correlated with short-term inflation expectations. Although MIDAS regressions gained popularity for their parsimonious treatment of high-frequency lags and successful out-of-sample performance, they struggle with the dimensionality issue posed by numerous high-frequency predictors, which may easily lead to overparameterization.

Boosted by the COVID-19 crisis and the big data boom in economics, this line of research has taken up but with an increased focus on ML methods to guard against overfitting in high-dimensional settings and improve nowcasting accuracy over traditional econometric frameworks.[2] Penalized MIDAS regressions evolved as a suitable strategy for performing variable selection in macroeconomic nowcasting (Marsilli, 2014; Siliverstovs, 2017; Uematsu and Tanaka, 2019; Mogliani and Simoni, 2021; Babii et al., 2021; Kohns and Potjagailo, 2023; Beck et al., 2023; Aliaj et al., 2023). Specifically, Borup et al. (2023) demonstrate that exploiting more recent daily Google Trends data via their proposed combination of the U-MIDAS approach with ML methods can substantially improve predictions of weekly initial claims while securing model interpretation in the era of big data. Besides, this mixed-frequency ML structure accommodates nonlinear ML-based predictive relationships, such as those analyzed in Richardson et al. (2021), Clark et al.

[2]Non-traditional high-frequency data such as web scraping, Google Search, and scanner data have also become viable sources to nowcast both headline inflation and disaggregated components, such as food prices (see Harchaoui and Janssen, 2018; Powell et al., 2018; Macias et al., 2023; Beck et al., 2023, to name only a few).

(2022) and Barbaglia et al. (2023).

Building on these trends, we develop guidelines for key modeling choices for producing accurate weekly nowcasts of inflation using a large set of macro-financial data within the mixed-frequency ML structure. Therefore, from a methodology standpoint, providing clear directives for practitioners and policymakers in this domain is the key objective of this paper. In addition, building upon the success of random forest models in forecasting US inflation (Medeiros et al., 2021) and their capability to address temporal nonlinearities when forecasting UK inflation (Joseph et al., 2021), we complement the existing ML applications in macroeconomic nowcasting. Specifically, we evaluate the effectiveness of tree-based methods for inflation nowcasts.

From a practical standpoint, we conduct a real-time empirical exercise based on Brazilian data, which encompasses recent decades marked by persistently high inflation rates. This sets us apart from the predominant focus on US or euro area inflation by existing literature. Notably, the presence of multiple episodes of rising inflation in recent Brazilian history allows us to gain insights that may be extrapolated to advanced economies undergoing unprecedented inflationary shocks not present in past data. To this end, we construct a novel real-time database from the Brazilian macroeconomy, which also features a variety of alternative high-frequency price indicators that are timely released by private agencies and closely monitored by professional forecasters. Furthermore, survey-based expectations have proven valuable to improve model-based nowcasts, particularly during periods of rising inflation (see, e.g., Banbura et al., 2021; Bańbura et al., 2023; Bobeica and Hartwig, 2023). In this sense, we integrate the daily SPF conducted by the Brazilian Central Bank (BCB).

Our empirical exercise produces weekly nowcasts for the monthly developments of the official headline CPI targeted by BCB's monetary policy decisions, which is released with an average delay of seven business days after the reporting month. We select 20 predictors to compose our real-time dataset. The predictors are either available at a higher frequency – and transformed into weekly time series containing the latest month-on-month signal – or sampled monthly but released throughout the reporting month. For model interpretation, we divide them into four categories: monthly price indicators, weekly price indicators, daily financial variables, and daily SPF expectations. To exploit the information in our real-time set of predictors while guarding against overfitting, we compare linear prediction models via shrinkage (the LASSO, Ridge, Elastic Net, and sparse-group LASSO) against nonlinear tree-based methods (Random Forest, Local Linear Forest, and Bayesian Additive Regression Trees).

Our findings underscore the effectiveness of shrinkage models to nowcasting inflation dynamics, with LASSO consistently surpassing tree-based methods in terms of RMSE. This is consistent with previous results for forecasting Brazilian inflation (see, e.g., Medeiros

et al., 2016; Garcia et al., 2017), which indicate that variable selection done via the LASSO outperforms at the very short horizon. Notably, LASSO predictions exhibit exceptional accuracy at longer nowcast horizons compared to SPF expectations. This reflects the tendency of professional forecasters to adjust their expectations more frequently as the information set expands within the reporting month. Additionally, we observe large nowcasting gains building up during the COVID-19 inflation surge, where professional forecasters underestimated the rapidly evolving inflationary environment.

Moreover, our analysis reveals a notable difference in the variables selected by the LASSO depending on the nowcast horizon. Specifically, at longer nowcast horizons, the selection tends to produce a relatively sparse structure with SPF expectations and weekly price indicators as the primary predictors. Conversely, as we approach shorter horizons, a denser model structure emerges, driven by the pronounced relevance of monthly price indicators. This shift reflects the increased availability of accurate contemporaneous inflation signals as the reporting month unfolds. Consequently, data releases on monthly price indicators diminish the relative importance of SPF expectations, although informed judgment remains highly influential, particularly in navigating the challenges posed by the COVID-19 crisis. Overall, while financial variables play a minor role, the combination of timely price indicators with SPF judgments proves critical in producing weekly inflation nowcasts.

Finally, a deeper investigation of key modeling choices within our mixed-frequency ML framework reveals the considerable impact of (i) accounting for SPF data in the predictor set, (ii) eliminating ragged edges, (iii) guiding model specifications by real-time data releases, and to a lesser extent, (iv) focusing solely on the most recent high-frequency signal. A baseline prediction model featuring these key elements yields predictive gains up to 60%. Notably, shrinkage-based predictions can highly benefit from using meaningful judgment in survey data and addressing the ragged-edge problem.

The paper proceeds as follows. Section 1.2 describes the real-time dataset of the Brazilian macroeconomy and how these macro-financial variables relate to the target variable. Section 1.3 outlines the nowcasting setup and provides an overview of the mixed-frequency ML strategies. Next, we present our empirical results in Section 1.4. This section also provides an interpretation of the best-performing fitted model and offers guidance on key modeling choices for constructing accurate weekly nowcasts using the real-time data flow. Finally, Section 1.5 concludes.

1.2 Data

To compute weekly nowcasts of inflation figures we select predictors that have two features: significant correlation with price developments and earlier availability in comparison to official inflation releases. We put together a novel real-time database of macro-financial series from the Brazilian economy tailored for inflation nowcasting. In this context, our dataset mainly consists of timely price indicators, financial variables, and experts' forecasts that carry predictive content about the current month's inflation rate.[3] Besides data on the target CPI variable, we organized publicly available information on price indicators released both by public and private institutions, financial indicators, and daily SPF with aggregate predictions for the target variable.[4] Our real-time dataset covers the period from June 2004 up to December 2022 ($T = 222$ monthly observations), whereas information on release dates is available from January 2013 onwards.

The official inflation measure in Brazil is known as the Broad National Consumer Price Index (IPCA), and concurrently, it serves as the reference for the inflation-targeting system in Brazil.[5] The IPCA is designed to reflect consumption patterns of urban households in major Brazilian cities that earn from 1 to 40 minimum wages (90% of urban population). The Brazilian statistical office publishes IPCA figures with an average lag of seven workdays after the end of the reporting month.

Figure 1.1 shows the IPCA evolution since mid-2001, shortly after the BCB adopted the inflation targeting regime. The year 2003 witnessed an escalation in political and economic risks following the election of the Workers' Party representative, triggering a foreign capital outflow that led to a strong exchange rate depreciation and domestic inflationary pressure. This was followed by a relatively calm period, marked by annual IPCA fluctuations around 5%. However, a return to double-digit inflation figures occurred during the political turmoil that started in 2013, leading to the impeachment of President Rousseff in early 2015. Following years of price stability with IPCA oscillating close to BCB's target, inflation surged again in the aftermath of the pandemic shock, similar to trends observed worldwide.

We use a total of 20 predictors in our empirical application, excluding the lags of IPCA.[6] These predictors can be divided into four categories: monthly price indicators,

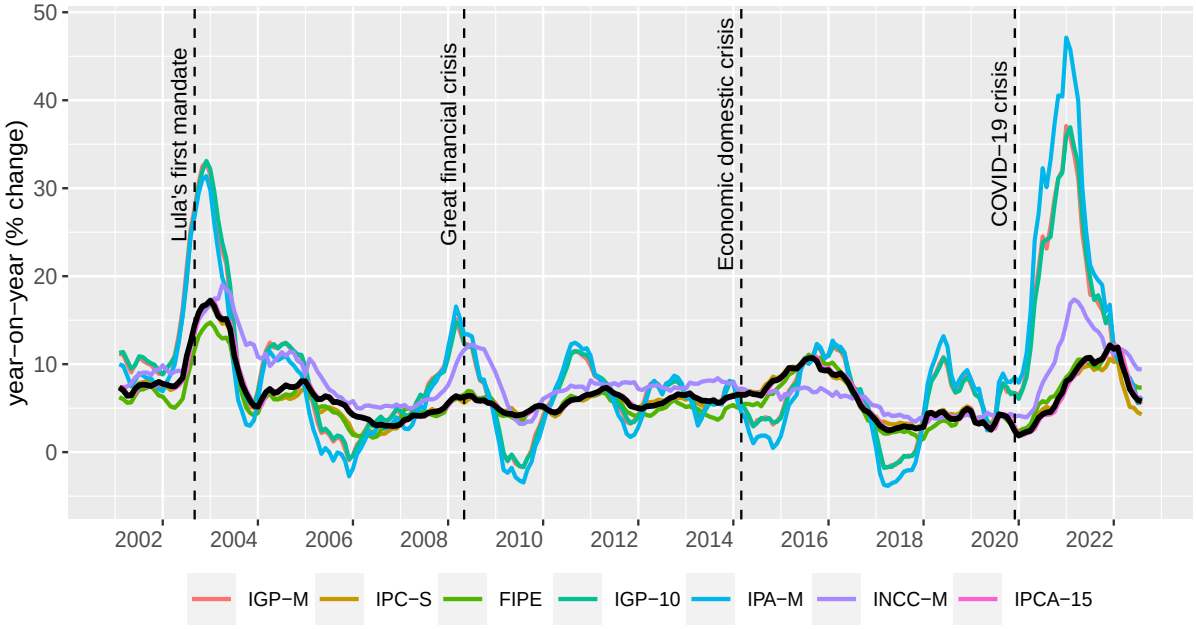
[3] We disregard monthly indicators of real economic activity for two reasons: (i) short or no availability before official releases of the target inflation and (ii) non-significant cross-correlations up to six lags with the target month-on-month inflation rate. Hence, economic activity variables do not fit our nowcasting purpose.

[4] Although our analysis focuses on price indicators and financial variables as the potential predictors for inflation, due to their high-frequency and timely attributes, the real-time database also comprises vintages and revisions of hard and survey-based data for economic activity (e.g., industrial production, unemployment rate, net payroll jobs, PMI manufacturing, retail and services indices, consumer and business confidence indicators, among others).

[5] Besides, a sizeable number of inflation-linked government bonds use the IPCA as their reference.

[6] While more indicators could potentially correlate with IPCA, we have chosen a medium-sized dataset. This decision aligns with previous findings in the literature (see, e.g., Carriero et al., 2019b, 2020), who show that, for point and density forecasting/nowcasting of GDP growth and inflation, a wider array of predictors do not outperform models with only a few hand-picked predictors. Nonetheless, the potential high-dimensionality issue arising in our application is also connected to the choice of high-frequency lags included in the nowcasting model (see Section 1.3.1).

Figure 1.1: Time series of Brazilian price indicators, 2001 – 2022



Notes: The official Brazilian CPI (IPCA) is depicted in black while alternative price indicators are illustrated in colored lines. The full description of each indicator is available in Table 1.1.

weekly price indicators, daily financial variables, and daily expectations of professional forecasters. The data and publication dates are obtained from many sources, including the Brazilian Institute of Geography and Statistics (IBGE), BCB, Brazil Stock Exchange (B3), Getulio Vargas Foundation (FGV), Institute of Economic Research Foundation (Fipe), Brazilian National Agency of Petroleum, Natural Gas and Biofuels (ANP) and Bloomberg. Table 1.1 presents a summary of the selected predictors for IPCA dynamics, including the sampling period and publication lags.

The first group of predictors consists of five monthly price indicators primarily collected in urban areas of major Brazilian cities. These indices are sampled at the monthly frequency but released before the end of the reporting month and essentially differ in terms of the sampling period and targeted prices. For instance, IPCA-15 mimics IPCA itself in terms of methodology, but it reflects prices collected from the 16^{th} of the preceding month to the 15^{th} of the reporting month. Releases for this mid-month version of the IPCA become available with an average delay of 8 days (usually at the beginning of the 4^{th} week) and thus allow for early signals of IPCA dynamics. Additionally, we include a producer price index (PPI) termed IPA-M, which monitors inter-business transaction prices of agricultural and industrial products, and a construction cost index named INCC-M. The remaining two indices, IGP-M and IGP-10, are the weighted average of the IPA-M (60%), IPC-S (30%, FGV's weekly CPI measure presented below), and the INCC-M (10%), diverging only by their sampling periods.

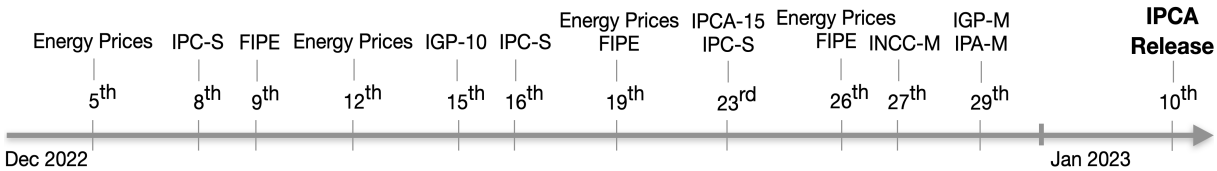
The above five monthly indices are also displayed in Figure 1.1. While closely correlated

with IPCA, some exhibit greater volatility, especially in turbulent times. For example, IPA-M – and consequently, IGP-M and IGP-10 – are markedly affected by the large volatility in exchange rates observed during the initial year of Lula's administration, the Great Financial Crisis, and the pandemic. The INCC-M, related to construction costs, is in general higher than the IPCA, but presents a lower amplitude than the majority of the other indexes.

The second group of predictors contains six timely indicators of consumer and energy prices sampled at the weekly frequency and published with a lag of one or two days after the closing of a given week. The IPC-S and FIPE intend to closely mirror the IPCA at a higher frequency – as shown in Figure 1.1 – but respectively accounting for consumption baskets of earnings in the range of 1-33 and 1-10 minimum wages.[7] Moreover, we include prices of major energy components: diesel, gasoline, ethanol fuel, and liquefied natural gas. These prices are collected by surveys of the wholesale fuel price practiced by retailers of around 500 cities nationwide.[8]

Figure 1.2 illustrates the timeline of real-time data releases of the above price indicators in December 2022. As shown, IPCA figures came out on the 10th of January 2023, but data releases of the selected predictors mostly occur throughout the reporting month. For example, given that energy prices become available after the closing of a calendar week, the first release is on 5 December while the subsequent numbers are provided on the following Mondays. IPC-S and FIPE become available shortly after the closing of a four-week collection system ending on four set dates (07, 15, 22 and end-of-month).[9] Hereby these numbers are first released on the 8th and 9th, followed by releases on the 16th and 19th, and so on, which is extremely quick for international standards. Turning to monthly indicators, data on IGP-10 and IPCA-15 come out relatively early in the month – around the third week – whereas INCC-M, IGP-M and IPA-M follow next before the month ends.

Figure 1.2: Release calendar of Brazilian price indicators in December 2022



The third group of predictors contains daily information from financial markets, including movements in the yield curve or interest rate spreads, commodity and stock price

[7] The sampling procedure of FIPE only accounts for households living in São Paulo city.

[8] Compared to information on raw oil prices available in financial markets, these surveys have the advantage that distribution and retail margins are fully accounted for.

[9] This means that the computation of these indices considers the average of prices collected during the four weeks preceding the closing date.

indices, and exchange rates.[10] The choice of these financial market variables is motivated by their timely information about short-term inflation expectations and findings in the literature on inflation forecasting. For example, Modugno (2013), Monteforte and Moretti (2013), and Breitung and Roling (2015) show that relevant commodities (e.g., crude oil prices) and financial assets are among the most reliable indicators of inflation changes. Furthermore, central banks and practitioners monitor daily financial variables to forecast the state of the macroeconomy (Andreou et al., 2013).

Finally, we use expert information from the SPF conducted by the BCB, also known as the FOCUS survey. It started in the late 90s, together with the implementation of the inflation-targeting regime in Brazil. Participation in the survey is limited to banks, asset managers, companies linked to real economic sectors, brokers, and consultancies, who have to be pre-screened by the BCB. These institutions can continuously provide their short and long-run expectations regarding key macroeconomic indicators such as GDP, inflation, and exchange rate, among others. The BCB releases daily aggregate statistics of the SPF, with a delay of one business day, as well as a Top 5 ranking with the best-performing forecasting institutions divided across indicators and forecast horizons.

Historically, there are over 100 active participants in the SPF survey. The median of these experts' forecasts for IPCA dynamics is closely monitored by market participants, especially via the weekly handout report released by the BCB every Monday morning with data up to the previous Friday (Marques, 2012). We use the median of SPF expectations as both a predictor in our models as well as a benchmark to compare our nowcasts. As an additional benchmark, we compare our predictions against the median forecast produced by the Top 5 forecasters. The BCB ranks the Top 5 participant institutions based on previous months' performance. Hence, after obtaining the best five institutions, for each indicator and horizon, the BCB averages their forecasts for the current month[11].

[10] The stock index (IBOV) corresponds to the B3 Index, while interest rates are derived from Brazilian interbank deposit future contracts negotiated at B3, ultimately linked to treasury bills issued by the BCB.

[11] Note that there is no "data leakage" given that the ranking is computed based on the past. For instance, it might be that the current Top 5 institutions are not the ones that will produce the best forecasts at the current period.

Table 1.1: Real-time database of Brazilian macro-financial indicators

Series	Mnemonic	Reference period	Publication timing	Avg. delay	Starting date	Source
Target inflation variable						
Broad national CPI	IPCA	full month t	2nd week, following month	7	2003M1	IBGE
Monthly price indicators						
IPCA - extended	IPCA-15	16^{th}_{t-1} to 15^{th}_{t}	3rd/4th week, reporting month	8	2003M1	IBGE
General market price index	IGP-M	21^{st}_{t-1} to 20^{th}_{t}	last week, reporting month	7	2003M1	FGV
General price index - 10	IGP-10	11^{th}_{t-1} to 10^{th}_{t}	2nd/3rd week, reporting month	4	2003M1	FGV
Wholesale market PPI	IPA-M	21^{st}_{t-1} to 20^{th}_{t}	last week, reporting month	7	2003M1	FGV
National construction cost	INCC-M	21^{st}_{t-1} to 20^{th}_{t}	last week, reporting month	5	2003M1	FGV
Weekly price indicators						
FGV's CPI	IPC-S	four-week	1st day, following week	1	2003M2	FGV
Fipe's CPI	FIPE	four-week	2nd day, following week	2	2003M1	Fipe
Diesel prices	DIESEL	full week	1st day, following week	1	2004M5W2	ANP
Gasoline prices	GAS	full week	1st day, following week	1	2004M5W2	ANP
Ethanol fuel prices	ETOH	full week	1st day, following week	1	2004M5W2	ANP
Liquefied natural gas prices	LNG	full week	1st day, following week	1	2004M5W2	ANP
Daily financial variables						
Short-term interest rates	SELIC	end of day	real-time	0	2003M1	BCB
Brazilian Real/U$$ forex	FOREX	end of day	real-time	0	2003M1	BCB
Bovespa stock price index	IBOV	end of day	real-time	0	2003M1	B3
Electric utilities index	IEE	end of day	real-time	0	2003M1	B3
DI-rates (10Y maturity)*	DI10	end of day	real-time	0	2004M1	B3
DI-spread (10Y minus 3M)*	SPREAD	end of day	real-time	0	2004M1	B3
Bloomberg commodity index	BCOM	end of day	real-time	0	2003M1	Bloomberg
Daily expectations from the FOCUS survey of professional forecasters						
IPCA nowcasts (median)	SPF	full day	subsequent day	1	2003M1	BCB

Notes: This table reports the full list of time series selected for the nowcasting exercise. The reference period relates to the data collection period. The publication timing provides the regular release calendar for the reference period while the average delay stands for the publishing lags (in business days). The variables are not seasonally adjusted and transformed into month-on-month (MoM) % change to guarantee stationarity of the time series; the only exceptions are the interest rates series (SELIC, DI10 and SPREAD) which are transformed into monthly changes. MoM transformations for high-frequency variables consider the same reference week or day from the preceding month. *DI-rates are yields of Brazilian interbank deposit future contracts negotiated at B3.

1.3 Methodology

Our nowcasting model follows an unrestricted mixed-frequency structure combined with ML methods that guard against overfitting in a high-dimensional setting. Our methodology is essentially divided into two components: (i) the general nowcasting setup, describing the functional form of how the mixed-frequency dataset will be organized and specifying the information used at each nowcast date, and (ii) the classes of ML methods employed to produce the nowcasts. This mixed-frequency ML structure enables us to treat separately the real-time flow of information from each predictor, thereby facilitating model interpretation, while improving nowcasting accuracy by harnessing the power of ML methods.

1.3.1 Nowcasting setup

To fix ideas, we aim to nowcast monthly inflation rates with predictors sampled at the daily, weekly, and monthly frequencies. Let $\pi_t = 100(P_t/P_{t-1} - 1)$ denote the month-on-month inflation rate, where P_t is the price level in month t. A generic high-frequency (daily or weekly) macro-financial variable is given by $x_t^{(w)}$ and can be sampled w times more frequently than the target π_t. Moreover, x_t represents a generic monthly price indicator, with the sampling process extending over t but disclosed before π_t. In this sense, time indices $t = 1, \ldots, T$ act as the common frequency between π_t and predictors $x_t^{(w)}$ and x_t.

Suppose we would like to update our nowcasts at the weekly frequency. Specifically, at four different points within the month: days 8, 15, 22, and end-of-month.[12] Given the mixed-frequency environment, we take a stance on how to incorporate high-frequency information on these four nowcast days. We start by assuming a fixed monthly-to-weekly combination, with a frequency ratio of $w = 4$, to accommodate weekly updates of the nowcast.[13] Hence, at the end of month t, the information set also includes the following K-dimensional vectors of high-frequency predictors: $\boldsymbol{x}_t^{(w)}, \boldsymbol{x}_{t-\frac{1}{w}}^{(w)}, \ldots, \boldsymbol{x}_{t-\frac{w-1}{w}}^{(w)}$, where $t - j/w$ denotes the j^{th} past high-frequency period for $j = 0, \ldots, w-1$. More precisely, t corresponds to end-of-month observations; $t - 1/4$ is the next to end-of-month, and thus day 22; $t - 2/4$, day 15; and $t - 3/4$, day 8. As a result, the forecast horizon h respectively becomes j/w.

Next, we must address the frequency mismatch between daily and weekly data, along with the non-synchronous nature of macroeconomic data releases. We transform the daily information from financial predictors and the SPF data into weekly time series containing

[12] This particular choice of days allows us to control for the problem of overlapping calendar weeks across consecutive months and the heterogeneous number of days in different months.

[13] The choice for weekly updates of the nowcast with a fixed monthly/weekly mixture also avoids a higher proliferation of parameters arising from a higher frequency mismatch in a model that would combine monthly and daily variables.

the latest month-on-month rates available on the nowcast day.[14] Data on our weekly predictors follow different sampling strategies but become available with a minimal lag of one or two days. Hence, by day 8 of the reporting month t, we assume immediate access to the first week's contemporaneous data. For example, IPC-S data covering the first week of t is reliably published on the first day following the closure of that week – typically on the 8th or the 9th/10th if the closing date is a Friday/Saturday (see Figure 1.2). Consequently, this data regularly integrates our information set in $t - 3/4$. For the remaining nowcast days within t, we shift forward the latest released contemporaneous month-on-month signal if the corresponding weekly data is not yet published.

The general prediction model for the nowcast horizon $h = j/w$ is given by

$$\pi_{t|t-h} = f^h\left(\pi_{t-1}, \boldsymbol{x}_t^h, \boldsymbol{d}_t, \boldsymbol{x}_{t-h}^{(w)}, \ldots, \boldsymbol{x}_{t-h-p/w}^{(w)}; \boldsymbol{\theta}^h\right) + \varepsilon_t^h, \tag{1.1}$$

whereas the autoregressive term accounts for temporal dependence in π_t[15]; $\boldsymbol{x}_t^h$ is a horizon-specific J^h-dimensional vector of monthly predictors, thereby sampled at the same frequency as π_t; the set of 11 monthly dummy variables $\boldsymbol{d}_t$ capture potential seasonal patterns in price dynamics; and high-frequency predictors with data up to the nowcast date $t - h$ and corresponding lags are respectively denoted by $\boldsymbol{x}_{t-h}^{(w)}, \ldots, \boldsymbol{x}_{t-h-p/w}^{(w)}$. In addition, $\boldsymbol{\theta}^h$ is a vector of model parameters specific to the prediction function f^h at horizon h; and ε_t^h is a zero-mean disturbance term.

In the general form, model (1.1) includes $p \geq 0$ relevant high-frequency lags to construct the nowcast at any horizon h. For instance, assuming that we stand at day 8 of month t, the nowcast horizon is $h = 3/4$ and we might use the high-frequency lags $\boldsymbol{x}_{t-3/4}^{(w)}, \boldsymbol{x}_{t-1}^{(w)}, \boldsymbol{x}_{t-5/4}^{(w)}, \ldots, \boldsymbol{x}_{t-p/4}^{(w)}$. Likewise, if end-of-month observations are available, the nowcast horizon is $h = 0$ and the predictors $\boldsymbol{x}_t^{(w)}, \boldsymbol{x}_{t-1/4}^{(w)}, \ldots, \boldsymbol{x}_{t-p/4}^{(w)}$ might be included.[16] Hereby the baseline specification only incorporates the most recent month-on-month high-frequency signal by setting $p = 0$, although one might choose $p = w - 1$ to account for all contemporaneous high-frequency signals when nowcasting at the end-of-month (see Section 1.4.3), or even $p > w - 1$ to include lags that span over past and distant months.

Since our prediction model assigns individual coefficients to each of the high-frequency predictors in $\boldsymbol{x}_t^{(w)}$ and its associated lags, a linear specification of (1.1) can be seen as a U-MIDAS model. Foroni et al. (2015) argue that a fairly small frequency mismatch, such as our monthly-to-weekly mixture, favors the adoption of the U-MIDAS over restricted MIDAS regressions with tightly specified lag polynomials that perform nonlinear temporal aggregation of high-frequency lags (see also Ghysels and Marcellino, 2018). Consequently, the U-MIDAS approach allows for flexible estimation of the individual effects of high-

[14] The month-on-month transformations of daily and weekly data are taken by referencing the same day in the previous month. This also ensures the stationarity of the variables.

[15] If needed, additional lags of π_t can be included.

[16] See Appendix 1.A for an explicit representation of the high-frequency component of (1.1) in matrix form.

frequency lags on the target while facilitating the interpretability of the model.

It is worth emphasizing that $\boldsymbol{x}_t^h$ only incorporates monthly predictors with available contemporaneous information at the time of the nowcast, resulting in a horizon-specific dimension J^h. For example, given that data on IPCA-15 is usually published between the 19$^{\text{th}}$ and 23$^{\text{rd}}$ of the reporting month t, IPCA-15 will not be included in $\boldsymbol{x}_t^h$ when nowcasts are made on days 8 and 15. This implies that the predictor space will exhibit reduced dimensionality at longer nowcast horizons. Hence, we constantly assess the real-time data availability of monthly predictors by the time of the nowcast and adjust the general specification (1.1) accordingly. This approach mitigates the risk of generating imprecise nowcasts from assigning non-zero and relevant coefficients to monthly predictors lacking new information to construct the nowcast for π_t.[17] Furthermore, it avoids ragged edges in the modeling.

The flexibility of model (1.1), however, comes at the cost of overparameterization as the count K of high-frequency predictors and their lags p rise. Specifically, the model features $K(p+1)+J^h+13$ parameters for a specific horizon h, including the intercept. In macroeconomic contexts, the effective sample size might be relatively short compared to the number of parameters, posing challenges for conventional estimation methods and leading to high estimation uncertainty. To address this high-dimensional prediction problem, we implement the mixed-frequency ML strategy (see Borup et al., 2023) by incorporating a wide range of ML methods that allow for flexible estimation of the coefficients while still guarding against overfitting.

1.3.2 Machine learning methods

The mixed-frequency ML strategy can be applied to both linear and nonlinear prediction models. The ML methods we implement have been enjoying growing popularity within economics and are distinguished between two classes: linear shrinkage and nonlinear tree-based methods. In the first group, we have the Elastic Net (ENet) regression and its two special cases, LASSO and Ridge. As an alternative to these standard methods, we apply the sparse-group LASSO estimator with MIDAS structure, a novel approach introduced by (Babii et al., 2021). This method has the advantage of acknowledging the serial dependence across different high-frequency lags. Turning to tree-based methods, we implement the Random Forest (RF), Local Linear Forest (LLF) – both in its solo form and the ensemble prediction with a LASSO pre-selection of predictors – and the Bayesian Additive Regression Trees (BART). Table 1.2 provides an overview of these methods and the corresponding tuning parameters.

[17] One might also include lags of $\boldsymbol{x}_t^h$ in Eq. (1.1), but we use the autoregressive term to fully capture the potential serial correlation in π_t.

Table 1.2: Summary of the ML competing models

Model	Short name	Reference	R function (package)	Tuning parameters and cross-validation
Least absolute shrinkage and selection operator	LASSO	Tibshirani (1996)	glmnet (glmnet) and trainControl, train (caret)	λ using time series cross-validation
Ridge	Ridge	Hoerl and Kennard (1970)	glmnet (glmnet) and trainControl, train (caret)	λ using time series cross-validation
Elastic Net	ENet	Zou and Hastie (2005)	glmnet (glmnet) and trainControl, train (caret)	α, λ using time series cross-validation
Sparse-Group LASSO	sg-LASSO	Babii et al. (2021)	cv.sgl.fit (midasml)	α, λ using time series cross-validation
Random Forest	RF	Breiman (2001)	randomForest (randomForest)	number of skip-sampled predictors to split the tree (mtry) equal to the maximum between number of predictors divided by three and one
Local Linear Forest	LLF	Friedberg et al. (2020)	ll_regression_forest (grf)	We used default values for sample fraction (0.5), number of trees (2000), mtry (min{number of predictors$^{1/2}$ + 20, number of predictors}), minimum node size (5), honesty fraction (0.5), honest prune leaves (1), α (0.05), imbalance penalty (0)
Bayesian Additive Regression Trees	BART	Chipman et al. (2012)	rbart (rbart)	200 trees, 1000 posterior simulations after burn-in (100), d=0.95, probability of death = 0.7

To set the stage to formally outline the ML methods, we denote by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_t)'$ the target inflation series up to t. The low-frequency predictor set specific to horizon h is denoted by $\boldsymbol{x} = (\boldsymbol{x}_1^{h'}, \ldots, \boldsymbol{x}_t^{h'})$. The $t \times K(p+1)$ predictor set of high-frequency data is given by $\boldsymbol{x^{(w)}} = (\boldsymbol{x_{-h}^{(w)}}, \boldsymbol{x_{-h-1/4}^{(w)}}, \ldots, \boldsymbol{x_{-h-p/w}^{(w)}})$, whereas $\boldsymbol{x_{-h}^{(w)}} = (\boldsymbol{x}_{1-h}^{(w)}, \boldsymbol{x}_{2-h}^{(w)}, \ldots, \boldsymbol{x}_{t-h}^{(w)})'$ denotes the $t \times K$ high-frequency set associated with lag $t-h$. The general predictors matrix is then given by $\boldsymbol{X} = (\iota, \boldsymbol{\pi}_{-1}, \boldsymbol{x}, \boldsymbol{d}, \boldsymbol{x^{(w)}})$, where ι accounts for the intercept, $\boldsymbol{\pi}_{-1}$ is the first lag of $\boldsymbol{\pi}$ and $\boldsymbol{d}$ comprises the seasonal deterministic dummies. For convenience, we drop the superscript h from the vector of model parameters $\boldsymbol{\theta}$.

Shrinkage methods

Shrinkage methods are penalized regression schemes that identify the relevant predictors from a large dataset. This targeted selection aims to improve forecasting precision at the cost of a slight increase in bias. The ENet estimator, proposed by (Zou and Hastie, 2005), solves the penalized least-squares problem:

$$\hat{\boldsymbol{\theta}} = \min_{\hat{\boldsymbol{\theta}}} ||\boldsymbol{\pi} - \boldsymbol{X}\boldsymbol{\theta}||^2 + \lambda \left(\alpha \left|\boldsymbol{\theta}\right|_1 + \frac{(1-\alpha)}{2} ||\boldsymbol{\theta}||^2 \right), \tag{1.2}$$

where $\alpha \in (0, 1]$ is a weight hyperparameter that interpolates between LASSO ($\alpha = 1$) and Ridge regression (as $\alpha \to 0$). Hence, LASSO penalizes the sum of absolute coefficients via the shrinking penalty using the ℓ_1-norm while Ridge penalizes the sum of squared coefficients via the ℓ_2-norm. The regularization hyperparameter λ controls the amount of shrinkage in the parameter space $\boldsymbol{\theta}$. Hence, estimator (1.2) shrinks coefficients of irrelevant predictors toward zero. Because the penalty term of ENet and LASSO include the ℓ_1-norm, they can perform variable selection and thus yielding a sparse and parsimonious model that facilitates interpretation. In contrast, coefficients estimated via Ridge regression never equal zero, yielding a dense model.

Babii et al. (2021) argue that high-dimensional mixed-frequency representations with multiple high-frequency lags ($p > 0$ using our notation) involve certain data structures that once taken into account should lead to increased performance out-of-sample. These structures relate to groups covering the relevant lags of a single high-frequency predictor. In this sense, the sg-LASSO with MIDAS structure selects not only the relevant predictors for the target but also the appropriate lag structure of each high-frequency predictor. This structured sparsity constitutes the key feature of sg-LASSO and a refinement of the unstructured LASSO, which fails to acknowledge serial dependence across high-frequency lags and tends to arbitrarily select one lag from the group (see "irrepresentable condition" in Zhao and Yu, 2006).

The sg-LASSO solves the penalized regression problem:

$$\hat{\boldsymbol{\theta}} = \min_{\hat{\boldsymbol{\theta}}} ||\boldsymbol{\pi} - \boldsymbol{X}\boldsymbol{\theta}||^2 + 2\lambda \left(\alpha |\boldsymbol{\theta}|_1 + (1-\alpha) ||\boldsymbol{\theta}||_{2,1}\right), \tag{1.3}$$

where $||\boldsymbol{\theta}||_{2,1} = \sum_{G\in\mathcal{G}} ||\boldsymbol{\theta}_G||$ is the group LASSO norm for a group structure $\mathcal{G}$ that comprises the $p+1$ lags of each high-frequency predictor.[18] This implies that sg-LASSO promotes sparsity between and within groups.[19]

Moreover, the high-frequency predictor set $\boldsymbol{x}^{(w)}$ in (2.12) is based on orthogonal Legendre polynomials of degree L that aggregate over the high-frequency lags of each predictor. They can be viewed as predetermined weights that alleviate overfitting by reducing the predictor-dimension in $\boldsymbol{x}^{(w)}$ from a factor of $(p+1)$ to L. In our empirical exercise, the sg-LASSO is implemented with $p = 3$ and $L = 1$. This means that four high-frequency lags are considered, which will be aggregated with equal weights for the Legendre polynomial of order $L = 0$ while $L = 1$ features an increasing linear function and thereby favors more distant lags.[20] In addition, we link these high-frequency lags to the contemporaneous information set only, giving rise to missing observations at the end of the sample (ragged-edge problem) when nowcasting on days 8, 15 and 22. Hereby we replace the ragged edges with random-walk updates of the latest month-on-month information available at the time of the nowcast. Therefore, in terms of model specification, sg-LASSO departs from standard shrinkage based on Eq. (1.2) in two aspects: the number of high-frequency lags ($p = 3$ rather than $p = 0$) and the presence of ragged edges. Finally, having estimated the parameters using either Equation (1.2) or (2.12), we can form a nowcast $\hat{\pi}$ by taking a new set of observations $\tilde{\boldsymbol{X}}$ and multiplying by $\hat{\boldsymbol{\theta}}$.

The shrinkage hyperparameters λ and α are tuned in a data-driven manner using time series cross-validation, whereas we set the grid values $(0, 0.25, 0.5, 0.75, 1)$ for α. Differently from the standard cross-validation procedure, in which folds are randomly se-

[18] $\alpha \in [0, 1]$ determines the relative importance of LASSO-sparsity and the group structure.

[19] Note that our application requires us to assume that each monthly predictor in $\boldsymbol{X}$ represents a whole group in (2.12).

[20] The choice of $L = 1$ delivers similar results compared to $L = 2$ but at a lower computational cost.

lected assuming that observations are independently and identically distributed, time series cross-validation splits the training dataset into time slices that retain the chronological order. Therefore, time series cross-validation takes place sequentially and avoids using future observations to fit the model (for a review, see Arlot and Celisse, 2010; Goulet Coulombe et al., 2022; Bergmeir et al., 2018). In our empirical exercise, we start with a 36-month initial fixed window with sequential folds of 12 months.

Tree-based methods

Tree models are based on decision trees, which are nonparametric methods that recursively divide the predictor space according to a pre-determined splitting rule. First proposed by Breiman (2001), random forests are an extension of decision trees in which the results from several non-correlated (or with very small correlation) trees randomly chosen are aggregated to form a prediction. The predictions of the trees in a forest are averaged in such a way that decreases the variance of the final predictions while maintaining the flexibility of the trees. Specifically, for a random forest with B trees, an univariate prediction is given by

$$\hat{\pi}(\tilde{\boldsymbol{X}}_m) = \frac{1}{B}\sum_{b=1}^{B}\hat{\pi}_b(\tilde{\boldsymbol{X}}_m), \tag{1.4}$$

where $\hat{\pi}_b(\tilde{\boldsymbol{X}}_m)$ is the prediction of the b-th tree using new data $\tilde{\boldsymbol{X}}_m$, and m here denotes a subset of all available predictors. RF can deal with high dimensional data without suffering from the curse of dimensionality, but in comparison to a single tree, the forests lack interpretability (James et al., 2013). Nonetheless, random forests have shown to be highly competitive against other ML methods and traditional econometric frameworks when used to forecast inflation (see Medeiros et al., 2021; Araujo and Gaglianone, 2023, among others).

The LLF method proposed by Friedberg et al. (2020) is the combination of a random forest with a local linear regression. In general terms, it combines the RF ability to deal with high-dimensional and nonlinearities with the smoothness of a local linear regression. It is a two-step approach in which the random forest is used to obtain weights for observations that will be later used in the local linear regression with a ridge-type penalty.

To find the weights using a random forest, we start from (1.4):

$$\begin{aligned}\hat{\pi}\left(\tilde{\boldsymbol{X}}_m\right) &= \frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{K_b}\boldsymbol{\theta}_{k,b}1_{\tilde{\boldsymbol{X}}_m\in\mathcal{J}_{k,b}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{\boldsymbol{X}_i\in\mathcal{J}_b(\tilde{\boldsymbol{X}}_m)}\frac{\pi_i}{\mathcal{J}_b(\tilde{\boldsymbol{X}}_m)} \\ &= \frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\frac{\pi_i 1_{\boldsymbol{X}_i\in\mathcal{J}_b(\tilde{\boldsymbol{X}}_m)}}{\mathcal{J}_b(\tilde{\boldsymbol{X}}_m)} = \sum_{i=1}^{n}\alpha_i(\tilde{\boldsymbol{X}}_m)\pi_i,\end{aligned} \tag{1.5}$$

where $1_{\tilde{\boldsymbol{X}}_m \in \mathcal{J}_{k,b}}$ is an indicator function denoting that $\tilde{\boldsymbol{X}}_m$ belongs to the region $\mathcal{J}_k$ in tree b, $\theta_{k,b}$ is a parameter, and $\cdot$ denotes the cardinality of a set. The quantity π_i denotes a response paired with X_i (from the in-sample information), from which n points are available. The term $\alpha_i(\tilde{\boldsymbol{X}}_m)$ is called forest weight and denotes the fraction of trees that allocates $\tilde{\boldsymbol{X}}_m$ in the same leaf as the predictor vector $\boldsymbol{X}_i$. In Eq. (1.5), the regression forest will assign higher weights to sample points closer to $\tilde{\boldsymbol{X}}_m$ since the prediction is an average over a set of trees. The forests can adapt the weights, such that a predictor that has little relation with π_i will appear less frequently when making splits (Athey et al., 2019).

The second step is a local linear regression. Specifically, $\pi(\tilde{\boldsymbol{X}}_m)$ will be the local average, which can be estimated together with a $\boldsymbol{\theta}(\tilde{\boldsymbol{X}}_m)$ through the following optimization problem:

$$\begin{pmatrix} \hat{\pi}\left(\tilde{\boldsymbol{X}}_m\right) \\ \hat{\boldsymbol{\theta}}\left(\tilde{\boldsymbol{X}}_m\right) \end{pmatrix} = \underset{\pi,\boldsymbol{\theta}}{\arg\min} \left\{ \sum_{i=1}^{n} \alpha_i\left(\tilde{\boldsymbol{X}}_m\right) \left(\pi_i - \pi\left(\tilde{\boldsymbol{X}}_m\right) - \left(\boldsymbol{X}_i - \tilde{\boldsymbol{X}}_m\right) \boldsymbol{\theta}\left(\tilde{\boldsymbol{X}}_m\right)\right)^2 + \lambda \left\| \boldsymbol{\theta}\left(\tilde{\boldsymbol{X}}_m\right) \right\|_2^2 \right\}, \tag{1.6}$$

where $\hat{\pi}\left(\tilde{\boldsymbol{X}}_m\right)$ is still a prediction for a new point but with the slope of the local linear regression $\boldsymbol{\theta}\left(\tilde{\boldsymbol{X}}_m\right)$, which corrects for the local trend in $\boldsymbol{X}_i - \tilde{\boldsymbol{X}}_m$. The LLF prediction is then based on the intercept $\hat{\pi}\left(\tilde{\boldsymbol{X}}_m\right)$ while the parameter vector $\boldsymbol{\theta}$ is neglected at this stage. Note that the penalization term $\lambda \left\| \theta\left(\tilde{\boldsymbol{X}}_m\right) \right\|_2^2$ plays a role in avoiding overfitting to the local trend and λ is typically chosen via cross-validation. As a result, the LLF can effectively approximate smooth functions through local regression without becoming infeasible with a growing number of predictors.

Additionally, we implement a combination of LASSO with LLF, as suggested by Friedberg et al. (2020). The pre-selection step via the LASSO might lead to improved predictive performance given that it helps mitigate the curse of dimensionality and better handles multicollinearity by selecting the most informative predictor among the group. Chinn et al. (2023) offer a broader discussion on multi-step nowcasting approaches composed of pre-selection and factor extraction before the estimation of tree-based models.

Chipman et al. (2012) introduce the BART method, which can be viewed as the Bayesian counterpart to random forests. BART predictions are derived from several trees, but opposed to RF, are here sequentially estimated using the residuals from the preceding tree as the dependent variable. Hence, each subsequent tree attempts to capture the remaining variability not explained by the previous trees. In general terms, each Bayesian (regression) tree is defined by $\boldsymbol{\mathcal{T}}$, a collection of interior nodes; and $\boldsymbol{\mathcal{M}}$ a set of

parameter values that are associated with the terminal nodes. The set $\boldsymbol{\mathcal{T}}$ is also called tree structure and contains the information on the topology of the trees: whether a node is terminal or not and how to make splits in non-terminal nodes.

A BART defines a function $g(\boldsymbol{X}_i, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}})$ which maps a row $\boldsymbol{X}_i$ (from the predictor matrix $\boldsymbol{X}$) to a particular $\theta_j \in \mathcal{M},\ j \in 1, \ldots, \boldsymbol{\mathcal{M}}$. Predictions from individual trees form the final BART prediction and are obtained by sampling from the posterior distribution. We closely mirror the prior specification used in Chipman et al. (2012). This implies a uniform prior to determine both the variable for a split and the corresponding cutpoint. A conjugate normal prior is used for the predictions on the terminal nodes and a conjugate inverse χ^2-square for the (constant) error term of the model. Finally, the probability of growing another layer in a tree is given by $\alpha(1+d)^{-\beta}$, where d is the current depth of the tree, while $\alpha \in (0, 1)$ and $\beta \in^+$ are hyperparameters.

In our empirical exercise, we set the hyperparameters of the above tree-based methods to default values respectively recommended by Breiman (2001), Friedberg et al. (2020) and Chipman et al. (2012).[21] To construct BART predictions we estimate 200 trees using 1000 posterior draws, with 100 draws as burn-in. For the tree structure, we use $\alpha = 0.95$ and $\boldsymbol{\theta} = 2$, which penalizes bigger trees. For the conjugate normal prior of the predictions, we centered the prior at 0 and the variance is equal to $0.5/k(\sqrt{m})$ where $k = 2$ and $m = 200$ denote the number of trees. For the variance prior, the hyperparameters ν and λ of the $(\nu\lambda)/\chi_\nu$ distribution are obtained from the standard deviation of the response variable in the estimation sample and a factor of 10, respectively.

1.4 Empirical results and discussion

In this section, we investigate the performance of our mixed-frequency ML models for nowcasting Brazilian inflation using a real-time dataset with macro-financial predictors that span from June 2004 to December 2022. For the out-of-sample evaluation, we focus on the interval from January 2013 to the end of our sample using an expanding window scheme. This evaluation sample is constrained by data availability, as the release calendar for the entire dataset is only available from January 2013 onward. Nonetheless, it includes two of the most inflationary periods in Brazil's recent history: the economic domestic crisis of 2014 and the COVID-19 pandemic.

To assess the accuracy of our IPCA predictions, we compare them against SPF expectations – both the median and the Top 5 – published by the BCB. We update our nowcasts on a set of fixed days (8, 15, 22 and end-of-month) using the most recent increments of monthly and high-frequency (weekly and daily) data respecting the release calendar.

[21] Tuning hyperparameters via time series cross-validation results in a lower nowcasting performance and substantially increased computational burden. These results are available upon request.

While model estimation is based on month-on-month transformations of variables, we use year-on-year IPCA rates as our ultimate metric for performance evaluation. Consequently, we adjust our model-derived nowcasts for month-on-month IPCA rates before comparison with actual realizations of the target. Our findings highlight the superior performance of shrinkage methods over tree-based methods. Moreover, a deeper analysis of key modeling choices in Eq. (1.1) reinforces the importance of eliminating ragged edges in a real-time setup and to account for some degree of informed judgment in SPF data.

1.4.1 Out-of-sample results

To compare the nowcasting performance across ML models, we use the root mean squared error (RMSE) of a competing model M_i relative to the benchmark SPF's median expectations at the nowcast horizon h. The RMSE is defined as follows

$$\text{RMSE}_{M_i,h} = \sqrt{\frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} e_{t,M_i,h}^2}\,, \tag{1.7}$$

where $e_{t,M_i,h} = \pi_t - \hat{\pi}_{t|t-h;M_i}$ is the corresponding nowcasting error with information up to $t-h$.[22] To test whether nowcasts generated by the ML model M_i are statistically different from the benchmark, we conduct the Diebold-Mariano (DM) test (Diebold and Mariano, 1995).

Table 1.3 reports the nowcasting performance of competing models, evaluated in terms of RMSE relative to the benchmark, whereas the rows refer to the nowcast horizon and, consequently, the within-month information set. The results underscore the superior performance of shrinkage methods across all nowcast horizons, highlighting the efficacy of employing penalized regressions alongside a comprehensive real-time dataset. Specifically, standard shrinkage via the LASSO, Ridge and ENet consistently yields lower RMSE values, resulting in statistically significant gains of 8.5% up to 17% compared to the median SPF expectations. While the LASSO generally outperforms, Ridge shows slightly better performance when nowcasts are made on day 22. Relative to the Top 5 SPF benchmark, these predictive gains range from 4% to 15%, indicating a substantial difference despite their status as the best-performing institutions before each nowcasting round. Therefore, we match those results of Medeiros et al. (2016) and Garcia et al. (2017) for forecasting Brazilian inflation, which found that techniques based on LASSO outperform at the very short horizon.

Across information sets, the performance of our ML models relative to the benchmark generally increases with the nowcast horizon. On days 8 and 15, for example, standard

[22] The nowcasting evaluation using the mean absolute error (MAE) slightly changes compared to the RMSE metric. This implies that our results are not affected by a few large errors, making them robust to outliers and asymmetries.

Table 1.3: RMSE: ML methods relative to the SPF benchmark

	SPF		Shrinkage-based models				Tree-based models			
Horizon	**Median**	**Top 5**	**LASSO**	**Ridge**	**ENet**	**sg-LASSO**	**RF**	**LLF**	**BART**	**LASSO-LLF**
Day 8	1	0.932	0.830**	0.856*	0.842**	0.930*	1.027	0.965	0.963	0.910
Day 15	1	1.014	0.865**	0.879	0.870*	0.955	1.089	1.035	1.033	1.011
Day 22	1	0.942*	0.833*	0.830*	0.833*	0.920	1.247	1.046	1.134	0.983
End-of-month	1	0.951	0.915	0.974	0.915	0.936	1.399	1.371	1.310	1.042

Notes: The table reports the RMSE for each competing model relative to the survey of professional forecasters (SPF, median). Results for the Diebold and Mariano (1995) test in the event of outperformance relative to the benchmark are indicated by the symbols * (5% level) and ** (1% level).

shrinkage methods can respectively cut the nowcast errors by 16% and 13.5% on average. Based on the absolute RMSE, this translates into 4.5 and 2.8 basis points of higher accuracy for tracking the year-on-year inflation target. It is worth noting the significant decline in nowcasting gains for the end-of-month horizon. Exclusively at this horizon, our model-based nowcasts have not exhibited statistically significant differences from the benchmark. This trend aligns with the timing of price indicator releases, which predominantly occur towards the end of the month, particularly the IPCA-15. Such a dynamic likely prompts professional forecasters to increase the frequency of their updates, narrowing the information advantage exploited by our timely nowcasts at longer horizons.

The dominance of LASSO among shrinkage methods indicates the need for a more aggressive variable selection to nowcasting inflation dynamics. This inherent feature of LASSO is particularly advantageous in dealing with the high degree of collinearity among the price indicators within our dataset. Despite cross-validation tuning of the hyperparameter α in ENet, both Ridge and ENet produce more evenly distributed estimates across those highly correlated predictors, resulting in slightly increased overfitting in this context. Furthermore, the sg-LASSO yields a considerably lower precision relative to standard shrinkage for longer horizons (day 8 up to day 22), possibly affected by the presence of ragged edges (see Section 1.4.3). For end-of-month horizons, where ragged edges are eliminated, sg-LASSO outperforms Ridge and narrows the gap against LASSO and ENet. This highlights the importance of an assertive selection of predictors and suggests that our baseline setting is not high-dimensional enough for sg-LASSO to thrive.

What drives the poor nowcasting results of tree-based methods? Although the flexibility of these methods allows for potentially detecting turning points and complex nonlinear dynamics in the data, their data-hungry characteristic leads to poor performance in our setup. Notable exceptions are observed for the longest horizon of day 8, where LLF, BART and LASSO-LLF yield RMSE reductions between 3.5% and 9% relative to the benchmark, tough statistical significance is not achieved. The slightly improved performance of LASSO-LLF corroborates the hypothesis that tree-based methods might be ill-equipped to handle the limited samples of macroeconomic time series; pre-selecting strong predictors from a large dataset works best and goes in line with the recommenda-

tion from Friedberg et al. (2020) when dealing with large datasets for the LLF. Moreover, these results suggest the absence of relevant temporal nonlinearities in Brazilian data.

The findings in Table 1.3 prompt the question of whether relative performance is constant throughout the evaluation period or largely affected by inflationary shocks. To gain further insights into the evolution of loss differentials, we report the cumulative sum of squared forecast error:

$$\text{CUMSFE}_{M_i,h} = -\sum_{t=t_0}^{t_1} \left(e^2_{t,M_i,h} - e^2_{t,M_{\text{SPF}},h}\right). \tag{1.8}$$

A positive value of CUMSFE indicates an outperformance of the ML model M_i relative to the benchmark median SPF expectations for horizon h and from period t_0 up to t_1. Negative values imply the opposite.

Figure 1.3 exhibits CUMSFE developments for shrinkage methods across different nowcast horizons. It becomes crystal clear that the inflationary period following the COVID-19 crisis is a game changer in terms of loss differentials. Particularly, large nowcasting gains build up from September 2020. During this period of persistent high inflation, we observe the largest jumps in CUMSFE for nowcasts made on days 8 and 15 using the LASSO. Moderate gains are also achieved when nowcasting at shorter horizons. It is worth emphasizing that these findings during the pandemic mostly drive performance evaluation based on full-sample metrics, as in Table 1.3. For instance, our previous hypothesis based on the performance across information sets – professional forecasters tend to update their expectations more frequently as the information set increases within the reporting month – is a trend predominantly observed within the context of the COVID-19 crisis.

Turning to the years preceding the pandemic, differences in predictive accuracy between shrinkage methods and SPF expectations are modest across all shrinkage methods. LASSO and sg-LASSO prove the most reliable models by consistently keeping an edge relative to the benchmark for most horizons. In contrast, nowcasts purely based on tree-based models can be highly detrimental during calm times, as shown in Figure 1.6 of Appendix 1.B. The LASSO-LLF performs roughly on par with SPF expectations, reinforcing the idea that a pre-selection step proves beneficial for tree-based methods. Nonetheless, amidst the COVID-19 crisis, most tree-based models exhibit a rising competitive advantage over the benchmark on days 8 and 15.

Figure 1.3: CUMSFE: shrinkage methods versus the SPF benchmark

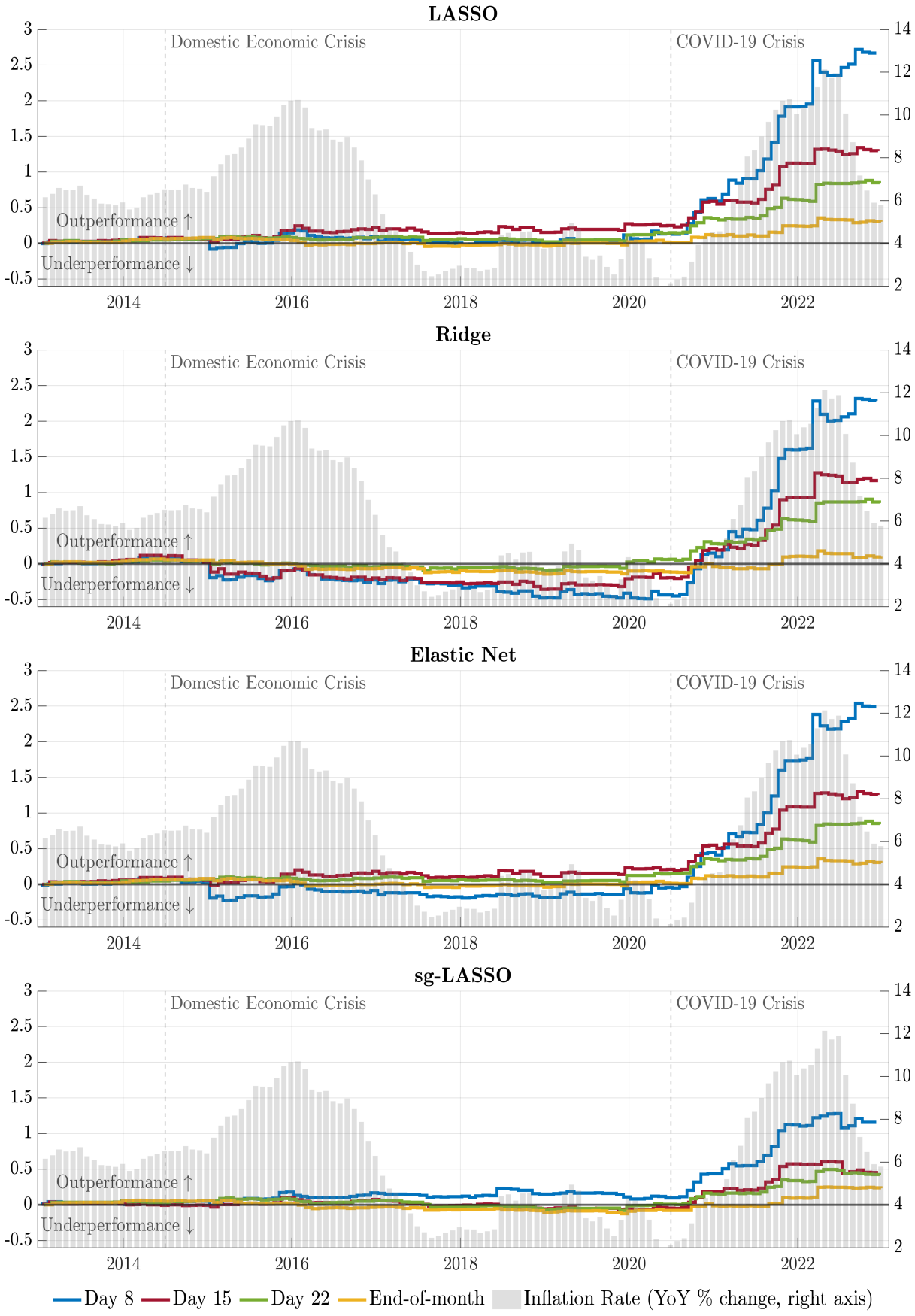


Figure 1.7 in Appendix 1.B reports the fluctuation test, introduced by Giacomini and Rossi (2010), and reaffirms the previous analysis. Predictive gains relative to SPF expectations change substantially over time, depending on the model and horizon, and are prominent in the aftermath of the pandemic. Standard shrinkage via the LASSO, Ridge and ENet deliver occasional significant gains at the 10% level throughout 2021. Other models also produce statistically significant gains during this period: sg-LASSO on day 8, and LASSO-LLF on both day 8 and end-of-month. Besides, the picture reveals a clear discrepancy between shrinkage- and tree-based models, as expected from previous results. Finally, a higher dispersion of prediction accuracy across models can be observed during turbulent times such as the domestic economic crisis of 2014 and the pandemic.

1.4.2 Interpreting the best-performing model

Based on the variable selection performed by our most effective strategy, we investigate the relative importance of the selected predictors fitted via the LASSO. Heatmaps illustrating the evolution of coefficient estimates at each nowcast horizon are presented in Figure 1.8 of Appendix 1.B. The x-axis denotes a nowcasting round in the evaluation sample, while predictors are displayed in the y-axis. Consequently, the coefficient value associated with a given predictor on a specific date within the evaluation sample determines the color intensity reflected in the graph.

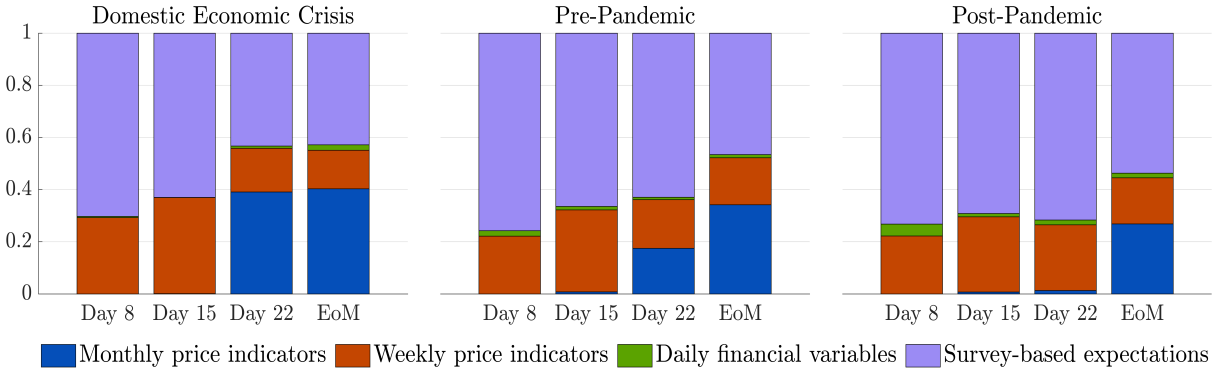
Comparing all panels in Figure 1.8, we observe that LASSO prompts a fairly sparse structure at higher nowcast horizons while a more dense structure prevails at horizons approaching the end of the month. Two factors contribute to this pattern: (i) increased availability of data on monthly price indicators as the month unfolds, and (ii) signals associated with price developments in the reporting month t become more accurate as month-on-month rates rely less on the information set from $t-1$. Not surprisingly, on days 8 and 15 (panels 1.8a and 1.8b), SPF expectations stand out as the primary predictor, with average coefficient estimates near 0.7 but exceeding 0.8 towards the end of the sample. Alongside SPF, the high-frequency price indicators IPC-S and FIPE are consistently selected across the entire evaluation sample, albeit with comparatively smaller coefficients – e.g., on average, 0.16 and 0.1 respectively for day 15. At the same time, energy prices, interest rate variables and commodity prices regularly enter the forecasting model, though with modest coefficient values.

As the horizons approach the end of the month, the low-frequency but timely indicator IPCA-15 takes on enormous importance, with coefficient estimates reaching 0.5 in many cases. Conversely, SPF expectations lose a substantial portion of their relevance. One plausible hypothesis is that professional forecasters adjust their survey responses in response to the release of this indicator.

Using a more aggregate approach, we assess the joint relevance of each class of predictors

across different nowcast horizons and sub-periods. As a variable-importance measure, Figure 1.4 depicts the weighted sum of absolute LASSO estimates grouped into four categories of predictors as described in Table 1.1: monthly price indicators, weekly price indicators, daily financial variables, and daily SPF expectations. As shown previously, SPF expectations, closely trailed by weekly price indicators (particularly IPC-S and FIPE), exert the most substantial impact on shaping our model-based nowcasts. This suggests that SPF expectations not only incorporate up-to-date information from our dataset but also integrate informed judgment that extends beyond relying solely on hard predictors for inflation. However, as recent data on monthly and weekly price indicators becomes available throughout the reporting month, their relevance in model estimation rises, subsequently diminishing SPF's weights as we approach the end-of-month horizon. Particularly, the availability of contemporaneous data on monthly price indicators after the third week typically elevates their relative importance when nowcasting on day 22 and end-of-month. On the other hand, financial variables play a minor role in shaping our model-based nowcasts due to their limited informativeness regarding current inflation dynamics, especially when compared to the signals already present in the dataset.

Figure 1.4: Variable relevance via coefficient estimates using LASSO



Notes: This Figure reports the weighted sum of absolute coefficient estimates fitted via the LASSO and grouped into different categories of predictors (see Table 1.1) on days 8, 15, 22 and end-of-month (EoM). The "Domestic Economic Crisis" covers the period from 2013 to 2016 while March 2020 divides the "Pre-Pandemic" period and the start of the COVID-19 crisis ("Post-Pandemic").

Turning to different sub-periods, the informed judgment in SPF data appears to weigh more heavily on nowcasting the inflation surge following the pandemic. This tendency toward increasing SPF weights is already evident in the calm period preceding the pandemic, particularly when nowcasts are made on day 22. In this case, monthly price indicators lose significant ground relative to SPF and weekly price indicators. This suggests that SPF encompasses more robust signals about the target dynamics as we advance in the sample, possibly stemming from a more timely update of forecasts made by specialists as new information on other predictors is released. Additionally, off-model information proves more valuable during turbulent times, especially if the nature of post-pandemic

price spikes differs from the nature of past inflationary shocks in the sample. For example, the inflationary wave induced by 2014's domestic economic crisis was more accurately anticipated by relying solely on hard signals from price indicators.

The natural question that follows is what part of the information set mostly contributes to the outperformance relative to the tough SPF benchmark. To address this question, we replicate our recursive exercise using the SPF nowcasting errors as the dependent variable in LASSO regressions and plot the period-wise estimates in Figure 1.9 of Appendix 1.B. Surprisingly, professional forecasters mainly overlook recent data increments of FIPE when nowcasting at longer horizons. Other weekly price indicators are also partially disregarded across all horizons but to a smaller extent. As for financial variables, they hardly contribute to explaining expert's errors; except for occasional minor effects of Bloomberg's commodity index and interest rate movements (SELIC, DI10 and SPREAD) for shorter horizons. Although a slightly negative intercept estimate shows up across the board, we observe a consistent downward bias of SPF expectations. Notably during the pandemic period after March 2020, where the average nowcast error based on year-on-year percentage points yields 0.14 for end-of-month nowcasts compared to only 0.003 in the pre-pandemic sample. Moreover, it appears that experts do not fully adjust for their past errors given the significant effect of the lagged dependent variable on day 8 and end-of-month.

Therefore, SPF nowcast errors can be partially predicted with the relevant information set available at the nowcast date, explaining the additional improvements reported in Section 1.4.1. But what if we add these LASSO-based forecasts for SPF nowcast errors back to the SPF expectations to obtain implied nowcasts for the IPCA target? In terms of relative RMSE as in Table 1.3, this modeling strategy fares better than our previous nowcasts on days 8 and 15. More precisely, respective predictive gains of 18.5% and 16.5% are now obtained compared to the SPF benchmark. For the remaining horizons, the nowcasting precision significantly drops when competing against the best-performing models in Table 1.3.

1.4.3 A deeper assessment of key modeling features: guiding accurate inflation nowcasts

What drives the accuracy of our inflation nowcasts? To explain why the baseline mixed-frequency ML structure introduced in Section 1.3.1 coupled with a highly informative dataset successfully outperforms tough benchmarks, we assess the value added of key modeling choices in Eq. (1.1). More specifically, we investigate three aspects: the impact of the high-frequency lag choice; the impact of eliminating ragged edges; and the impact of using SPF expectations in the predictor set. It is noteworthy to recall that our baseline

specification of (1.1), to which we compare the alternative strategies, features the following choices: only the most recent high-frequency data enters the model by setting $p = 0$, there is no ragged-edge problem, and we include the SPF as a predictor.

First, we investigate whether past month-on-month high-frequency regressors carry predictive value beyond the most recent signal available at the nowcast date $t-h$. This is done by extending our baseline choice of $p = 0$ to account for $p = 3$ high-frequency lags. Hence, our alternative specification here includes the four most recent high-frequency information $\boldsymbol{x}_{t-h}^{(w)}, \ldots, \boldsymbol{x}_{t-h-3/4}^{(w)}$. For example, if we stand at 31 December, we include the high-frequency signals stemming from $\boldsymbol{x}_{31\text{ Dec}}^{(w)}, \boldsymbol{x}_{22\text{ Dec}}^{(w)}, \boldsymbol{x}_{15\text{ Dec}}^{(w)}, \boldsymbol{x}_{8\text{ Dec}}^{(w)}$ rather than just $\boldsymbol{x}_{31\text{ Dec}}^{(w)}$. In general terms, we increase the high-frequency predictor set by a factor of 4.

The impact in terms of RMSE from incorporating these additional high-frequency lags can be seen in Figure 1.5a. The plot indicates that RMSE values generally deteriorate with the inclusion of additional high-frequency lags. This pattern is more pronounced across longer nowcast horizons (days 8, 15 and 22) and tree-based models such as LLF. However, at the end-of-month (yellow dots) a marginal increased gain of 2-5% in RMSE is observed across shrinkage methods, most notably for Ridge regression. Therefore, the last available high-frequency predictor on its own already carries the relevant signal for updating the nowcasts throughout the reporting month, except at end-of-month, whereas including lagged high-frequency information is slightly favorable.

Secondly, we keep the alternative assumption of $p = 3$, however, we redesign the model to incorporate exclusively contemporaneous high-frequency lags $\boldsymbol{x}_{t}^{(w)}, \ldots, \boldsymbol{x}_{t-3/4}^{(w)}$ at any nowcast day in month t. Additionally, we assume that the low-frequency predictor set must include all monthly price indicators in the dataset at any horizon h, regardless of their publication timing. These choices give rise to missing observations at the end of the sample (ragged-edge problem) when nowcasting on days 8, 15 and 22. As for sg-LASSO, we complete the ragged edges with random-walk nowcasts based on the most recently released information.

Figure 1.5b points to a considerable worsening of the nowcast precision when the model suffers from the ragged-edge problem. Except for the RF model in the first week of the month, the loss in performance is consistent within both classes of ML methods. Shrinkage methods, particularly at shorter nowcast horizons, exhibit the most pronounced susceptibility to ragged edges, experiencing average losses approaching 60%. These findings confirm the consensus in the MIDAS literature suggesting that ragged edges worsen the forecasting properties of the model, especially in the very short run (see, e.g., Marcellino and Schumacher, 2010; Andreou et al., 2013; Monteforte and Moretti, 2013).

The reasons for this underperformance are twofold. Firstly, during model estimation, substantial weight is assigned to monthly predictors, which can only bring outdated information from $t-1$ to construct the nowcast for π_t, particularly at longer horizons.

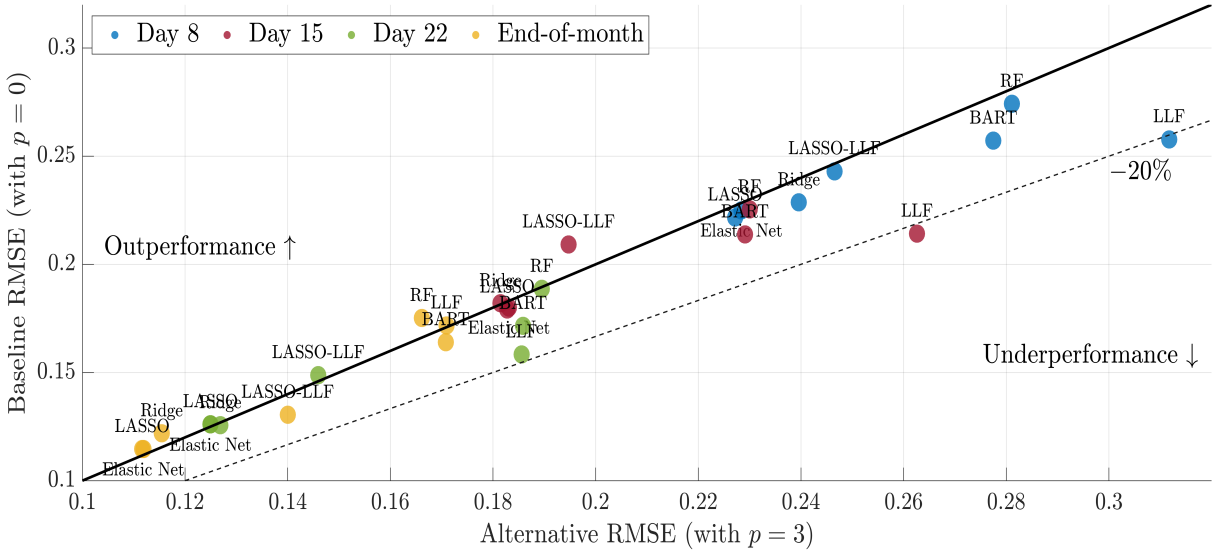
Secondly, the multicollinearity arising from the inclusion of $p = 3$ high-frequency lags in the predictor set somewhat disorients ML methods, hindering their ability to identify accurate high-frequency signals.

Third, we investigate the benefits of incorporating some degree of informed judgment entailed in SPF median expectations. Professional forecasters do not solely rely on models to form their expectations about short-run inflation dynamics but these can also be attributed to judgment, particularly in challenging times such as the COVID-19 crisis where purely model-based forecasts are adversely affected. Since our baseline strategy includes SPF as a high-frequency predictor, we compare it against the alternative specification that excludes any SPF information from the predictor set. Notably, Figure 1.5c shows that adding meaningful off-model information from SPF leads to sizeable nowcasting advantages. Specifically, predictive gains are substantially higher across shrinkage methods, averaging from 27% on day 8 to 36% at end-of-month. This indicates that SPF information on inflation expectations can better discipline parametric model structures.

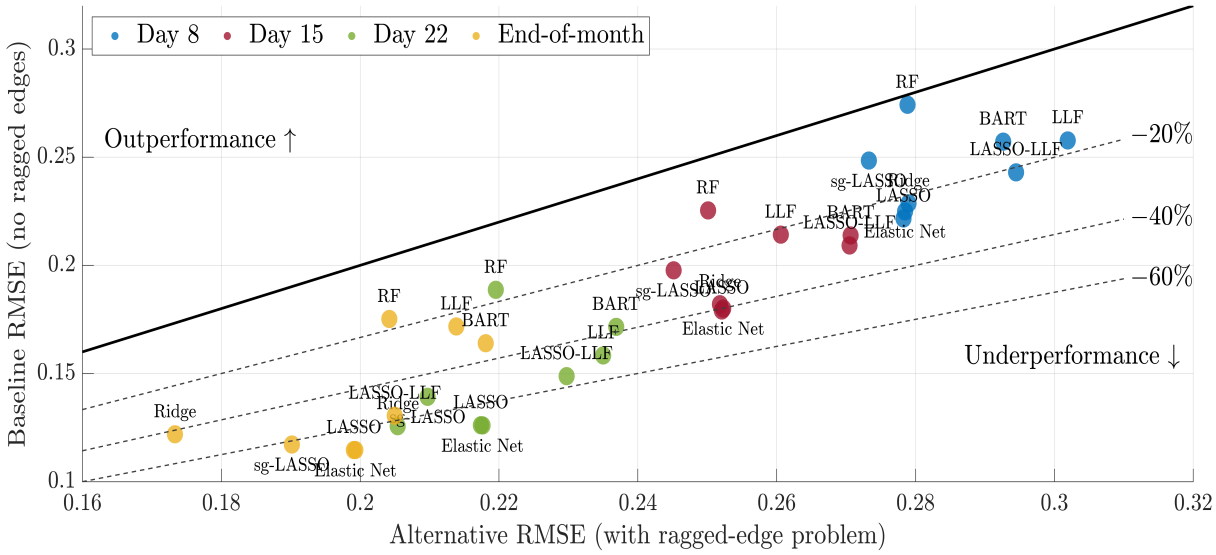
Amidst the turbulence induced by the pandemic, professional forecasters tended to underestimate the inflation surge. Nevertheless, informed judgment in SPF data carried relevant information about rapidly unfolding inflationary trends beyond what was reflected by other predictors in our dataset. In addition, the previous discussion on Figure 1.4 reveals a growing relevance of SPF in constructing LASSO-based nowcasts, especially post-2021. Consequently, environments marked by highly unpredictable and elevated inflation, like the pandemic, are best suited for enriching model-based inflation nowcasts with SPF expectations. These findings align with most of the previous studies that investigate the value added of SPF expectations into model-based forecasts (see, e.g., Bańbura et al., 2023; Bobeica and Hartwig, 2023).

In summary, superior nowcasting accuracy predominantly stems from the combination of a well-designed mixed-frequency ML structure with carefully selected predictors that include some degree of informed judgment in SPF expectations. Specifically, the prediction model must be free from ragged edges. This is first attained through high-frequency leads, preferably focusing solely on the last available high-frequency signal conveyed by $\boldsymbol{x}_{t-h}^{(w)}$. Furthermore, the inclusion of monthly predictors in the model specification should be guided by their real-time data releases; in particular, only those with available contemporaneous data by the nowcast date.

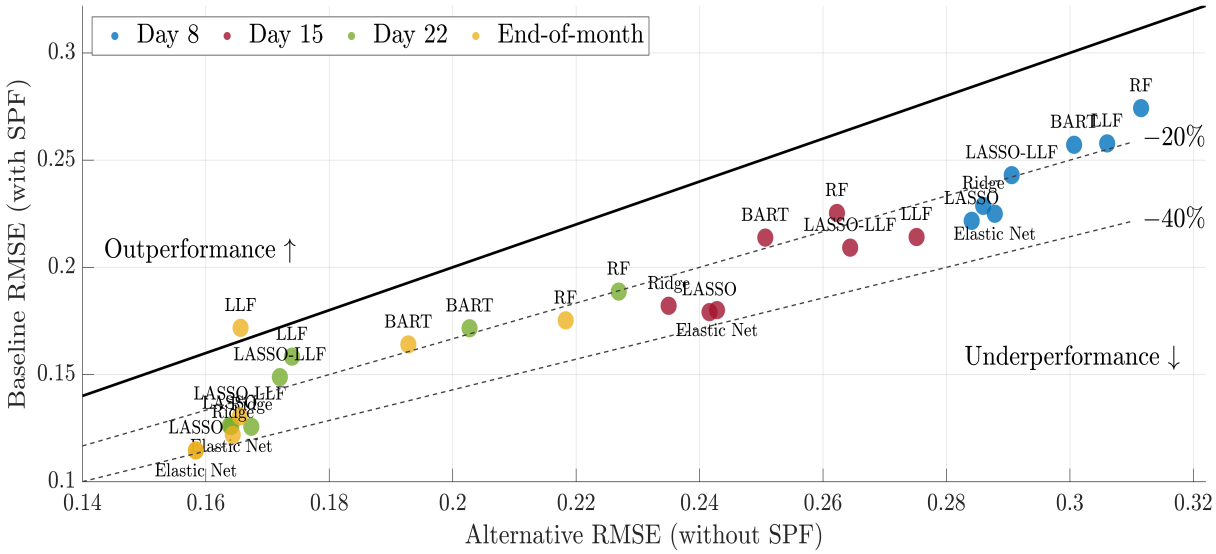
Figure 1.5: Absolute RMSE: alternative versus the baseline specification



(a) Alternative (with $p = 3$ high-frequency lags) versus the baseline (with $p = 0$)



(b) Alternative (with ragged-edge problem and $p = 3$) versus the baseline (no ragged edges)



(c) Alternative (without SPF) versus the baseline (including SPF as predictor)

Notes: This Figure reports the absolute RMSE of the alternative specification versus the baseline model specified in Section 1.3.1. Points below (above) the 45-degree reference line, in solid black, indicate an underperformance (outperformance) of the alternative specification for a given competing ML method and nowcast horizon.

1.5 Summary and conclusions

Machine learning methods have recently gained considerable traction as standard tools for macroeconomic nowcasting, offering an effective solution to handle the increasing availability of high-frequency information stemming from all parts of the economy. In the wake of disruptive events like the COVID-19 pandemic, the demand for such nowcasts has intensified. Yet there remains a notable gap in harnessing ML methods to leverage high-frequency signals for real-time inflation nowcasting.

To address this gap, our study compares shrinkage methods with tree-based models in an environment characterized by persistently high inflation. Our empirical exercise underscores the importance of a well-designed mixed-frequency ML framework to construct robust inflation nowcasts that consistently outperform SPF expectations, with major nowcasting benefits during the COVID-19 inflation surge. Moreover, we show that good nowcasts depend on variable selection performed via the LASSO combined with accurate timely signals from price indicators and informed judgment entailed in SPF data. The findings highlight the adaptability of shrinkage methods to produce accurate nowcasts across different horizons while tree-based methods lead to poor performance due to the limited time series sample and the plausible absence of temporal nonlinearities in our setup. Overall, the timely and high-frequency character of the Brazilian real-time dataset offers valuable insights for policymakers and practitioners seeking to refine their inflation forecasting capabilities in uncertain economic landscapes.

Variable importance analysis via the LASSO fitted coefficients shows that model se-

lection heavily depends on the contemporaneous information set available at the time of the nowcast. Specifically, at longer nowcast horizons, a more sparse model delivers higher predictive gains compared to the benchmark, while exploiting early information from weekly price indicators and SPF expectations. At shorter horizons, shrinkage-based models yield a denser structure that also assigns substantial importance to monthly price indicators, which only enter the predictor set when their contemporaneous signal becomes available. In general, financial variables play a minor role but the combination of timely price indicators with SPF judgments proves highly influential in shaping our model-based nowcasts.

Furthermore, our study sheds light on key modeling choices in a mixed-frequency ML framework. The results suggest implementing the following strategies to achieve higher performance: (i) account for expert judgment in the predictor set, (ii) make the prediction model free of ragged edges, (iii) align the model specification with the release calendar of monthly predictors, and (iv) prioritize the most recent high-frequency signal available in the information set. With our framework, we can significantly improve upon SPF expectations, even outperforming the Top 5 SPF institutions, which are widely regarded as the most challenging benchmark for forecasting Brazilian inflation dynamics. As a fruitful avenue for further research, one could expand our analysis to encompass additional classes of ML methods and contrast them with traditional econometric frameworks such as factor models and mixed-frequency Bayesian VARs. Moreover, one could assess the economic value of our nowcasting gains in monetary policy decisions and portfolio allocation strategies.

Appendix 1

1.A Mixed-frequency framework in matrix form

For expositional simplicity, let us reduce the general multiple-predictors specification (1.1) to the single generic high-frequency predictor $x_t^{(w)}$ and neglect both the low-frequency predictors and seasonal dummies. From there, assume the latest data release for the target variable is associated with a given month t. Based on the high-frequency information set available up to the nowcast point, say $t+1-h$, and pre-sample information $\{\pi_0, x_0^{(w)}, x_{0-1/4}^{(w)}, \ldots, x_{0-p/4}^{(w)}\}$, one can construct the nowcast for π_{t+1} at horizon $h = j/w$, with $j \in \{0,1,2,3\}$, by using the following matrix representation for model estimation:

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_t \end{bmatrix} = \begin{bmatrix} 1 & \pi_0 & x_{1-h}^{(w)} & x_{1-h-\frac{1}{4}}^{(w)} & x_{1-h-\frac{2}{4}}^{(w)} & \cdots & x_{1-h-\frac{p}{4}}^{(w)} \\ 1 & \pi_1 & x_{2-h}^{(w)} & x_{2-h-\frac{1}{4}}^{(w)} & x_{2-h-\frac{2}{4}}^{(w)} & \cdots & x_{2-h-\frac{p}{4}}^{(w)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \pi_{t-1} & \underbrace{x_{t-h}^{(w)}}_{\text{nowcast day } (nd)} & \underbrace{x_{t-h-\frac{1}{4}}^{(w)}}_{nd-\frac{1}{4}} & \underbrace{x_{t-h-\frac{2}{4}}^{(w)}}_{nd-\frac{2}{4}} & \cdots & \underbrace{x_{t-h-\frac{p}{4}}^{(w)}}_{nd-\frac{p}{4}} \end{bmatrix} \begin{bmatrix} c \\ \rho_1 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_{p+1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix} \quad (1.9)$$

For example, suppose we stand at day 15 of December and we want to construct the nowcast for π_{Dec} assuming a general high-frequency lag order p. In this case, the forecast horizon is $h = 2/4$ and we estimate the model using monthly data until November and weekly data until 15 November. To account for the lags p, the last high-frequency observations in (1.9) will respectively be associated with 15 November, 8 November, 31 October, 22 October, 15 October, and so on up to the corresponding lag-length p. From there, the nowcast for π_{Dec} is constructed using the estimated coefficients and all the low- and high-frequency information available until 15 December.

Ultimately, note that (1.9) makes explicit that the general prediction model is still written at the monthly frequency but accounting for the w high-frequency time increments within each common period t. The nowcast for the inflation rate at periods $t+1, \ldots, T$ can then be updated regularly using the high-frequency data increments that become available after t and well before official releases of the target inflation rate.

1.B Supplementary results

Figure 1.6: CUMSFE: tree-based methods versus the SPF benchmark

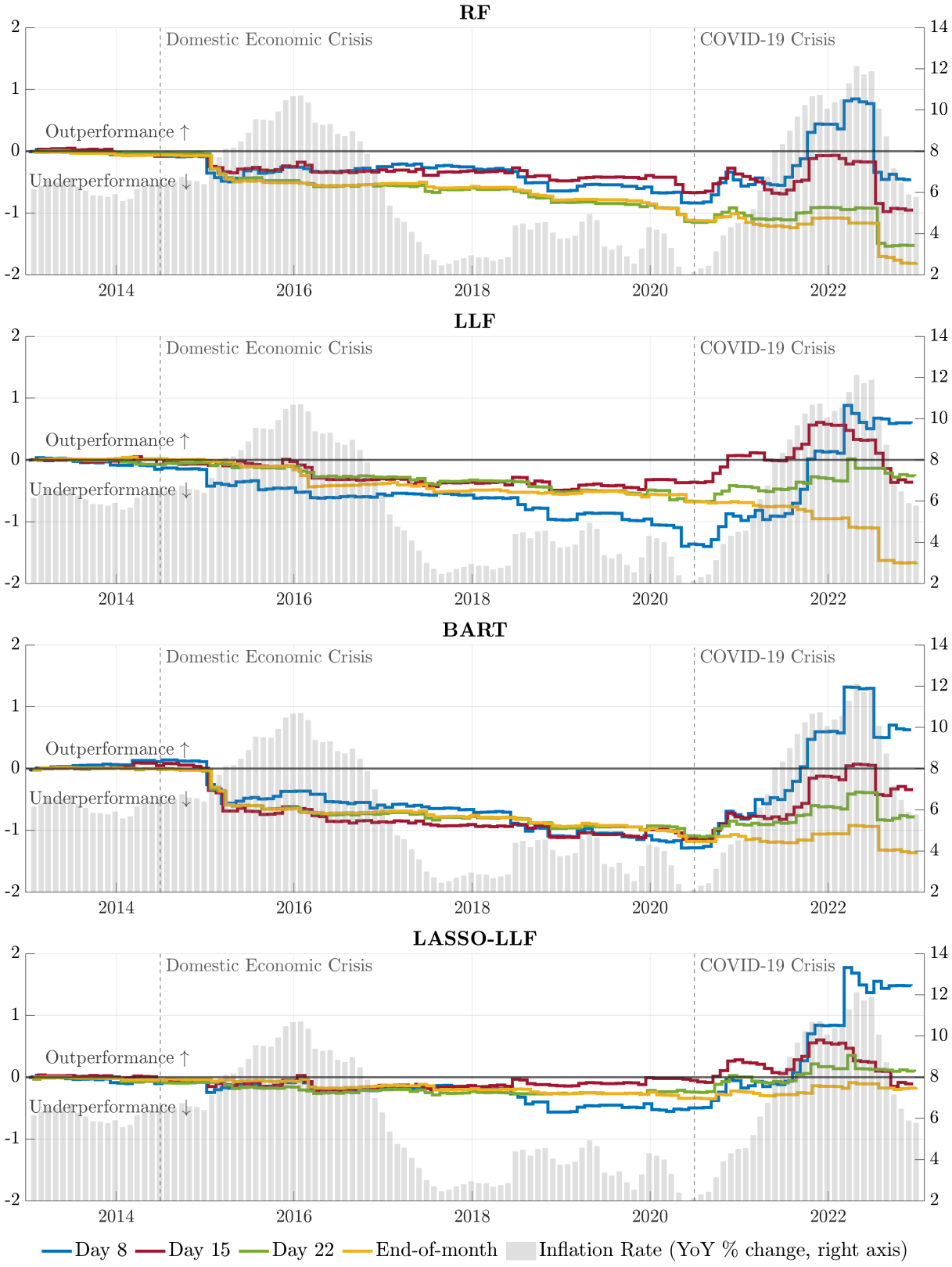
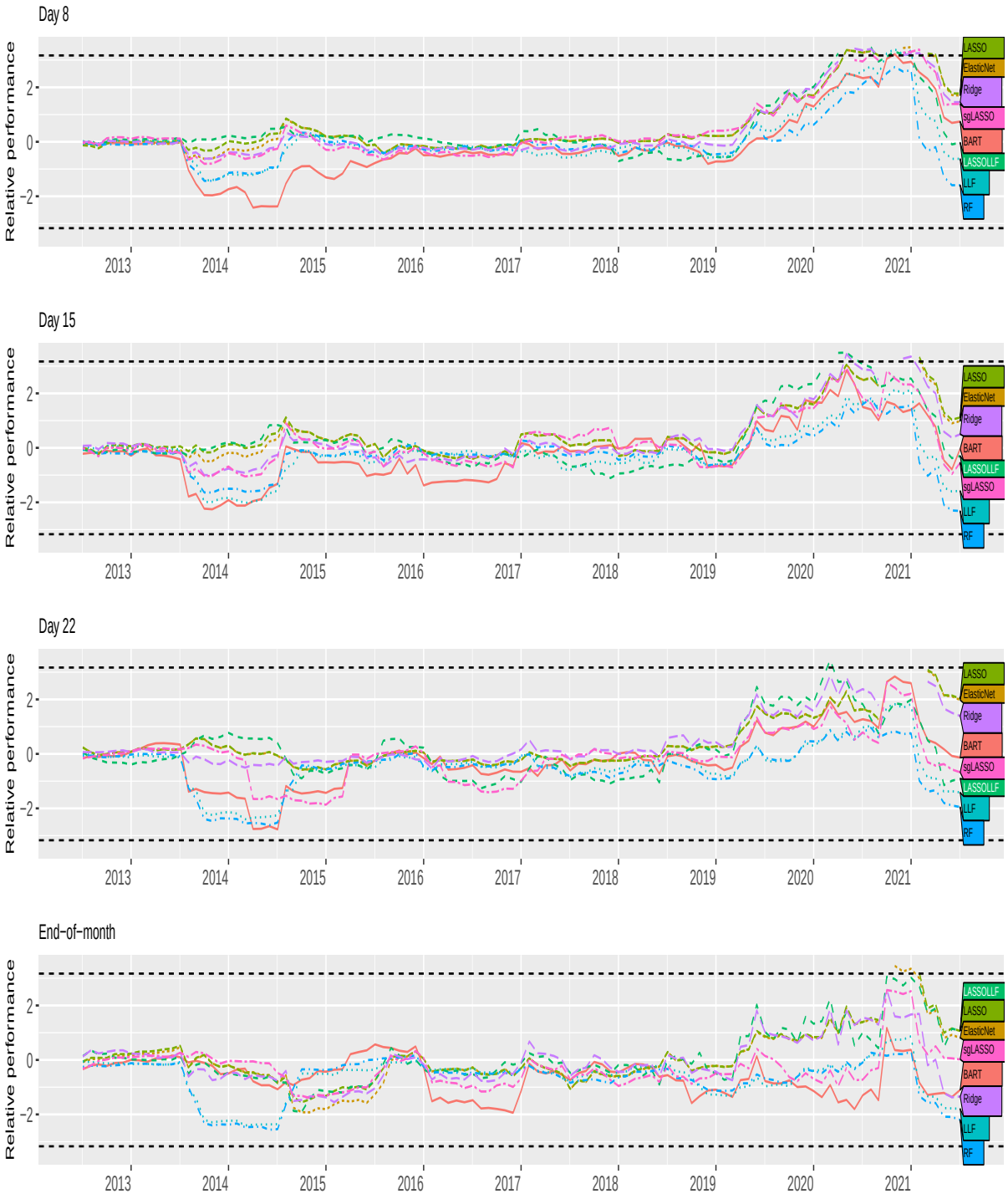
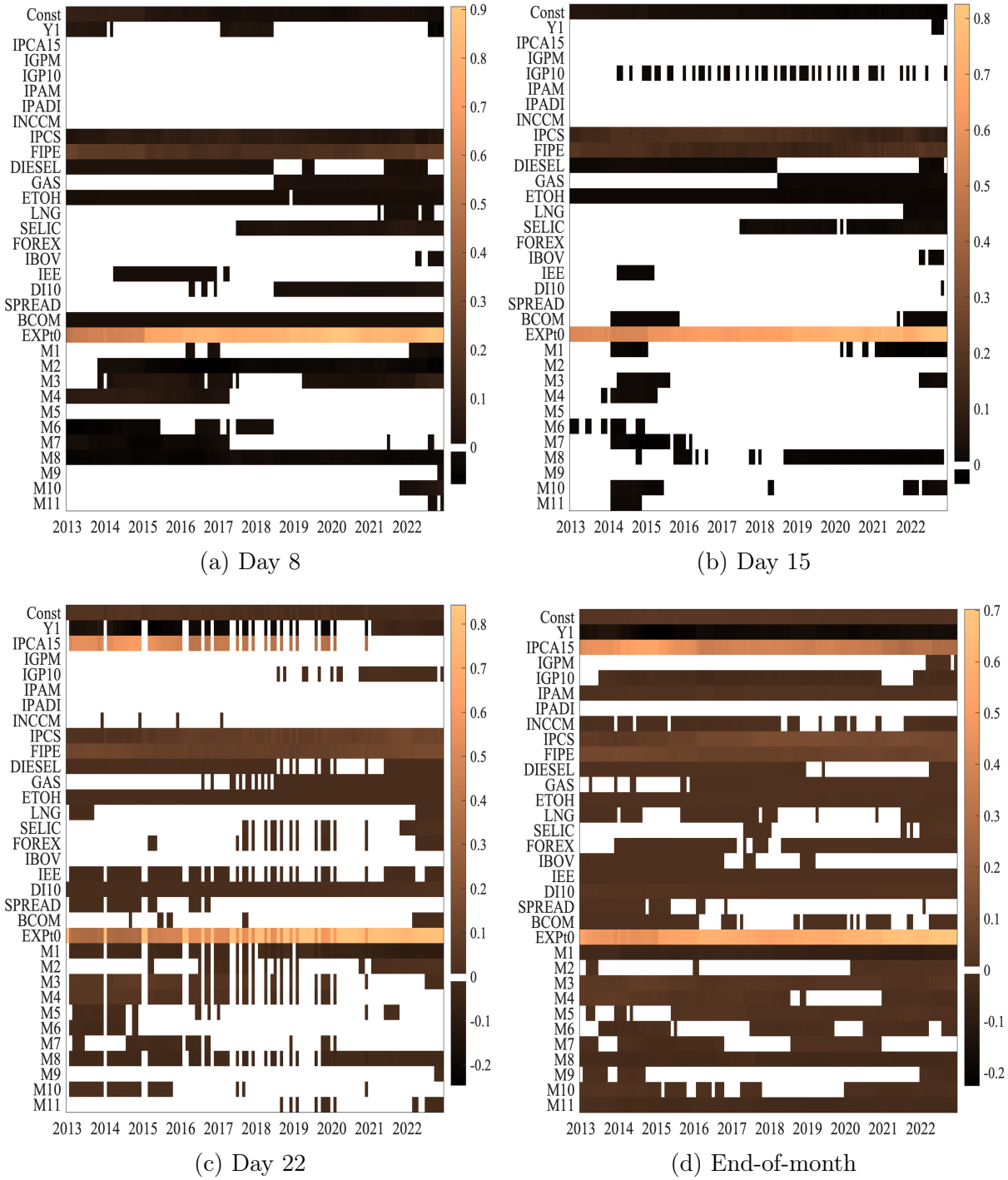


Figure 1.7: Fluctuation test: ML competing models versus the SPF benchmark



Notes: This Figure reports the fluctuation test from Giacomini and Rossi (2010) based on the squared loss differential between a machine learning method and SPF nowcasts. Areas between the horizontal dashed lines correspond to the 90% confidence interval of the two-sided statistical test. We used as window parameters of the test $\mu = 0.1$ and five for the number of lags in the variance of the DM test.

Figure 1.8: Heatmap of coefficient estimates using LASSO



(a) Day 8

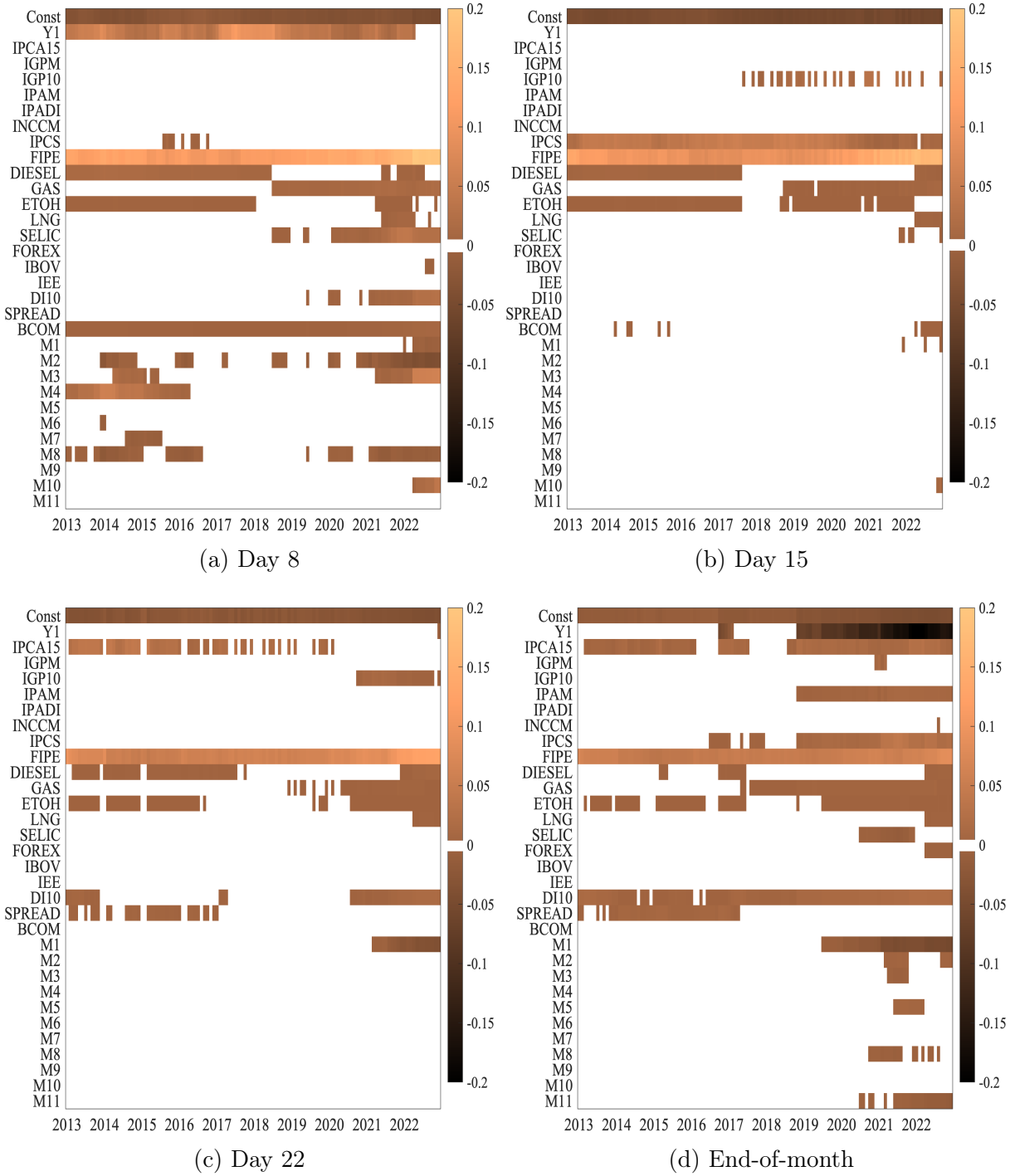
(b) Day 15

(c) Day 22

(d) End-of-month

Notes: This Figure depicts heatmaps of LASSO-fitted coefficients over the evaluation period. Empty cells represent a coefficient estimate equal to zero, and thus a predictor that has not been selected at the estimation round t in the evaluation period.

Figure 1.9: Heatmap of coefficient estimates using LASSO on the SPF nowcasting errors



(a) Day 8

(b) Day 15

(c) Day 22

(d) End-of-month

Notes: This Figure depicts heatmaps of LASSO-fitted coefficients using SPF nowcasting errors as the dependent variable. Empty cells represent a coefficient estimate equal to zero, and thus a predictor that has not been selected at the estimation round t in the evaluation period.

Chapter 2

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

Abstract

We study how millions of highly granular and weekly household scanner data combined with novel machine learning techniques can help to improve the nowcast of monthly German inflation in real-time. Our nowcasting exercise targets three hierarchy levels of the official consumer price index. First, we construct a large set of weekly scanner-based price indices at the lowest aggregation level underlying official German inflation, such as those of butter and coffee beans. We show that these indices track their official counterparts extremely well. Within a mixed-frequency modeling framework, we also demonstrate that these scanner-based price indices improve inflation nowcasts at this very narrow level, notably already after the first seven days of a month. Second, we apply shrinkage estimators to exploit the large set of scanner-based price indices in nowcasting product groups such as processed and unprocessed food. This yields substantial predictive gains compared to a time series benchmark model. Finally, we nowcast headline inflation. Adding high-frequency information on energy and travel services, we construct highly competitive nowcasting models that are on par with, or even outperform, survey-based inflation expectations that are notoriously difficult to beat.

Keywords: inflation nowcasting, scanner price data, machine learning, mixed-frequency modeling.

JEL classification: E31, C55, E37, C53.

This study is coauthored by Günter W. Beck, Kai Carstensen, Jan-Oliver Menz and Elisabeth Wieland.

It is published as the Deutsche Bundesbank Discussion Paper No 34/2023.

2.1 Introduction

The economic shock induced by the COVID-19 pandemic posed unprecedented challenges to policymakers and triggered an enormous demand for reliable real-time information about the state of the economy, including inflation dynamics. A similarly strong need for a timely and continuous flow of information about ongoing price developments arose after Russia's invasion of Ukraine in February 2022, when an immense inflationary wave started to unfold. Whereas official macroeconomic statistics are only available with a certain time lag and at fixed intervals, non-traditional high-frequency data such as web scraping and scanner data in combination with machine learning (ML) techniques represent a promising toolkit for policymakers to monitor ongoing and potentially disrupting developments in real time and to make better-informed decisions in such situations (Tissot and de Beer, 2020, Doerr et al., 2021).

The usefulness of real-time information is not constrained to extraordinary periods, however, as it allows more generally for a faster identification of the current state of the economy, thereby enabling a timelier and more targeted response by policymakers. It can also aid in quantifying the impact of policy measures more precisely (Buda et al., 2023) and possibly adjusting them more swiftly. Furthermore, in the absence of timely official data, there is a risk that economic agents, by basing their decisions on private data sources, may amplify idiosyncratic shocks. Having high-frequency data at hand, policymakers can dampen such effects via a regular data-driven communication strategy (see, e.g., Assenmacher et al., 2021).

The aim of this paper is to analyze how the combination of non-standard high-frequency price data with state-of-the-art machine learning methods helps to nowcast inflation in real time. More specifically, we demonstrate that weekly household scanner data improve inflation nowcasts, both at a very high level of disaggregation and for major product groups and headline inflation. Our data stem from the household panel of the market research company GfK and contain daily purchases of fast-moving consumer goods (henceforth denoted as GFK:FMCG) at the barcode level, primarily covering food, beverages, personal and household care products. Using the comprehensive product descriptions available in the dataset, we transform the highly granular daily price information into weekly price indices that closely match the official price indices at the most disaggregate level used in the German consumption basket, the so-called COICOP-10 item level.[1] With the help of a recursive out-of-sample nowcast experiment, we then document that the application of mixed data sampling (MIDAS) and machine learning techniques to these data (combined with official monthly inflation series and some complementary high-frequency data) yields highly informative nowcasts as soon as after only seven days of a month. Not surprisingly,

[1] The COICOP-10 item level provides a considerably more detailed disaggregation than the COICOP-5 level used for the classification of euro area inflation.

the nowcast accuracy increases with the number of days included. After 14 and particularly after 21 days, the headline inflation nowcasts even outperform standard surveys of market expectations that are notoriously difficult to beat.

Our nowcasting exercise proceeds in three steps that relate to the three hierarchy levels of the German consumption basket our project focuses on: highly disaggregate COICOP-10 items, product groups, and headline inflation. We start at the COICOP-10 level and construct weekly price indices from the granular GFK:FMCG data with the help of time-product dummy regressions used successfully both in the literature (de Haan et al., 2021) and at statistical offices (Eurostat, 2022). For each COICOP-10 item, we then specify a U-MIDAS model (Ghysels et al., 2004) that uses the weekly index to predict, on days 7, 14, 21, and 28 of a month, its official counterpart, which is measured at a monthly frequency. We document that this approach reduces the nowcast error substantially relative to a univariate time series benchmark model. The advantage is particularly pronounced for COICOP-10 items classified as unprocessed fruit and vegetables and dairy products and fat, for which we achieve root mean squared error (RMSE) reductions in the range of 40%-60%. Importantly, large nowcasting gains accrue even if only the scanner data of the first seven days of a month are included.

In the second step, we focus on product groups such as unprocessed and processed food that are regularly monitored by policymakers and market participants and for which we have weekly scanner data information available. As these product groups consist of large numbers of COICOP-10 items, direct nowcasting models of group-specific inflation rates need to efficiently integrate a multitude of weekly GFK:FMCG price indices, which is why we resort to shrinkage estimators from the machine learning toolkit. We find that, compared to a standard time series benchmark model, the group-specific inflation nowcasts of both unprocessed food and processed food benefit considerably from adding the weekly information. Specifically, we document reductions in the relative RMSE of up to 25%. When considering more disaggregate product groups, the relative forecasting gains are particularly large for dairy products and fat (reduction in RMSE of roughly 45% to 55%), unprocessed fruit and vegetables (reduction of around 20% to almost 40%), processed meat and fish (reduction of more than 25% to almost 40%), and unprocessed meat, fish and eggs (reduction of nearly 20% to 25%). Again, the superiority of our ML-based approach becomes apparent as early as after day 7 of a month.

In the final step, we construct nowcasts of headline inflation. To this end, we split the German consumption basket into six components – unprocessed food, processed food, energy, package holidays, non-energy industrial goods (NEIG), and services – which we consider separately. Specifically, we fit a mixed-frequency machine learning model directly to the monthly inflation rate of each component, produce nowcasts, and aggregate the nowcasts to headline inflation by applying the official HICP weighting scheme. To the set

of weekly predictors, we add price quotes for energy and package holidays, which are two of the most volatile and difficult-to-predict inflation components of the German HICP. The resulting headline inflation nowcasts consistently outperform not only a time series benchmark approach but also Bloomberg market expectations if at least information up to day 14 is included.

To study the merits of aggregate versus disaggregate inflation nowcasting, we supplement the direct machine learning models with a bottom-up nowcasting approach which works as follows. For COICOP-10 items matched by a weekly predictor, we apply the U-MIDAS model discussed in the first step. For the remaining COICOP-10 items, we fit a time series benchmark model. We then aggregate the item-level inflation nowcasts to headline inflation nowcasts by the official HICP weighting scheme. We document that this bottom-up approach is slightly outperformed by the direct machine learning approach in normal periods but dominates it in turbulent times. In addition, it has proved highly competitive compared to market expectations, even during the inflation hike of 2022. From this, we conclude that in terms of inflation nowcasting, direct machine learning models are difficult to beat in normal times but do not necessarily adapt quickly enough to large shocks. Overall, this suggests that there is not a single nowcasting method which uniformly outperforms all its competitors. Rather, it is the careful integration of informative high-frequency data into nowcasting models that makes a difference.

The outline of this paper is as follows. Section 2.2 discusses related research and emphasizes our contribution in some detail. Section 2.3 explains the high-frequency scanner data and how we derive price indices that mirror the official indices published by the statistical office. In Section 2.4, we describe our nowcasting strategy and in Section 2.5, we report the results. Robustness checks are presented in Section 2.6, while Section 2.7 concludes.

2.2 Literature review

Our paper relates to several strands of the literature. First, the use of scanner data for economic research can be dated back to as early as the 1970s. In an excellent survey article, Dubois et al. (2022) report how both household and retail scanner data have been fruitfully exploited in research on firm and consumer behavior. Within the area of consumer prices, scanner data have been used to compute household-specific inflation (Kaplan and Schulhofer-Wohl, 2017; Jaravel, 2019), assesses inequality across countries (Beck and Jaravel, 2021), study price-setting strategies of firms (Butters et al., 2022; Karadi et al., 2023), estimate price elasticities of consumer demand (Beck and Lein, 2020), track the effects of the COVID-19-related lockdown on prices and product variety (Jaravel and O'Connell, 2020), and measure the cross-border effects on prices within the

euro area (Messner et al., 2023). In addition, scanner data supplemented with survey questionnaires (D'Acunto et al., 2021) and survey-based information treatments (Weber et al., 2022) can help to reveal the manner in which the way inflation expectations are formed and affect spending plans. Our paper contributes to this literature by showing that household scanner data can also successfully be employed to nowcast headline inflation.

Second, our paper relates to a burgeoning body of literature that was spurred on the COVID-19 pandemic and seeks to construct high-frequency measures of existing low-frequency macroeconomic series. Examples include real-time indicators of house prices (Anenberg and Laufer, 2017), GDP (Eraslan and Götz, 2021) and – using individual bank account information – private consumption (Buda et al., 2022). Regarding inflation, the billion prices project initiated by Cavallo and Rigobon (2016) has shown that online prices can successfully be used to build a high-frequency price index that closely mirrors the official headline inflation published by statistical offices. In addition, using household scanner data for the UK, Jaravel and O'Connell (2020) provide a measure of food price inflation that closely tracks the official number at least on an annual frequency during the COVID-19 period. Studies that draw on web-scraped data to examine price effects in this period include Watanabe (2020), Cavallo and Kryvtsov (2023) and Stelmasiak et al. (2023). Non-standard, high-frequency data have also turned out to be useful in the study of price and consumption effects following natural disasters as, e.g., in Cavallo et al. (2014), Gagnon and López-Salido (2019) or Watanabe (2020). Furthermore, using household scanner data for the UK, Jaravel and O'Connell (2020) provide a measure of food price inflation that closely tracks the official number at least on an annual frequency, and Alvarez and Lein (2020) offer an online inflation measure for Switzerland by combining web-scraped prices with consumption weights derived from debit card transactions. We add to this literature by using scanner data to construct weekly inflation measures for unprocessed food, processed food, and non-durable goods, both on the aggregate level and for detailed subcomponents such as "butter", "coffee beans", and "sanitary cleaner". Since our data also contains information on quantities, we are able to use real-time weights in the construction of price indices. The availability of such detailed inflation series is particularly useful in times of crises, as we have illustrated with respect to the price development of selected food items following Russia's invasion of Ukraine (Beck et al., 2022).

Third, we contribute to the literature on nowcasting key macroeconomic variables. Typically, research in this field has focused on providing monthly GDP estimates; however, fuelled by the COVID-19 pandemic and the war in Ukraine, there has been increased interest in producing high-frequency measures of monthly data, such as inflation. Some studies in this field use traditional data sources such as weekly gasoline or commodity prices (Modugno, 2013; Breitung and Roling, 2015; Knotek and Zaman, 2017; Clark et al.,

2022; Aliaj et al., 2023) and report robust forecasting and nowcasting gains compared to econometric benchmark models or market expectations. Another branch of the literature uses web-scraped price data to predict aggregate and disaggregate food price inflation (Macias et al., 2023; Powell et al., 2018) and headline inflation (Harchaoui and Janssen, 2018; Aparicio and Bertolotto, 2020), again documenting improved forecasting accuracy. We add to this literature by showing that scanner data is a very promising candidate for the nowcasting of inflation at both the aggregate and the disaggregate level.

Fourth, our paper relates to an earlier body of literature that addressed the question of whether it would pay off to forecast headline inflation by explicitly using subcomponents or even the full breakdown of the inflation rate as inputs (Hendry and Hubrich, 2011; Ibarra, 2012; Espasa and Mayo-Burgos, 2013; Bermingham and D'Agostino, 2013) Generally, the studies show that it does indeed help to take disaggregate information into account. Recently, this line of research has been taken up again by Joseph et al. (2022) who use disaggregate inflation data combined with machine learning methods to forecast headline inflation in the UK. Related to this academic debate, central banks have always forecasted different components of the inflation rate, both for statistical reasons and for understanding the underlying price dynamics (Benalal et al., 2004; Capistrán et al., 2010; Huwiler and Kaufmann, 2013; Giannone et al., 2014; Ulgazi and Vertier, 2022). By using the German inflation rate which is the one with the most detailed breakdown worldwide, we show that combining disaggregate inflation nowcasts into an aggregate nowcast for headline inflation is a highly competitive approach.

From a methodological perspective, our paper relates to recent advances in machine learning that seek to improve inflation forecasts by exploiting large datasets. While this literature dates back at least to Stock and Watson (1999), simple univariate models have been found to be very difficult to beat (Atkeson and Ohanian, 2001; Stock and Watson, 2007). Recent results are more promising. In particular, Garcia et al. (2017), Medeiros et al. (2021) and Babii et al. (2021) apply a multitude of machine learning techniques to large macroeconomic datasets and demonstrate that this can lead to notable forecasting gains (see also Paranhos, 2021; Li et al., 2022; Goulet Coulombe et al., 2022; Hauzenberger et al., 2023). In a similar vein, Joseph et al. (2022), Botha et al. (2022), and Barkan et al. (2022) show that inflation forecasts can also benefit from applying machine learning to large sets of highly disaggregate price indices. Our findings add to this body of research by revealing that machine learning tools provide an effective solution for handling a large set of disaggregate price series in a mixed-frequency setting. In particular, we show that it pays off to combine shrinkage methods with the weekly GFK:FMCG dataset to produce higher quality nowcasts for major product groups, such as for unprocessed and processed food, and headline inflation.

Finally, our approach of nowcasting inflation at a very disaggregate level with the help

of machine learning models also helps to produce robust results during severe crises. As discussed by Bańbura et al. (2023), adjusting standard forecasting models to cope with the effects of large shocks is no mean feat. As we will show, modeling inflation at a disaggregate level helps to take into account very specific policy measures taken during the pandemic and the recent energy price hike, both of which only affect a limited number of disaggregate price series.

2.3 Data

We base our nowcasts on a weekly dataset of consumer prices from three different sources. Most importantly, we use household scanner data for daily purchases of fast-moving consumer goods, mainly consisting of food items and non-durable goods. In addition, we collect weekly energy prices from the European Commission and daily transaction data for travel services from a private travel booking system provider. We combine this high-frequency dataset with monthly inflation rates compiled by the German statistical office from 2003 until 2022. This section provides details about the data sources and the data transformations necessary to construct high-frequency price indices that match the official HICP series as closely as possible.[2]

2.3.1 Fast-moving consumer goods (GFK:FMCG)

The most comprehensive part of our dataset consists of home scan purchases of fast-moving consumer goods (FMCG) collected from private households by the market research company GfK. It records, at a daily frequency, the purchases of around 30,000 households which constitute a representative sample of the German household population, and mainly includes food, beverages, and personal care items. The dataset starts in January 2003 and covers around 200,000 different products and an average of around 30 million observations per year. Table (2.1) provides an illustration of the structure of the GFK:FMCG data using "butter" as an example. In addition to the price paid on a particular day and the barcode of a single product, the dataset contains detailed information about products and retailers.

For the purpose of our study, this dataset offers several advantages. First, scanner prices stem from actual transactions and should thus closely co-move with official prices sampled from a representative product bundle by the statistical office. Second, information on prices and purchases are recorded by shopping day, i.e., it can be used to construct high-frequency series. The data is also well-suited for nowcasting purposes as weekly updates are already made available by GfK on the following Monday. Third, the dataset includes

[2] A list of all data sources used is provided in Table 2.6 of Appendix 2.A.1.

Table 2.1: Illustrative example of the household scanner data (GFK:FMCG)

GfK Household Panel							Mapping to COICOP Classification	
Household	Product Description	Barcode	Quantities	Sales	Retailer	Purchase Date	10-digit Code	Product-Category
1	Green Hill Butter 250g	400123123123	1	3.39 €	A	28.11.2022	0115100100	Butter
2	Lovely Butter 250g	400456456456	2	6.58 €	B	01.12.2022	0115100100	Butter
3	Lovely Butter 250g	400456456456	1	3.39 €	C	01.12.2022	0115100100	Butter
4	Green Hill Butter 250g	400123123123	1	3.29 €	B	02.12.2022	0115100100	Butter
5	Green Hill Butter 250g	400123123123	2	6.98 €	A	03.12.2022	0115100100	Butter
6	Sunny Sunflower oil 1l	100445566123	1	2.29 €	B	01.12.2022	0115400100	Sunflower oil
7	Blossom Sunflower oil 1l	100112233123	1	3.99 €	C	01.12.2022	0115400100	Sunflower oil
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Note: Fictitious entries.

quantities, i.e., the number of units bought of a particular product. This information is important as we can use it to construct consumption weights that increase the correlation with the official price indices. Fourth, information is provided at the barcode level, which allows us to take into account composition effects.[3] For example, if a particular brand of butter is purchased more frequently, it will receive a higher weight in the construction of the scanner-based price index of butter. While this effect might temporarily lead to deviations from the official price index, which is based on a consumption basket that is fixed in the short run, official weights are adjusted regularly and thus should converge with those derived from actual shopping behavior.

In our empirical exercise, we construct nowcasts both for headline inflation as well as for some product groups such as "unprocessed food", "processed food", or "non-durable goods', and, if possible, for the most detailed inflation series available. According to the COICOP system,[4] the German inflation rate can be broken down into different categories, ranging from goods, services and energy to more detailed components such as "vegetables" or "mineral water". In the euro area, the most detailed harmonized breakdown of the inflation rate is the so-called COICOP-5 level, whereas in Germany, inflation can be disaggregated further into the COICOP-10 level,[5] which, however, poses some challenges. First, the COICOP-10 series are compiled within the system of the national CPI and not the HICP. Hence, methodological differences between both concepts have to be taken into account when aggregating inflation nowcasts from the bottom up.[6] Second, the CPI is typically revised every five years with the introduction of a new base year by including new product groups and removing outdated ones. Therefore, some COICOP-10 series are only available on a shorter time period, which we try to overcome by extending them backwards using corresponding price indices from the previous base years.

Keeping this in mind, we match the products of the scanner data to the corresponding

[3] Mostly, the barcode is given by the "Global Trade Identification Number" (GTIN), whereas GfK assigns a unique ID to products such as fresh food or private labels for which no GTIN is available.

[4] COICOP stands for "Classification of Individual Consumption by Purpose"; see Eurostat (2018) for details.

[5] For example, the COICOP-5 component "cheese and curd" is decomposed into the five COICOP-10 groups "hard cheese", "sliced cheese", "soft cheese", "curd" and "cream cheese". Indices computed at this lowest-available level are generally denoted as "elementary indices"; see IMF, ILO, OECD, Eurostat, UNECE, and The World Bank (2020), Chapter 1.

[6] The HICP is a chain-linked price index where weights are updated each year whereas the CPI is a fixed-base index where weights are updated only every five years. Moreover, the CPI includes prices for gambling and owner-occupied housing. Despite their methodological differences, headline rates of the HICP and CPI behave rather similarly over time.

COICOP-10 items, making use both of the product categories and the detailed product descriptions included in the GfK household panel as well as the item descriptions contained in the HICP manual.

Based on the mapped scanner data, we proceed as follows. First, in each COICOP-10 component, the raw price $p_{i,d}$ of a given product i bought on a particular day d is defined as:

$$p_{i,d}^{raw} = \frac{sales_{i,d}}{unit_{i,d}}, \tag{2.1}$$

where $sales_{i,d}$ are the total expenses in euro for a given item and $unit_{i,d}$ denotes the number of items bought. By this, we obtain for each item and time period a sample of unit-value observations. Moreover, we omit outliers which are below and above the 1st and 99th percentiles of the price distribution at the COICOP-10 level. Due to the large number of households in the dataset, a specific item is often bought several times per day, in which case we compute the average price per item and day. Finally, we transform the data from daily into weekly frequency by defining four weeks per month such that the first week consists of day 1 until day 7, the second week from day 8 to 14, the third week from day 15 to 21 and the fourth week from day 22 to day 28. By using only 28 days of a given month, we tackle the problem of a shorter February and leap years.

We compile scanner-based price indices at the COICOP-10 level by running weighted time-product dummy (TPD) regressions. This method, proposed by Diewert (2005), is widely used in official price statistics to construct price indices from scanner or web-scraped data (de Haan et al., 2021; Eurostat, 2022). Specifically, for each week $t = 0, \ldots, T$ and product $i = 1, \ldots, N$, we fit the equation

$$\ln p_{i,t} = \beta^0 + \sum_{\tau=1}^{T} \delta^\tau d_{i,t}^\tau + \sum_{j=1}^{N-1} \gamma^j D_i^j + \varepsilon_{i,t}, \tag{2.2}$$

where $D_{i,t}^j$ represents a product dummy which takes the value 1 if $i = j$ (as identified by its barcode) and 0 otherwise, and $d_{i,t}^\tau$ denotes a time dummy which takes the value 1 if $t = \tau$ and 0 otherwise. Weights are given by the total expenses, $sales_{i,t}$, for a given product. This increases the price effect of popular products compared to those that are bought less frequently. Note that missing prices for a given item are treated in a similar way as in official price statistics: if an item is only temporarily missing, its last price is carried forward up to eight weeks, before it is replaced with another product.

For each week $t = 0, ..., T$, we estimate a price index from the exponential of the coefficient on the respective time dummy, such that:

$$I_{TPD}^{0,t} = 100 \times \exp\left(\hat{\delta}^t\right). \tag{2.3}$$

To mimic a real-time compilation of scanner-based price indices, we estimate equation (2.2) on a rolling window of 49 weeks, which covers at least one full year of scanner data (i.e., one month consisting of four weeks only). For example, the first estimation window covers period 1 to period 49 (providing a price index of the same length), the second estimation window period 2 to period 50, and so on. The linking of this sequence of 49-period price indices is performed in the spirit of a mean splice. By linking subsequent index values to the existing one, this yields a non-revisable real-time price index such that:

$$I_{TPD}^{0,t} = \prod_{k=t-\lambda}^{t-1} \left(I_{TPD}^{0,k} \times I_{[t-w+1,t]}^{k,t} \right)^{\frac{1}{\lambda}}, \tag{2.4}$$

where w denotes the window size (49 weeks) and λ is an overlapping linking period, which we set to eight weeks.[7] Hence, in our example above, the index in period 50 is obtained as a geometric average of the pairwise changes of the period-50 index value to the index values in periods 42 to 49 obtained from the second estimation window, each multiplied by the index value of the corresponding overlapping period from the first estimation window.

2.3.2 Energy and travel services

Unexpected price changes for energy and package holidays regularly contribute strongly to forecast errors of German headline inflation. Therefore, we also match these components with high-frequency information available in almost real-time.

The energy component of the HICP consists of 14 price indices at the COICOP-10 level[8] that we try to match with the high-frequency price indicators discussed in the following. Most importantly, we use weekly price data from the Weekly Oil Bulletin (WOB) by the European Commission that has proven invaluable in previous work (Modugno, 2013; Aliaj et al., 2023). The WOB contains weekly information starting in 2005 about average fuel prices at the pump (Diesel, Supergrade petrol) and household-size deliveries of heating oil. The prices include duties and taxes and are thus much more closely related to consumer price indices than futures prices traded on financial markets, as used by Breitung and Roling (2015) and Knotek and Zaman (2017). Nevertheless, we include the European

[7] In the context of a rolling time window, a mean splice includes all overlapping periods as a linking period (Eurostat, 2022). However, in our weekly application, this yields 48 overlapping periods. We therefore opted to link over a shorter period to ease the computational burden. We also computed alternative price indicators based on different index concepts and splicing methods. Overall, the TPD model with splicing over eight weeks gave the best results in terms of in-sample correlations.

[8] These items are "Electricity" (COICOP no. 0451010000), "Natural gas, excl. share in the costs" (0452103000), "Share in the costs for gas central heating" (0452105100), "Liquefied gas, charging of a tank container" (0452200200), "Heating oil" (0453001100), "Share in the costs for oil central heating" (0453005100), "Coal briquettes" (0454100200), "Firewood, wood pellets or the like" (0454900100), "District heating" (0455002200), "Diesel fuel, cetane number below 60" (0722100100), "Diesel fuel, cetane number 60 and more" (0722100300), "Supergrade petrol, 95 octane" (0722201100), "Supergrade petrol, 98 octane and more" (0722204300), and "Liquefied petroleum gas" (0722301100).

Gas Spot Index (EGSI) as we have no other high-frequency information for the household gas supply.

We also use a few more daily energy price series from various sources as described in Table 2.6 of Appendix 2.A.1. All German gasoline stations have to report their intraday price changes of super and diesel fuels to the Market Transparency Unit for Fuels. We access this database via the data provider "Tankerkoenig" and take unweighted averages of all prices reported within a day. We also include daily measures of heating oil and wood pellet prices. We turn all daily series into weekly by taking an unweighted average of days 1-7 (week 1), 8-14 (week 2), 15-21 (week 3), and 22-28 (week 4).

Regarding travel services, we use high-frequency data from the travel booking system provider Amadeus that include transaction prices for package holidays, which is defined as a combination of flight and accommodation services. Prices for international package holidays represent an important component of the German HICP because of their relatively large weight in the consumer basket and their relatively high volatility.[9] The Amadeus dataset starts in 2012 and is available at a daily frequency.[10] We provide more details about the construction of the high-frequency index for package holidays in Section 2.7 of Appendix 2.A.3. Basically, the index is compiled by constructing a weighted average of the prices for the most relevant travel destinations and that focuses on last-minute bookings up to 14 days before the travel date.

2.3.3 Descriptive statistics

Table 2.2 provides an overview of the HICP components for which we are able to construct a corresponding high-frequency price index using household scanner data for the full sample 2003 to 2022. In total, we cover about 12% of the basket underlying the German headline inflation rate, including the vast majority of unprocessed and processed food items.[11] With regard to non-energy industrial goods (NEIG), we match a significant proportion of non-durables but are unable to do so for most of the semi-durable and durable goods. Except for package holidays, our dataset does not comprise high-frequency prices for services which make up about 50% of the German consumer basket. However, service prices are either determined by administrative measures, such as insurance or tuition fees, or are rather sticky (see Gautier et al., 2023). Therefore, we do not expect the lack of high-frequency prices for services to impair our nowcasting results too strongly.

In Figure 2.1, we compare the year-over-year inflation rates of the HICP with its scanner-based counterparts by considering the product groups "food" and "non-durable

[9] In 2023, package holidays make up 3.5% of the German inflation rate, compared to 0.2% in France and 0.5% in Italy.

[10] The dataset has been used by Henn et al. (2019) for price measurement and by Nagengast et al. (2021) to estimate the exchange rate pass-through.

[11] Tobacco products are not available in the scanner data set since households are typically reluctant to reveal reliable information about their actual purchases.

Table 2.2: Mapping between high-frequency price data and the German HICP

	HICP		Scanner data	
Component	COICOP-10s	Weight	COICOP-10s	Weight
Unprocessed food	38	2.4	30	2.0
Fruit	8	0.7	6	0.5
Vegetables	11	0.7	9	0.6
Meat & eggs	15	0.9	15	0.9
Fish	4	0.1	0	-
Processed food	142	11.1	116	8.1
Fruit	7	0.2	5	0.1
Vegetables	12	0.4	11	0.4
Meat	13	1.1	11	0.9
Fish	7	0.2	4	0.1
Bread & cereals	25	1.5	23	1.4
Dairy products & fat	18	1.5	14	1.4
Beverages	29	2.9	23	2.7
Other food products	28	1.2	25	1.0
Tobacco	3	2.1	0	-
NEIG	302	23.0	39	1.8
Non-durables	75	5.9	36	1.8
Semi-durables	139	8.7	3	0.1
Durables	88	8.4	0	-
Total HICP	482	36.5	185	11.9

Note: The table reports the number of COICOP-10 price indices and the associated weights of different HICP components in addition to the number of series and the share for which we have household scanner data available over the full sample 2003 to 2022. Weights refer to CPI expenditure shares of the base year 2020.

goods".[12] Note that the HICP aggregates are derived from the underlying COICOP-10 series using only those series for which we also have high-frequency data available. While this distinction does not matter for food, energy and package holidays, it is relevant for non-durable goods given that we only cover about 40% of this aggregate.

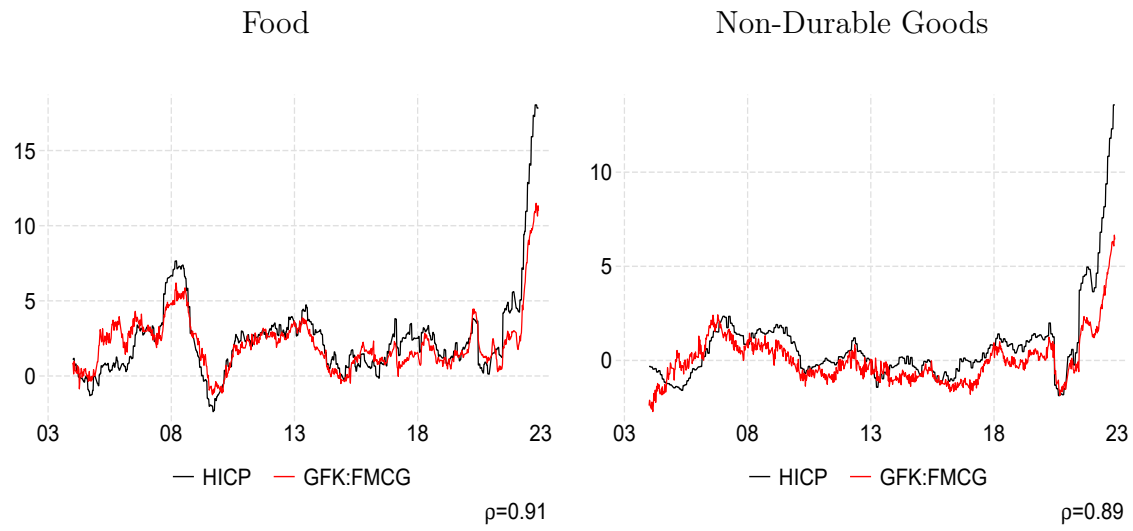
Overall, the co-movement between the weekly scanner and the monthly official inflation rates is very high. For food, non-durable consumer goods and energy, the correlation coefficient is about 0.9. Even if we use the HICP component for non-durable goods based on the full set of the underlying COICOP-10 series, the correlation still exceeds 0.5.[13]

It is worth noting that this high correlation of the aggregate inflation series masks

[12] The graphs for "energy" and "package holidays" are shown in Figure 2.6 of Appendix 2.A.2. An online appendix to this paper (available upon request) plots the high-frequency inflation rates together with the official rates for all of the about 180 product groups which could be mapped to scanner data.

[13] Correlations are computed by carrying forward the monthly inflation rates to each of the four weeks per month. This approach mirrors our forecasting setup more closely than aggregating the weekly series into the monthly frequency in the first step.

Figure 2.1: German HICP inflation and high-frequency scanner counterparts



Note: The figure shows year-over-year inflation rates (% change) for HICP subcomponents aggregated using all of the corresponding COICOP 10-digit level series for which we have high-frequency data available. "HICP" refers to the aggregates using official COICOP 10-digit series, and "GFK:FMCG" refers to the scanner data for fast-moving consumer goods. ρ reports the correlation coefficient between both series.

some considerable heterogeneity at the COICOP-10 level. This is illustrated for a selected number of product groups in Figure 2.7 that highlight some general patterns.[14] First, as the indices for "cucumbers", "cherries" and "whole milk" illustrate, official inflation rates tend to exhibit larger fluctuations compared to their high-frequency counterpart. This could be due to the fact that the statistical office samples only a subset of products, whereas scanner-based indices use all products being bought by households, which might smooth out extreme price changes. On the other hand, we also observe that for other COICOP-10 components indices of high-frequency scanner data are more volatile than the official series. This is, e.g., the case for "multivitamin juice" or "salt". This could stem from sales or special offers that are more likely to be included in the high-frequency scanner dataset rather than the monthly HICP data. Finally, towards the end of the sample, official inflation series have increased much more than their high-frequency counterpart. This points to substitution effects that arise if households switch from expensive products to cheaper ones.

2.4 Nowcasting strategy

Our nowcasting exercise proceeds in three steps. In the first step, we evaluate at the highly disaggregate COICOP-10 item level how well the weekly GFK:FMCG price indices

[14]The graphs for all COICOP-10 components are provided in an online appendix to this paper (available upon request).

predict their current month's official HICP counterparts. For that purpose, we employ the well-established MIDAS framework advocated by Ghysels et al. (2004) that takes into account the different frequencies of the data.[15] We compare this approach to an autoregressive benchmark with seasonal dummy variables. We are thereby able to quantify the informational advantage of adding weekly scanner data to nowcast monthly inflation at the most granular level.

In the second step, we nowcast the inflation rates of product groups of the HICP consumption basket that are closely monitored by central banks, market observers, and business cycle experts and for which we have disaggregate weekly GFK:FMCG price information available. The most important of these groups are unprocessed food, processed food, and non-energy industrial goods. As they consist of large numbers of disaggregates – for example, unprocessed food is an aggregate of 142 COICOP-10 items – we implement machine learning techniques like the elastic net and the sparse group LASSO to achieve dimensionality reduction via shrinkage and produce direct inflation nowcasts at the product group level in a high-dimensional mixed-frequency environment. Again, our benchmark is a standard autoregressive model augmented with seasonal dummies. The setup allows us to study the extent to which the combination of high-frequency external information and up-to-date machine learning techniques can improve inflation nowcasts.

In the third step, we construct nowcasts of headline inflation. Here, we face the obstacle that the GFK:FMCG data cover only a part of the HICP consumption basket. Therefore, we add high-frequency data on price developments of energy and package holidays, which for Germany are the two most volatile and difficult-to-predict inflation components omitted so far. Rather than attempting to directly relate headline inflation to the multitude of dissaggregate information compiled, we use the product-group approach also applied in the previous step. Specifically, we target the following six components that make up to headline inflation: unprocessed food, processed food, energy, package holidays, non-energy industrial goods (NEIG), and services. For each of these components, we fit a machine learning model that includes all relevant disaggregate weekly information. We then compute headline nowcasts from component nowcasts by applying the official HICP weighting scheme. We complement the machine learning strategy with a bottom-up approach that generates inflation nowcasts exclusively at the COICOP-10 level as described in the first step and subsequently aggregates them, again using the official HICP weights. In this setup, we predict COICOP-10 items not matched by high-frequency covariates with the help of the seasonal-dummy autoregressive benchmark model. We compare these two approaches to market expectations, which have been shown by Bańbura et al. (2023) to be a very challenging benchmark.

[15] Alternatively, we could have used a dynamic factor model or mixed-frequency VAR approach estimated by means of a Kalman filter or Bayesian methods (see, e.g., Modugno, 2013; Cimadomo et al., 2022); however, in our setting, the much simpler MIDAS framework is sufficient.

We conduct a recursive out-of-sample nowcast experiment that covers the period of January 2016 to December 2022. This evaluation sample is dictated by data availability as quite a few official COICOP-10 price indices start only in January 2015. Nevertheless, it includes both the COVID-19 crisis and the recent surge in inflation, arguably not only two of the most challenging periods for inflation forecasting in recent history, but also especially hard for statistical models that are fitted to the data in normal times and do not incorporate all the structural information professional forecasters took into account. The following paragraphs describe our nowcasting strategy in more detail.

2.4.1 Benchmark nowcasts

We define the month-over-month inflation rate as $\pi^M_{c,t} = 100 \times (P_{c,t}/P_{c,t-1} - 1)$, and the year-over-year inflation rate as $\pi^A_{c,t} = 100 \times (P_{c,t}/P_{c,t-12} - 1)$, where $P_{c,t}$ denotes the HICP index of the COICOP-10 item $c = 1, \ldots, 644$ in month t. Then, we compute model-based benchmark nowcasts from a seasonal dummy autoregressive (SD-AR) model of the following form:

$$\pi^M_{c,t} = \alpha_0 + \sum_{j=1}^{p} \rho_j \, \pi^M_{c,t-j} + \sum_{s=1}^{13} \gamma_{c,s} \, d_{c,s,t} + \varepsilon_{c,t}, \tag{2.5}$$

Since we work with data that are not adjusted for seasonal effects, we augment the AR model with 11 monthly dummies $d_{c,1,t}, \ldots, d_{c,11,t}$, defining December as the reference case. We also add an Easter dummy, $d_{c,12,t}$, and a Pentecost dummy, $d_{c,13,t}$, to capture the specifics of the German public holiday season, which in turn affect the consumption patterns of German households.[16] Throughout the paper, we use the year-over-year inflation rate as our target variable; hence, we transform the resulting nowcasts for the month-over-month rates accordingly. The model specification is, in each estimation step, guided by the Bayesian Information Criterion (BIC). This includes the autoregressive lag order $p \in \{1, \ldots, 12\}$ and the decision of whether or not to include the dummies.[17]

2.4.2 Step 1: nowcasting item-level inflation rates

For each COICOP-10 item c for which we have weekly GFK:FMCG data available, we estimate a MIDAS model of the monthly HICP inflation rate, $\pi^M_{c,t}$, to account for the

[16] The Easter dummy measures how many days of the two Easter weeks are in March and April, while the Pentecost dummy measures how many of the three Pentecost days (Saturday to Monday) are in May and June.

[17] We also use a local mean model that consists only of an intercept estimated from the last twelve observations to accommodate the specifics of some administered prices. In addition, for seven administered price indices of various medical and veterinary services, driving license fees and motor vehicle registration fees we do not fit a model to the month-over-month inflation rate as these indices are known to change only rarely and then stay constant for years. Instead, we use a random walk forecast of the level. In fact, we have evaluated more generally the forecast accuracy of AR models estimated in year-over-year rates as well as random walk models specified both in price levels and inflation rates, but have found that – with the exception of the seven cases just listed – the SD-AR model yields the best overall results for the COICOP-10 items, the subcomponents, and headline inflation.

mixed-frequency environment. We use the baseline specification

$$\phi(L)\,\pi^{M}_{c,t+h} = \alpha_{0,h} + \beta_{c,h}\,B(L^{1/m};\theta)\,x^{(m)}_{c,t} + \sum_{i=s}^{13}\gamma_{c,s}\,d_{c,s,t+h} + \varepsilon_{c,t+h}, \tag{2.6}$$

where the subscript $t = 1, \ldots, T$ denotes the monthly time index and the superscript m denotes the high-frequency ratio within a month. The predictors include the weekly GFK:FMCG inflation rate, $x^{(m)}_{c,t}$, sampled four times more frequently than the target variable, and the set of monthly dummy variables, $d_{c,1,t}, \ldots, d_{c,13,t}$, defined above. To account for temporal dependence, we specify an autoregressive polynomial, $\phi(L)$, with lag order $\{1, 12\}$ and a distributed lag polynomial, $B(L^{1/m};\theta)$, that aggregates the m high-frequency lags to the common frequency t.

For each month, we use high-frequency information only from its first four weeks, so that we have a fixed week-to-month ratio of $m = 4$ as required by the MIDAS model. The first week spans the seven initial days of a given month t, the second week includes days 8 to 14, and so on up to day 28. This strategy controls for the problem of overlapping calendar weeks across consecutive months and the heterogeneous number of days in different months. It is also in line with the typical convention in official price statistics according to which prices are primarily collected during the first three to at most four weeks of a month, implying that prices observed after the 28th day of a month hardly enter the HICP.

To integrate weekly observations into a monthly model, we use the following notation. We denote the four weekly observations by $x_t, x_{t-\frac{1}{4}}, x_{t-\frac{2}{4}}, x_{t-\frac{3}{4}}$, where x_t denotes the high-frequency inflation rate of the 4th week in month t over the same week of month $t-1$, $x_{t-\frac{1}{4}}$ refers to the inflation rate measured in the 3rd week and so on up to the 1st week of t.[18] Since the choice of $m = 4$ with a single predictor does not lead to a proliferation of parameters in equation (2.6), we implement the unrestricted MIDAS approach (hereafter *U-MIDAS*) with OLS estimation (see Foroni et al., 2015; Ghysels and Marcellino, 2018, for more details).[19]

How do we deal with the ragged-edge feature of our mixed-frequency dataset? In each month t, we use the four information sets available on days 7, 14, 21 and 28 to nowcast the inflation rate $\pi^{M}_{c,t}$ of some item c which is matched by GFK:FMCG data. It is important to keep in mind that the official COICOP-10 inflation rates are only released with a two-week delay following the end of the reference month. Hence, nowcasts made on day 7 of month t use official inflation rates up to month $t-2$ and an estimate of month $t-1$ derived from the weekly data which we denote by $\hat{\pi}^{M}_{c,t-1}$. They also include the first weekly observation $x_{c,t-\frac{3}{4}}$ and fill in the remaining three weeks with this latest observation available. Nowcasts

[18] See Appendix 2.B.1 for a more formal description of the baseline model (2.6) and its matrix notation.

[19] In our setting, the U-MIDAS approach delivers results that are roughly equivalent to nonlinear choices of the high-frequency aggregation scheme via $B(L^{1/m};\theta)$, such as the exponential Almon lag specification.

made on day 14 use the official inflation rate of the previous month $\pi^M_{c,t-1}$, which at that day has just been published, and add the first two weekly observations $x_{c,t-\frac{3}{4}}$ and $x_{c,t-\frac{2}{4}}$. Again, they use the latter to fill in the two missing weeks. This updating scheme is repeated on days 21 and 28. We list the information sets in the following:

Day 7: $\left(\hat{\pi}^M_{c,t-1}, x_{c,t-\frac{3}{4}}, x_{c,t-\frac{3}{4}}, x_{c,t-\frac{3}{4}}, x_{c,t-\frac{3}{4}}\right)$

Day 14: $\left(\pi^M_{c,t-1}, x_{c,t-\frac{3}{4}}, x_{c,t-\frac{2}{4}}, x_{c,t-\frac{2}{4}}, x_{c,t-\frac{2}{4}}\right)$

Day 21: $\left(\pi^M_{c,t-1}, x_{c,t-\frac{3}{4}}, x_{c,t-\frac{2}{4}}, x_{c,t-\frac{1}{4}}, x_{c,t-\frac{1}{4}}\right)$

Day 28: $\left(\pi^M_{c,t-1}, \underbrace{x_{c,t-\frac{3}{4}}}_{\text{day 7}}, \underbrace{x_{c,t-\frac{2}{4}}}_{\text{day 14}}, \underbrace{x_{c,t-\frac{1}{4}}}_{\text{day 21}}, \underbrace{x_{c,t}}_{\text{day 28}}\right)$

Note that this within-month random-walk update does not require estimation of any parameter and works well in our sample.

2.4.3 Step 2: nowcasting product group-specific inflation

Policymakers, professional forecasters, and market participants regularly monitor the inflation rates of important product groups. Given the restricted coverage of the GFK:FMCG data, we focus on the high-level product groups unprocessed food, processed food and non-energy industrial goods, which receive considerable attention in the Eurosystem, and the more disaggregated low-level product groups unprocessed fruit and vegetables; unprocessed meat, fish and eggs; processed fruit and vegetables; processed meat, fish and eggs; bread and cereals; dairy products and fat; beverages and other food products; and non-durables.

A natural starting point for nowcasting at the product-group level is to treat all COICOP-10 series belonging to that group as relevant predictors. As the U-MIDAS setting is not suited to handling such a large set of predictors, we follow the related literature which successfully applied shrinkage estimators in such data-rich environments (see, for instance, Garcia et al., 2017; Medeiros et al., 2021; Joseph et al., 2022).

We try two modeling approaches. Our first approach avoids mixed frequencies by aggregating the weekly GFK:FMCG indicators, $x^{(m)}_{c,t}$, to the monthly frequency, which yields a single $x^{(M)}_{c,t}$. Then, we apply standard shrinkage methods to estimate nowcasting models of the group-specific target inflation rates.[20] Specifically, we use the least absolute shrinkage and selection operator (LASSO), the ridge and the elastic net estimator.[21] We proceed

[20] Applying penalized U-MIDAS regressions to the large set of predictors defined at the weekly frequency (four weekly series for each predictor) is also feasible; however, this approach does not recognize serial dependence across high-frequency lags and thereby may be subject to random selection. Zhao and Yu (2006) show that LASSO selects the true model consistently if and (almost) only if the irrelevant covariates are not highly correlated with the predictors in the true model ("irrepresentable condition").

[21] We use the elastic net without tuning the relative weights of the L1 and L2 norms. Instead, we impose equal weights.

as follows. Suppose we want to nowcast, on day 14 of month T, the official inflation rate, $\pi^M_{g,T}$, of product group g. We first regress $\pi^M_{g,t}$ on the full set of contemporaneous predictors $x^{(M)}_{c,t}$ belonging to group g using the sample $t = 1, \ldots, T-1$ of all monthly data available on that day. This yields a vector of estimated parameters $\hat{b}_g$. We then construct the day 14 estimates of $x^{(M)}_{c,T}$, $\forall c \in g$, as described in the previous step, substitute them on the right-hand side of the regression equation and compute a nowcast based on the estimated parameters $\hat{b}_g$.

The second approach applies the sparse-group LASSO (sg-LASSO) estimator proposed by Babii et al. (2021) and regresses the group-specific target inflation rate, $\pi^M_{g,t}$, directly on the large set of weekly GFK:FMCG inflation rates using orthogonal Legendre polynomials as the aggregation scheme. This approach has the advantage in that it performs shrinkage in a mixed-frequency rather than a low-frequency setting by recognizing serial dependence across different high-frequency lags, also taking into account the time series nature of the data.[22]

The tuning parameters of the aforementioned machine learning tools are determined in a data-driven manner using cross-validation to obtain optimal prediction performance.[23] Finally, to evaluate the nowcast precision of these machine learning approaches based on weekly GFK:FMCG information, we fit SD-AR benchmark models directly to the group-specific target inflation rates, $\pi^M_{g,t}$, from which we construct time series forecasts.

2.4.4 Step 3: nowcasting headline inflation

To nowcast headline inflation, we split it into the following six components: unprocessed food, processed food, non-energy industrial goods, energy, package holidays, and services. We nowcast each part separately and construct a headline nowcast by using the official HICP weighting scheme.

Bottom-up U-MIDAS approach. Our first nowcasting approach follows a bottom-up strategy. We estimate one nowcasting model for each COICOP-10 item and aggregate the item nowcasts to the six components of headline inflation using the official HICP weights. For those items that are matched by weekly GFK:FMCG data, we use the U-MIDAS model presented in Section 2.4.2. For all other items, except for energy and package holidays, we use the SD-AR model described in Section 2.4.1. This implies that

[22] We use a Legendre polynomial of degree $L = 0$ which attributes equal weights to all high-frequency lags and delivered similar results compared to other choices of L but at a lower computational cost (see Section 2.6). By contrast, the $L = 1$ polynomial leads to an increasing linear function and thereby favors more distant lags, $L = 2$ features higher weights to very recent and more distant lags, and so on. See Appendix 2.B.2 for a more formal description of the machine learning methods.

[23] We tune the hyperparameters of sg-LASSO via expanding cross-validation splitting the in-sample data into $k = 5$ folds and tests on the $k + 1^{\text{th}}$ fold so that it accounts for the time series nature of the data, although it only uses the end of the sample as the test set. Cross-validation of the standard shrinkage methods (LASSO, ridge and elastic net) also uses a training split of $k = 5$ folds but hereby assumes independent and identically distributed samples, which is also valid in a time series context provided the models yield uncorrelated errors (Bergmeir et al., 2018). For a review of these cross-validation methods, see Goulet Coulombe et al. (2022).

the nowcasts for almost all service items are based on the SD-AR model as we do not have any high-frequency indicator for services.

The energy component consists of 14 COICOP-10 items relating to household energy and fuels. Since the weekly energy prices discussed in Section 2.3.2 are not always good counterparts to these items, we proceed as follows. We first select, at each recursion of the nowcasting experiment, the weekly series most strongly correlated with the COICOP-10 item at hand from a relevant subset. We then run a U-MIDAS model including the selected series as predictor.[24]

Package holidays typically receive a weight of around 3%-4% in the overall HICP consumption basket (except in the years 2021 and 2022, when the COVID-19 pandemic reduced it to roughly 1%). Nevertheless, this component accounts for relevant fluctuations in headline inflation due to its high volatility, which is, unsurprisingly, dominated by a strong seasonal pattern. It consists of two COICOP-10 items, domestic and international package holidays, whereas international travels correspond to more than 95% of the total index. We model both series separately and account for the methodological change from January 2019[25], which revised them backward until 2015. Hence, we keep the real-time perspective of the exercise by producing nowcasts for the non-revised series until December 2018 – using June 2012 as the starting point of the sample due to a structural break in the seasonal pattern – while the revised series becomes the target as of January 2019 with a break-free sample starting in January 2015. Package domestic holidays are modeled with an AR structure with lags $\{1, 2, 6\}$ on the month-over-month growth rates augmented with seasonal dummies. For package international holidays, we fit the log level on the set of seasonal dummies and weekly Amadeus indices aggregated to the monthly frequency – more precisely, "last-minute" bookings (see Section 2.3.2).[26] Moreover, we correct the nowcasts of this model because time-invariant dummies cannot account for changes in the price level over time. Specifically, we adjust each nowcast by the nowcast error observed on average in the previous two months. Even though this model is specified solely in monthly data, we summarize it under the heading "U-MIDAS approach" to keep the labeling simple.

Direct machine learning approach. Our second nowcasting approach uses machine learning and directly targets the six components of headline inflation. Given the fairly

[24]Specifically, we use the following subsets to select from for each of the 14 COICOP-10 items: For "Electricity", "Natural gas, excl. share in the costs", "Share in the costs for gas central heating", and "District heating", we do not have well-matching weekly information available, which is why we select from the series provided by the WOB, the European Gas Spot Index (EGSI), and, for "District heating", a lagged moving average of heating oil prices. For "Liquefied gas, charging of a tank container" we select from LPG and heating oil prices. For "Heating oil", "Share in the costs for oil central heating", and "Coal briquettes", we select from heating oil prices provided by the WOB and Heizoel24. For "Firewood, wood pellets or the like", we use the series on pellet prices. For "Diesel fuel, cetane number below 60", "Diesel fuel, cetane number 60 and more", "Supergrade petrol, 95 octane", "Supergrade petrol, 98 octane and more" we select from the direct counterparts provided by the WOB and Tankerkoenig. For "Liquefied petroleum gas", we use the LPG prices provided by the WOB.

[25]For more details, see Box 5 in ECB Economic Bulletin Issue 2, 2019.

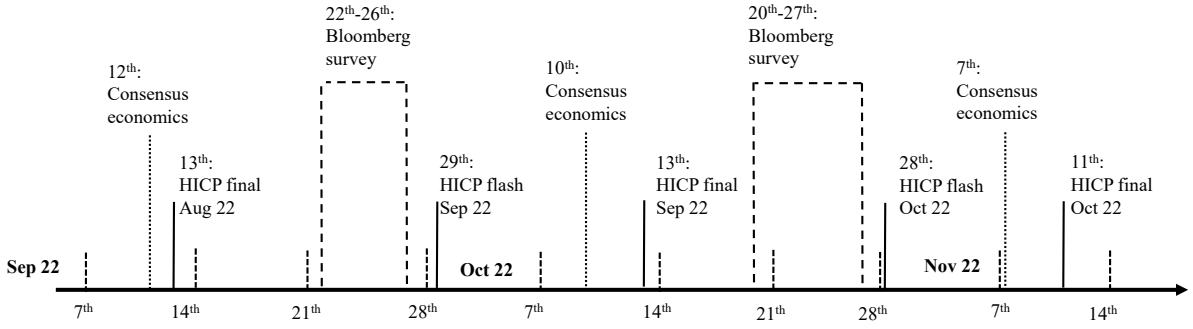
[26]The results are very similar when we use the first principal component of the Amadeus indices instead of their average.

good performance of the LASSO estimator within the horse race presented in Section 2.4.3, we replicate this ML strategy for unprocessed food, processed food, non-energy industrial goods and energy. Thereby, we proceed by first aggregating the weekly indicators of GFK:FMCG and ENERGY to the monthly frequency and specifying a LASSO regression of each target month-over-month inflation rate on the corresponding predictors. For package holidays, we directly model the COICOP-3 component that combines domestic and international packages, but using different models for both the non-revised series up to 2018 and the revised series as of 2019. In the former case, we implement an AR with lags $\{1, 4\}$ on the month-over-month growth rates augmented with seasonal dummies and monthly rates of the Amadeus series. For the revised series, we replicate the log level regression, introduced above. As for service inflation, we cannot fit a direct forecasting model as we do not have any relevant high-frequency indicator available. Therefore, we resort to the same bottom-up nowcasts derived from the SD-AR model used in the bottom-up nowcasting approach above.

Benchmark approach. For the six components of headline inflation, we use aggregated nowcasts of the SD-AR model applied to each item at the COICOP-10 level as the benchmark. For headline inflation, we also use market expectations provided by Bloomberg and Consensus Economics as the benchmark. Each month, these data providers ask economists mainly working in private banks to report their best estimate for headline inflation in Germany, along with many other macroeconomic variables. The Bloomberg survey generally takes place within the third week of a given month and is available since 2015. Consensus Economics has only recently started to ask survey participants to also provide an estimate for inflation in the current month. The survey is conducted within the second week of each month and data is available since March 2021.

In Figure 2.2, we summarize the timeline of the different data releases of market expectations and HICP releases between September 2022 and November 2022. At the end of each of the four weeks in each month, we produce a new nowcast based on incoming high-frequency data. Note that in the first week of September, the nowcasts need to use our previous estimate for inflation, since the official numbers for the COICOP-10 series have not yet been published. The statistical office publishes a first estimate for headline inflation at the end of the corresponding month, for example, on September 29th, followed by the final numbers in the subsequent month, in this case, on October 13th. Consensus Economics has surveyed its participants on September 12th, and Bloomberg between September 22th to September 26th.

Figure 2.2: Timeline of survey-based market expectations and German HICP data releases



Note: The figure shows, for the months September 2022 until October 2022, the four weeks ending at days 7, 4, 21, and 28, at the end of which we produce a nowcast based on new high-frequency data. In addition, it shows the data releases for inflation and survey periods of market expectations.

2.4.5 Real-time information set on non-standard policy measures, annual updates of HICP weights and imputed prices in times of crisis

Our evaluation period has been marked by extraordinary events as the outbreak of the COVID-19 pandemic in 2020 and Russia's invasion of Ukraine in 2022. These events brought with them a set of non-standard policy measures which also affected consumer prices, such as a temporary VAT cut in 2020 and one-time emergency aid measures for gas, heating and public transport. Notably, all these policy measures were communicated by the government well before their introduction and therefore included in professional forecasters' information set at that time. Hence, for a proper comparison with market-based inflation expectations, we include *a priori* information on the introduction and reversal of a given policy measure in our nowcasting models. More precisely, we apply ex-ante assumptions when nowcasting the months 2020M7 and 2021M1 (introduction and reversal of temporary VAT cut), 2022M7 and 2022M9 (introduction and ending of heavily reduced public transport tickets) as well as 2022M12 (one-time emergency aid for gas and heating).[27] If we exclude these *a priori* information on policy measures, our results for headline inflation would be obscured by larger nowcasting errors in those components where we do not have high-frequency price information (e.g. services and durable goods).

Finally, the COVID-19 pandemic also hampered price collection by statistical offices considerably. This is notably true for the prices of travel-related services which had to be estimated during lockdown periods, typically by using the month-over-month inflation rate of the previous (non-pandemic) year to reflect the seasonal pattern of these prices. Since this imputation procedure for package holidays was publicly known at that time, we take it into account when nowcasting the HICP subcomponent of package holidays.[28]

[27] Appendix 2.B.3 provides more details on the specific policy measures and how they are accounted for within our econometric framework.

[28] In Germany, the prices for package holidays were imputed during the period 2020M4-2020M6 and 2020M9-2021M5.

Moreover, disaggregate HICP weights are updated at the beginning of a year, but typically only published with the final January HICP figures in February. Hence, in our nowcasting models, we use the previous year's weight information when nowcasting the headline rate for January.

2.5 Results

We present the results of our recursive out-of-sample nowcast experiment in three steps. First, we demonstrate the usefulness of including weekly GFK:FMCG data in nowcasting models of HICP inflation at the COICOP-10 level, i.e., the most disaggregate available COICOP level. Then, we show that combining these data with machine learning methods yields superior forecasts at more aggregate product-group levels. Finally, we document that even headline inflation nowcasts benefit considerably when this approach is used, outperforming market expectations in most periods.

2.5.1 Results of the item-level inflation nowcasts

Table 2.3 reports the root mean squared error (RMSE) of the monthly inflation nowcasts by the U-MIDAS model relative to the benchmark SD-AR model for selected COICOP-10 items for which weekly GFK:FMCG data are available. We group the results into panels of product groups such as unprocessed fruit and vegetables, processed meat and fish, and dairy products and fat. Within each panel, the columns refer to the within-month information sets of day 7, 14, 21, and 28. For readability, each cell is colored in a heatmap style where darker colors indicate a lower relative RMSE and hence a better nowcasting performance of the U-MIDAS model compared to the SD-AR benchmark.

For many products and nowcasting days, feeding the weekly GFK:FMCG information in a U-MIDAS model reduces the nowcast error substantially relative to the benchmark. Consider, for example, the RMSE of nowcasting the official year-over-year inflation rate for sweet peppers, which is the first item in the upper left panel. Day 7 nowcasts with the U-MIDAS model yield a relative RMSE of 0.51, cutting the nowcast error almost by half. To understand this result, recall that on day 7 of a month t, the latest official information available pertains to month $t-2$. Hence, the SD-AR benchmark model effectively needs to generate a two-step ahead forecast. By contrast, the U-MIDAS model uses the GFK:FMCG data, which are complete for month $t-1$ and even include the first week of month t. This information advantage can be expected to produce a more accurate nowcast as long as the GFK:FMCG index is reasonably correlated with its official counterpart.

On day 14, the relative RMSE of the U-MIDAS model for sweet peppers increases

slightly to approximately 0.56. It remains around this level on days 21 and 28. Why is the superiority of incorporating GFK:FMCG data mitigated compared to day 7? On days 14, 21, and 28 of a month t, the official HICP of sweet peppers for month $t-1$ has already been published. The benchmark model thus only needs to generate a one-step ahead forecast so that the informational advantage of using GFK:FMCG data declines in this respect. At the same time, the U-MIDAS model can exploit additional weekly GFK:FMCG information of month t which may result in a more stable signal concerning this month. However, for sweet peppers – and for many other items – the effect of expanding the information set to days 22 and 28 does not lead to (notably) better nowcasts, which is probably related to the typical practice in official price statistics of recording many prices around the middle of a month.

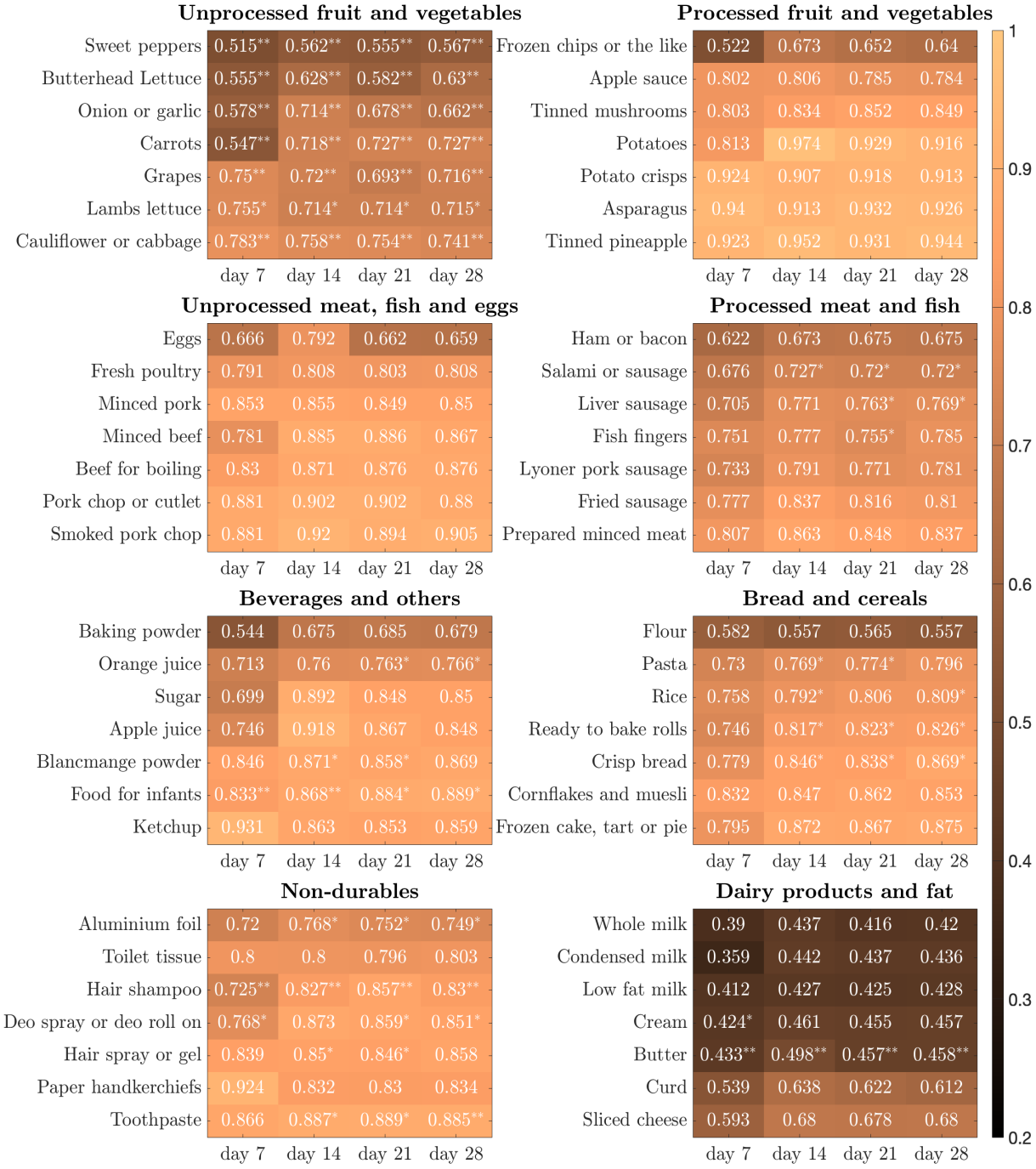
Across product groups, GFK:FMCG information improves the nowcasts at the item level particularly strongly for unprocessed fruit and vegetables as well as dairy products and fat, for which RMSE reductions in the range of 40%-60% can be achieved, also indicated by statistically significant results of the Diebold and Mariano (1995) test of equal predictive accuracy in many of these cases. The U-MIDAS model also exhibits very good nowcasting properties for many items of the other product groups.

What if we directly match the monthly GFK:FMCG inflation series to their official HICP counterparts using OLS? This basic forecasting method (hereafter *OLS match*) avoids using the MIDAS model and constructs the nowcasts by aggregating the available weekly GFK:FMCG information to the monthly frequency. Table 2.8 of Appendix 2.C.1 replicates the relative RMSE figures and reveals that nowcasting accuracy can be slightly improved across items that already perform exceptionally well using U-MIDAS. This applies to most items within the category unprocessed fruit and vegetables and dairy products. By contrast, U-MIDAS still yields smaller RMSE values across product groups that cannot surpass nowcasting gains of 50%. Therefore, improvements in relationship to the SD-AR benchmark are mostly stemming from precise weekly GFK:FMCG information.

So what drives nowcasting success at the item level? The most important factor is a close match between the GFK:FMCG and official price indices. Figure 2.3 displays the predictive gain of using the GFK:FMCG price indicators as a function of the in-sample fit with their official counterparts. Evidently, a higher correlation typically goes hand in hand with a smaller RMSE compared to the benchmark model. This negative relation also holds across different nowcasting days. On average, including the GFK:FMCG information starts to reduce the nowcast RMSE once the correlation between the month-over-month rates exceeds 0.4.

Figure 2.3 also shows some differences between high-level product groups. Processed food items (denoted by red dots) exhibit correlations between the GFK:FMCG and official

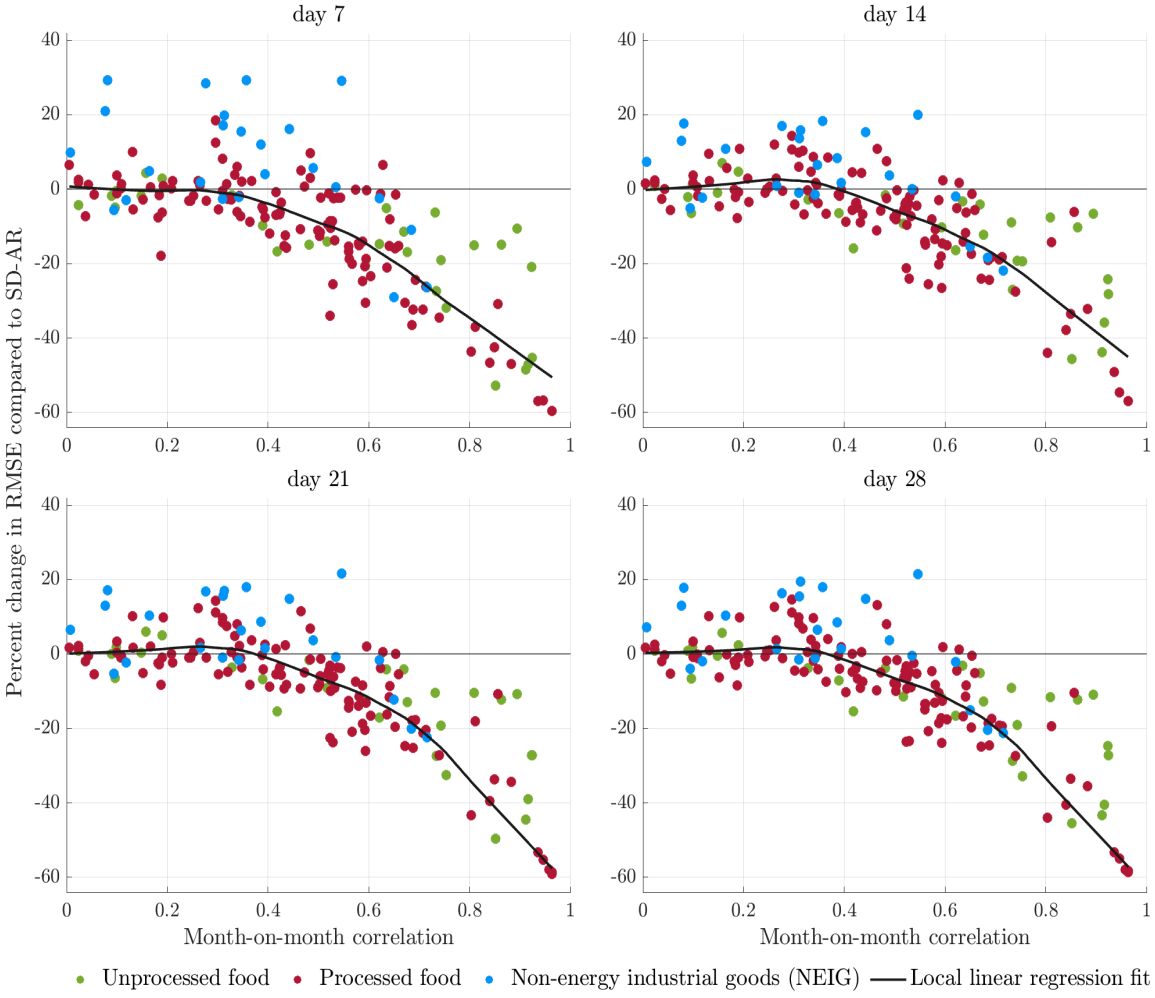
Table 2.3: RMSE for FMCG product-level inflation: U-MIDAS relative to the SD-AR benchmark



Sources: GfK household panel; own calculations.
Notes: The figure shows heatmaps of RMSE values for the U-MIDAS model relative to the SD-AR benchmark at nowcasting days 7, 14, 21 and 28 for the best-performing COICOP-10 items within selected FMCG product groups. Results for the Diebold and Mariano (1995) test in the event of outperformance relative to the SD-AR model are indicated by the symbols * (5% level) and ** (1% level).

price indices on the full range between zero and almost 1. They appear to be most representative of the relationship between correlation and nowcast gain just discussed. Most items of unprocessed food (denoted by green dots) show a strong or very strong correlation, but the nowcast benefit is more heterogeneous. By contrast, non-energy

Figure 2.3: Predictive gains of GFK:FMCG in relation to the fit with official counterparts



Sources: GfK household panel; own calculations.
Note: For each FMCG item at the COICOP-10 level, the figure shows the percentage change in RMSE of the U-MIDAS nowcasts (on days 7, 14, 21 and 28) compared to SD-AR as a function of the fit between GFK:FMCG indicators and their official counterpart based on correlations using month-over-month rates. Outliers at the 1st and 99th percentiles of the RMSE changes are removed.

industrial goods (blue dots) mostly exhibit moderate to low correlations between the GFK:FMCG and official price indices; thus, including the scanner data in many cases does not pay off. These products are typically characterized by a greater product variety than food items, which makes it more difficult to fit the official price indices.

2.5.2 Results of the product group-specific inflation nowcasts

The inflation rates of the three high-level product groups unprocessed food, processed food and non-energy industrial goods (NEIG) receive considerable attention in the Eurosystem. As described in Section 2.4.3, we use machine learning shrinkage methods to estimate direct nowcasting models for these product groups (and some subgroups) that successfully include the large set of regressors we have available, namely the underlying

weekly GFK:FMCG price indicators at the COICOP-10 level. We compare these models to SD-AR benchmark models fitted to the group-specific inflation rates.

Table 2.4 displays the RMSEs of the group-specific shrinkage models relative to their benchmark. The top panel refers to the three high-level product groups. With regard to unprocessed and processed food, the shrinkage models clearly outperform the benchmark with reductions in the RMSE between 15% and 25% on all nowcasting days. By contrast, for non-energy industrial goods (NEIG) the benchmark dominates. This outcome likely reflects the different coverage rates of the GFK:FMCG data across product groups. As shown in Table 2.2, 30 out of 38 COICOP-10 items of unprocessed food and 116 out of 142 COICOP-10 items of processed food are matched, but only 39 out of the 302 NEIG with semi-durables and durables almost lacking completely. In addition, even the relatively few matched NEIG items, mostly non-durables, do not correlate very strongly with their HICP counterparts, as reported in Figure 2.3.

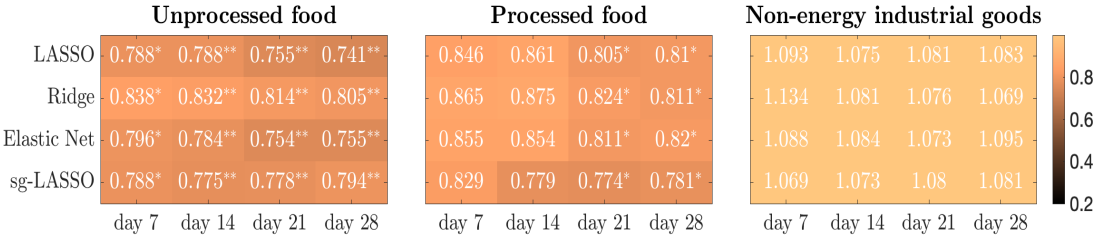
The bottom panel in Table 2.4 shows the results for the eight more disaggregated low-level product groups. In the large majority of cases, it pays off considerably to use shrinkage models that include weekly GFK:FMCG data. The advantage is particularly large for dairy products and fat (reduction in RMSE of roughly 45% to 55%), unprocessed fruit and vegetables (reduction of around 20% to almost 40%), processed meat and fish (reduction of more than 25% to almost 40%), and unprocessed meat, fish and eggs (reduction of nearly 20% to 25%). For processed fruit and vegetables, bread and cereals, and beverages and others, the nowcasting gains are more muted, but it is still generally beneficial to use shrinkage models. Only in the case of non-durables is there no clear difference to the benchmark, which likely again reflects the low correlation of the GFK:FMCG items at the COICOP-10 level with their HICP counterparts.

The weekly flow of information affects the nowcasting performance in a way very similar to the underlying COICOP-10 items discussed in the previous section. Most importantly, the information available at day 7 of a given month already turns out to be highly valuable. This likely reflects the fact that at day 7 of a month t, the benchmark model includes only official inflation rates of month $t-2$, while the shrinkage approaches use the full GFK:FMCG data of month $t-1$ and the first week of month t. The additional information exploited at day 14 typically further improves the nowcasts in absolute terms (see Figure 2.8 of Appendix 2.C.1 for the absolute forecast errors over time by nowcasting day), whereas this is not always the case relative to the benchmark, which on that day includes the official inflation rates of month $t-1$. Finally, the additional information gained in weeks 3 and 4 of a month is of minor quantitative importance.

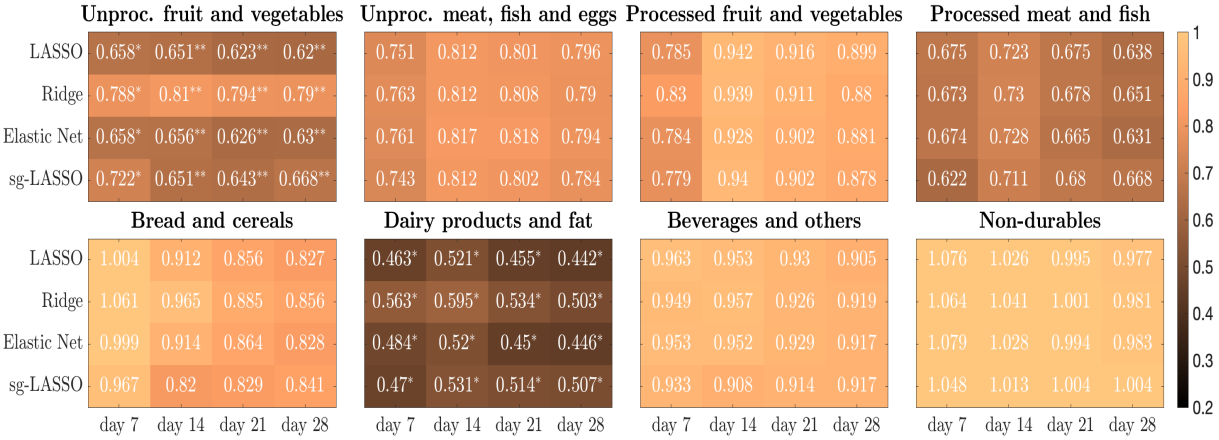
Concerning the different shrinkage approaches, the nowcasting results do not favor a single method. The general conclusion is that it is important to include the weekly GFK:FMCG dataset and make it usable in an appropriate way. To this end, standard

shrinkage methods (LASSO, ridge and elastic net) work generally as well as the sg-LASSO approach, which performs variable selection in a mixed-frequency setting and fully accounts for the time series nature of the dataset. Nevertheless, for processed food, which is the product group with by far the largest number of underlying COICOP-10 items that we match with GFK:FMCG data, the sg-LASSO is clearly superior. This may indicate that this approach is especially promising when it comes to very high-dimensional estimation and nowcasting settings.

Table 2.4: RMSE for FMCG product-group inflation: shrinkage methods relative to the SD-AR benchmark



(a) High-level product groups



(b) Product groups

Sources: GfK household panel; own calculations.
Notes: The figure shows heatmaps of RMSEs for nowcasts based on (i) shrinkage methods (LASSO, ridge and elastic net) and (ii) sg-LASSO relative to the SD-AR benchmark for FMCG higher-level components and subcomponents. Results for the Diebold and Mariano (1995) test in the event of outperformance relative to the benchmark are indicated by the symbols * (5% level) and ** (1% level).

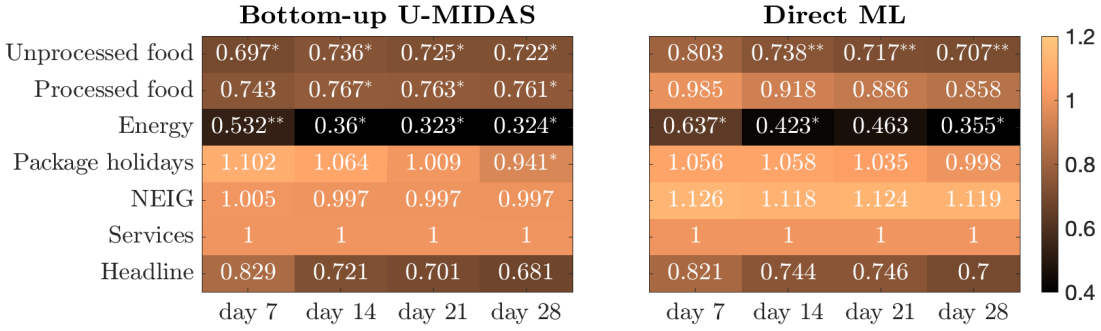
2.5.3 Results of the headline inflation nowcasts

Table 2.5 shows the RMSE of the bottom-up U-MIDAS approach and the direct machine learning approach relative to the benchmark for headline inflation and its six components. Let us first focus on the bottom-up U-MIDAS approach. Recall that it uses U-MIDAS models to nowcast each COICOP-10 item that is matched by weekly GFK:FMCG, energy price, or international package holiday data, and the SD-AR model for the remaining items before it aggregates all these nowcasts with the help of the official HICP weights.

By contrast, the benchmark model fits the SD-AR model to all COICOP-10 items, while it uses the same aggregation strategy. Hence, the relative RMSE tells us in a clear way how beneficial the inclusion of weekly external data for aggregate inflation forecasting is.

From the left panel of Table 2.5 we infer that the weekly external data make a considerable difference. On day 7, the informational advantage relates to unprocessed food, processed food, and energy, reducing the RMSE by roughly 30%, 25% and 47% respectively, whereas the SD-AR model is sufficient to nowcast package holidays, non-energy industrial goods, and services (for which we do not have any weekly indicators). This translates to a reduction in the RMSE of headline inflation by 18% compared to the benchmark. The informational advantage increases to levels close to 32% for headline inflation as more information accumulates over days 14, 21, and 28, mainly because nowcasts of the energy component and package holidays become more accurate. In fact, using the OLS match approach with monthly aggregated estimates of price changes – see Table 2.9 of Appendix 2.C.1 – leads to slightly improved performance for most components and headline inflation.

Table 2.5: RMSE of headline inflation and its components: bottom-up U-MIDAS and direct ML approaches relative to the bottom-up SD-AR benchmark



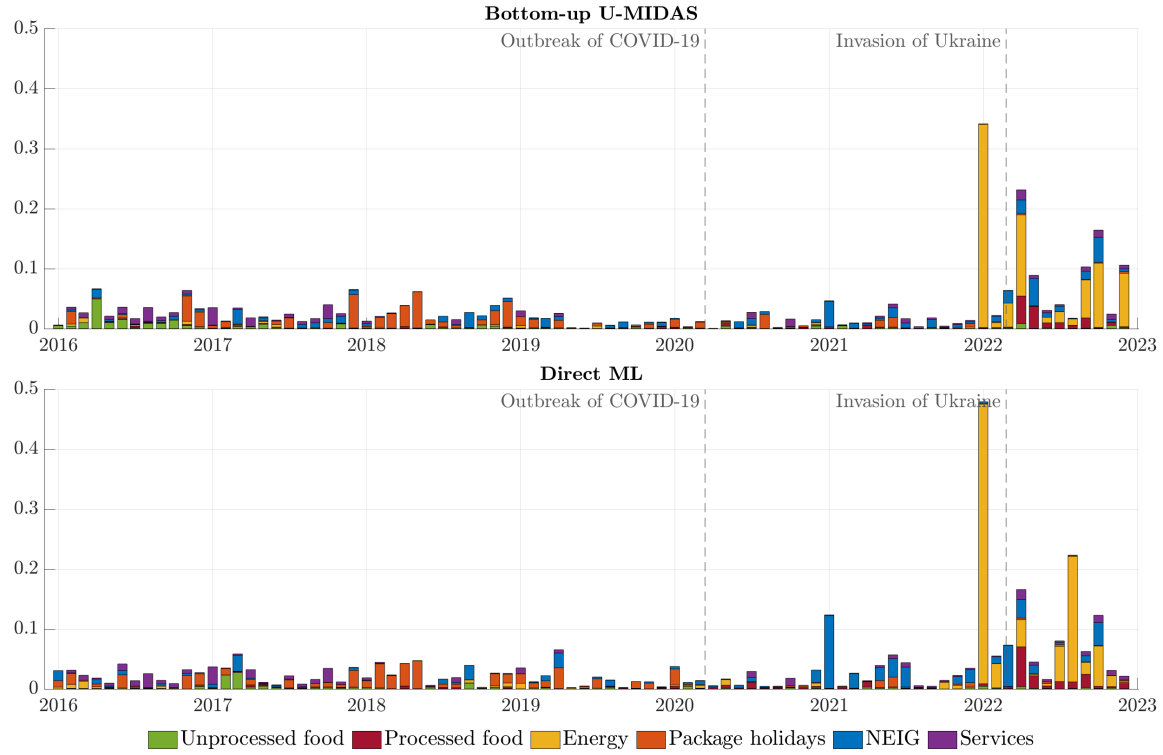
Sources: GfK household panel; European Commission's Weekly Oil Bulletin; Amadeus; own calculations.
Note: The figure shows heatmaps of RMSEs for nowcasts based on (i) the bottom-up U-MIDAS approach with aggregation via HICP weights and (ii) the direct machine learning relative to the benchmark approach, which is a bottom-up nowcast based on SD-AR models fitted at the COICOP-10 level. Results for the Diebold and Mariano (1995) test in the event of outperformance relative to the benchmark are indicated by the symbols * (5% level) and ** (1% level).

The right panel of Table 2.5 shows how the machine learning approach performs. Recall that it directly models the inflation rates of the six components as a function of all underlying weekly data and then aggregates these six nowcasts to headline inflation. Except for unprocessed food, the nowcasts generally deteriorate when compared to the bottom-up U-MIDAS both for the components (especially processed food) and to a smaller extent, for headline inflation. This result suggests that – in a setting where official aggregation schemes are known and easy to implement – there is no systematic advantage of employing ML models to estimate data-dependent aggregation weights. While the latter may, in theory, better reflect the dependence structure of the target series and its underlying predictors, we conjecture that in our case, the time series used to estimate them are neither

long nor stable enough to outweigh the associated increase in the nowcast variance.

Is this disadvantage of the direct machine learning approach related to specific periods? To shed light on this issue, Figure 2.4 displays the squared forecast errors of the six components of headline inflation over time, where we focus on day 28. Note that we multiply these squared forecast errors with their COICOP weights, which we use to aggregate the component-wise forecasts to the headline forecast, in order to obtain an impression of their overall relevance – keeping in mind, of course, that the weighted sum of the six squared forecast errors (which we show in the figure) is not equal to the square of the weighted sum of the six forecast errors (which amounts to the headline forecast error). Nevertheless, comparing the top and bottom panels shows that the energy shocks of the year 2022 are the main source of forecast errors for the U-MIDAS, but even more so for the direct machine learning approach. On a smaller scale but still relevant are the forecast errors relating to non-energy industrial goods – where the direct machine learning in particular exhibits weaknesses during the pandemic and thereafter – while package holidays account for the majority of the errors throughout normal times like those prior to the pandemic, especially around 2018.[29] These findings indicate that estimating data-dependent aggregation weights is particularly detrimental in turbulent times.

Figure 2.4: Contribution of HICP components to squared headline forecast errors



Sources: GfK household panel; European Commission's Weekly Oil Bulletin; Amadeus; own calculations.
Notes: The figure shows the weighted squared forecast errors of the six components of headline inflation on day 28 of a month for the bottom-up U-MIDAS approach (top panel) and the direct machine learning approach (bottom panel).

[29] See Figure 2.8 of Appendix 2.C.1 for a period-wise illustration of these forecast errors for GFK:FCMG groups.

Finally, we benchmark our model-based nowcasting results with expectations surveyed by Bloomberg and Consensus Economics among market participants. To that end, we report the cumulative sum of the loss differential of our model-based nowcasts versus the median survey expectations. Using the squared forecast error as our loss measure, we calculate the differential as

$$D_{t,ij} = -\sum_{\tau=1}^{t}(e^2_{\tau,M_i} - e^2_{\tau,S_j}), \qquad t = 1, \ldots, T, \tag{2.7}$$

where e_{t,M_i} denotes the nowcast error of model M_i (either U-MIDAS or direct machine learning) and e_{t,S_j} denotes the nowcast error of market survey S_j (either Bloomberg or Consensus Economics). A positive value of $D_{t,ij}$ indicates that model i outperforms survey j while negative values imply the opposite.

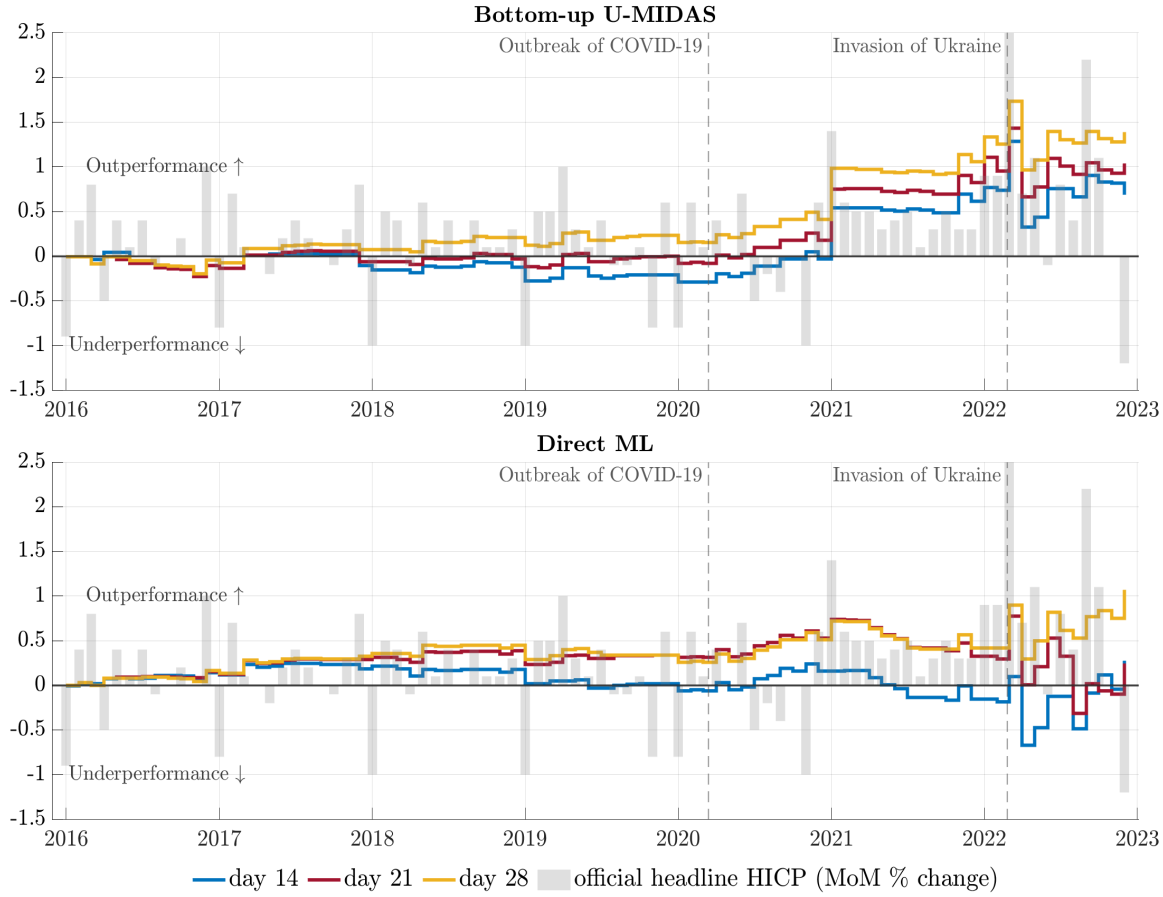
Figure 2.5 shows the evolution of the loss differentials over time for nowcast days 14, 21, and 28 which are competitive to the Bloomberg survey.[30] The bottom-up U-MIDAS is roughly on par with Bloomberg expectations until the pandemic hits, while the direct machine learning approach even outperforms consistently during this period if at least high-frequency information up to day 21 is used. In fact, the modest underperformance of the bottom-up U-MIDAS compared to direct ML until 2020 stems primarily from larger nowcast errors for unprocessed food (see Figure 2.4). Following the pandemic period, forecasting gains in relation to the Bloomberg benchmark accumulate gradually and consistently over time so that our modeling strategies clearly outperform if at least information up to day 14 is included for the bottom-up U-MIDAS and day 28 for the direct ML.

Surprisingly, in January 2021, our simple ex-ante assumption for the VAT impact (see Section 2.4.5) led to a considerable jump in the performance of the bottom-up U-MIDAS compared to Bloomberg expectations. Moreover, the rising inflation scenario following the pandemic in 2021 reveals itself to be a challenging period for the direct machine learning approach in terms of keeping its cumulative advantage, especially using only the information set of days 14 and 21. Similarly, Russia's invasion of Ukraine favors the Bloomberg survey on impact irrespective of the information set; nevertheless, the bottom-up U-MIDAS and the direct ML on day 28 quickly catch up with previous advantage levels. This means that efficiently exploiting daily and weekly external information from energy markets allows us to maintain the edge in relation to Bloomberg expectations throughout 2022. Across information sets, marginal gains in loss differential mostly persist when information up to days 21 and 28 are included. Overall, these findings suggest

[30] We leave out loss differentials for day 7 because the Bloomberg survey outperforms any model based on this very limited information set by far during the post-pandemic sample, compressing the scale of the axis measuring the loss differential so that the differences between days 14, 21, and 28 become difficult to distinguish. Figures that include day 7 are available upon request.

that estimating data-dependent aggregation weights based on ML methods improves the nowcasting precision of headline inflation in normal times but can be detrimental in unstable environments.

Figure 2.5: Cumulative sum of the squared forecast error differentials: bottom-up U-MIDAS and direct ML versus Bloomberg survey-based expectations



Sources: GfK household panel; European Commission's Weekly Oil Bulletin; Amadeus; Bloomberg survey; own calculations.
Notes: The figure shows, on the left axis, the cumulative sum of the squared forecast error differential of the bottom-up U-MIDAS approach (top panel) and the direct machine learning in relation (bottom panel), respectively, in comparison to Bloomberg survey-based expectations on days 14, 21 and 28. The gray bars represent official month-over-month percentage changes in headline inflation.

The survey results published by Consensus Economics have only been reported since March 2021, which precludes a comparison over the whole nowcasting sample. Nevertheless, in the 20 months available, the bottom-up U-MIDAS and the direct machine learning approaches strongly outperform the survey expectations, as shown in Figure 2.9 of Appendix 2.C.1.

2.6 Robustness analysis

We evaluate the robustness of the results reported in the previous section with respect to three aspects. First, we focus on the data input and investigate whether alternative methods of compiling GFK:FMCG price indices at the COICOP-10 level lead to different outcomes when compared to the (weighted) TPD regression described in Section 2.3.1. Second, given the high weekly volatility of some COICOP-10 item indices, we assess whether their predictive content improves by smoothing out volatile time series. Finally, we focus on the modeling strategy: (i) the robustness of the baseline U-MIDAS setting to alternative model specifications, extending the information set by non-contemporaneous weekly inflation rates and quantity indices; and (ii) the stability of product group-specific results to different folds of the cross-validation procedure in machine learning tools and different degrees of the sg-LASSO Legendre polynomial.

2.6.1 Alternative methods of compiling COICOP-10 price indices

In our baseline setting, we compute weekly price indices from the granular GFK:FMCG data employing a (weighted) TPD regression. In doing so, we follow the practice of statistical agencies worldwide that rely on this method due to its good results with respect to in-sample fit and nowcasting. To examine the extent to which this also holds for our GFK:FMCG data, we have implemented various alternative methods of computing weekly price indices.

We start by transforming daily prices into weekly aggregates using the arithmetic mean, the geometric mean and the median price. We do this both for the raw price data and the data excluding outliers. We consider several choices for the product sample underlying the computation of each price index constructed following the above-mentioned approaches. Due to the constant replacement and entry of new products in the GFK:FMCG dataset, price series can be highly volatile, with sudden spikes and jumps. Official price statistics address this issue by regularly selecting a basket of goods and services that is kept fixed for a specific period of time. Hence, we follow a similar approach by considering only prices of products that are available in each month of the entire sample since 2003, or within each month of the preceding two years. Next, we follow a standard time-dependent rule in official price collection and include only prices of products that have been bought between the 12th and the 18th of each month. Additionally, we focus on goods sold in discount shops, as these shops hold significant pricing power in Germany, causing other shops to adjust their prices accordingly. Finally, we employ a top-seller approach by selecting only those products with the highest market share that cover 50% of the market in the previous two years prior to the current year.

In total, combining the alternative aggregation methods and selection strategies de-

scribed above, we end up with more than 50 different price indices for each COICOP-10 item. Using only "pre-sample" data dating back to before January 2016, the start of our nowcasting experiment, we then compare the month-over-month inflation correlations of the various methods with their official counterparts. This allows us to rank them and determine the best method based on the highest in-sample correlation for each COICOP-10 series.

Next, we use the respectively best method for each COICOP-10 item to replicate the analysis in Section 2.5.1. The results in Figure 2.10 of Appendix 2.C.2 show that the predictive properties found using the TPD-based GFK:FMCG indicators are, across all nowcast horizons, generally not affected when implementing the best pre-sample method by COICOP-10 instead. In fact, the comparison reveals a slight decay in the predictive ability of the best method compared to the TPD approach, especially in cases where the baseline U-MIDAS combined with TPD indicators substantially outperforms the SD-AR benchmark. Overall, the findings support the use of the TPD method as the baseline method since the in-sample fit can only be marginally improved by alternative methods while delivering the best solution from an out-of-sample perspective.

2.6.2 Smoothing out volatile inflation series

As described in Section 2.3.3, some COICOP-10 price indices can be highly volatile, which is a common property exhibited by high-frequency data. To filter out short-term noise of volatile GFK:FMCG price series and then reassess their predictive properties, we consider moving average filters using one to four weeks of past data. Figure 2.11 of Appendix 2.C.2 plots the changes in terms of nowcast RMSE when applying the four-week moving average smoother as a function of the weekly volatility level of each COICOP-10 item. It turns out that the predictive ability of smoothed inflation series worsens on average for the most volatile group, namely unprocessed food items, on days 7 and 14. By contrast, as we approach end-of-month nowcast horizons, changes in RMSEs become less pronounced. These findings indicate that it is difficult to reduce the noise in GFK:FMCG prices without affecting the signal that is related to official inflation dynamics.

2.6.3 Alternative U-MIDAS specifications and ML hyperparameter choices

In the first robustness check, we enrich the U-MIDAS information set in (2.6) by including non-contemporaneous GFK:FMCG weekly inflation rates from periods $t-1$ and $t-2$. The results are summarized in Figure 2.12 of Appendix 2.C.2. They show that the RMSE is generally stable across COICOP-10 items after accounting for the weekly inflation rates of $t-1$. Modest RMSE improvements in the range of 10% to 15% can be mostly attributed

to items in which U-MIDAS already outperforms, to some extent, the SD-AR benchmark in the baseline scenario, especially when nowcasting on day 7. Unsurprisingly, the same pattern can be identified when additionally accounting for the high-frequency inflation lags of period $t-2$. Hence, incorporating past high-frequency information only leads to minor predictive gains, and mostly does so for processed food and NEIG items. This suggests that the autoregressive component in (2.6) sufficiently accounts for recent price dynamics while contemporaneous GFK:FMCG information constitutes the key signal for well-performing nowcasts. Moreover, a higher number of distributed lags in the U-MIDAS setting goes hand in hand with an increased nowcasting variance.

The second robustness check investigates whether GFK:FMCG quantity indices improve the quality of disaggregate inflation nowcasts beyond contemporaneous information of price indices. To this end, we add the month-over-month quantity indices at the weekly frequency, and their lags, as regressors in model (2.6). The RMSE practically remains unchanged across all COICOP-10 series, indicating that the inclusion of quantity information does not enhance the predictability of disaggregate inflation rates beyond what is already conveyed by GFK:FMCG price indices. As an exception, results for the COICOP-10 series "flour" display a noteworthy improvement when nowcasting days 22 and 28. This most likely reflects that the start of the war in Ukraine had strong effects on the average quantity-price relationship of "flour".

With the third robustness check, we examine whether the product group-specific nowcasting results discussed in Section 2.5.2 hold irrespective of our hyperparameter choices for the estimation of shrinkage methods. We start by increasing the degree of the sg-LASSO Legendre polynomial from $L = 0$ to $L = 1, 2$. The findings suggest a slight improvement in the precision of the nowcast in only a small number of cases where statistically significant outperformance is already achieved by sg-LASSO with $L = 0$ compared to SD-AR. Hence, it represents the optimal choice given that it promotes a higher dimensionality reduction and carries a smaller computational burden when estimating sg-LASSO coefficients. Similarly, we test for different folds of the cross-validation considering the grid set $k \in \{5, 10, 15, 20, 25\}$. Overall, these choices favor similar tuning parameters and model architectures, thus not altering the results for group-specific targets.

2.7 Conclusion

The recent decade has witnessed a burst in granular high-frequency information stemming from all parts of the economy. To be useful for policymakers and society at large, this vast amount of data needs to be processed with care and fed into appropriate models.

This paper demonstrates how pairing millions of household scanner data with state-of-the-art machine learning techniques yields highly competitive real-time inflation nowcasts

for Germany, both at a very disaggregate level as well as for major product groups and headline inflation. The guiding principle of our approach is to use the economic structure inherent in the construction of official price indices to organize and condense the information carried by granular purchase decisions at the household level before we open up the machine learning toolkit. This strategy is reflected in the three steps of our analysis: we start at the most disaggregate level possible, proceed to an intermediate level of product groups, and finally turn to headline inflation.

In the first step, we construct a set of more than 180 weekly price indices at the COICOP-10 level from the granular scanner data. The virtue of this approach is that the mapping from the detailed product descriptions available in the scanner data to the official consumption basket underlying German inflation is straightforward as the COICOP-10 level includes items such as "butter", "coffee beans", and "sanitary cleaner". We show that the scanner-based indices obtained in this way track their official counterparts well, especially for food items (with an average correlation of 0.9 for year-over-year inflation rates). When fed into a MIDAS model, they also improve disaggregate monthly inflation nowcasts, notably as soon as after the first seven days of a month.

In the second step, we turn to nowcasting product groups like unprocessed and processed food, taking all available weekly scanner-based indices into account. To this end, we apply shrinkage estimators to cope with the high-dimensional predictor set. Relative to a time series benchmark model, we obtain substantial RMSE predictive gains of up to 25%.

In the final step, we nowcast headline inflation. Once again, we use the COICOP architecture to do so. Specifically, we split headline inflation into six components which exhibit very heterogeneous time series properties and for which we can come up with different types of weekly predictors. We demonstrate that this approach yields highly competitive nowcasting models that are on par with, or even outperform, survey-based expectations, which are notoriously difficult to beat.

In summary, our strategy to combine fixed economic structures like the COICOP system with flexible machine learning tools turned out to provide us with accurate inflation nowcasts at different levels of aggregation. Our approach thereby exploits the virtue of granular data to provide an understanding of the disaggregate dynamics underlying overall inflation, whilst at the same time yielding valuable high-frequency real-time information about price developments of aggregates closely monitored by policymakers.

Looking ahead, given the considerable value added of high-frequency scanner data for inflation nowcasting documented in this paper, there is an urgent need to identify and exploit high-frequency information concerning those parts of the consumption basket underlying German inflation that are not covered by scanner data on fast-moving consumer goods. These include services, clothing and footwear, and the large range of items typically referred to as slow-moving consumer goods such as furniture and household appliances.

Appendix 2

2.A Supplementary information on data, descriptive statistics and data construction

2.A.1 Overview of data sources

Table 2.6: Data sources of high-frequency price series and official inflation in Germany

Variable	Source	Description
		Consumer Prices
Headline inflation	DESTATIS	Harmonised index of consumer prices (HICP), unadjusted data.
10-digit series	DESTATIS	German national consumer price index (CPI), unadjusted data.
		Fast-moving consumer goods (GFK:FMCG)
Prices	GFK	Scanner data recorded by private households on daily purchases.
		Travel services
Prices	AMADEUS	Transaction data on daily bookings of package holidays by German travelers.
		Energy prices
Euro super	Weekly Oil Bulletin	Average Monday pump price including duties and taxes
Diesel	Weekly Oil Bulletin	Average Monday pump price including duties and taxes
LPG motor fuel	Weekly Oil Bulletin	Average Monday pump price including duties and taxes
Heating oil	Weekly Oil Bulletin	Average price for deliveries of 2,000 to 5,000 liters incl. duties/taxes
Euro super	Tankerkoenig	Real-time prices of all gasoline stations in Germany
Diesel	Tankerkoenig	Real-time prices of all gasoline stations in Germany
Heating oil	Heizoel 24	Average daily heating oil price of participating delivery firms
Wood pellets	Holzpellets.net	Average daily wood pellets price of participating delivery firms
European Gas Spot Index (EGSI)	European Energy Exchange AG	Volume-weighted average price of all spot transactions concluded on the trading day. Values before October 2021 are mean values of the gas prices from the two former market areas GASPOOL and NetConnect Germany.
		Market expectations
HICP nowcast	BLOOMBERG	Survey among market participants and professional forecasters.
HICP nowcast	CONSENSUS	Survey among market participants and professional forecasters.

Notes: *DESTATIS*: Federal Statistical Office of Germany, *GFK*: Growth from Knowledge, *AMADEUS*: AMADEUS IT group, *HAVER*: Haver Analytics

2.A.2 Additional descriptive statistics of high-frequency scanner data

Table 2.7 summarizes some basic characteristics of the GFK:FMCG data. Overall, we observe an average price of 1.7 euro, with the highest prices being paid for durables, and the lowest for dairy products and fat. Over the full sample, we observe about 100,000 distinct products by month distributed fairly equally across HICP subcomponents. The only exceptions are unprocessed fish and durables, where product coverage is relatively low; however, in more recent years, product variety has increased for these products, too. On average, a product stays in the sample for about 190 days, whereas dairy products have the longest lifetime with about 407 days compared to 47 days for durable goods. About 43% of all purchases are made in discount shops, with the largest discounter share attributable to unprocessed fish, followed by unprocessed meat and fruit. The availability of consumption data in discount shops is very important for nowcasting official inflation in Germany, since large discounters typically act as price leaders in the market. Finally, about 74.4% of household purchases are spend on processed food, in particular beverages and dairy products and fats, followed by non-energy industrial goods (15.5%) and unprocessed food (9.6%).

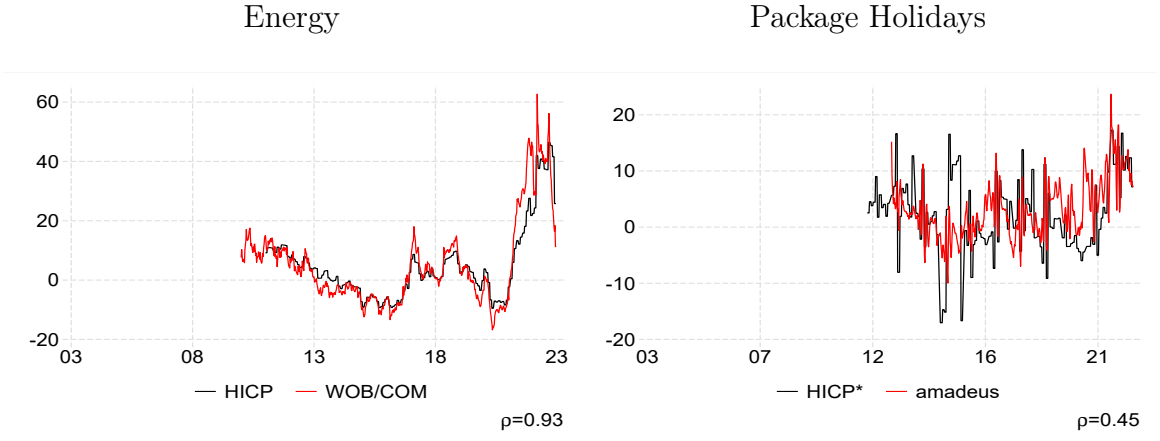
Finally, Figure 2.6 displays our supplementary high-frequency indicators on energy and package holidays, together with their official counterpart. For energy, comovement at the monthly frequency is nearly perfect, with a correlation coefficient of 0.93. The correlation for package holidays is also fairly high, at 0.45 if one adjusts the official inflation rate for a large jump in 2016 that reflects a statistical break due to the current HICP chain-linking practice (Dietrich et al., 2021).

Table 2.7: Summary statistics of GFK:FMCG

Component	mean	median	10^{th}	90^{th}	#ID	#T	#discount	#expenses
Unprocessed food	2.2	1.8	0.9	3.8	5,853	150	0.53	9.6
Fruit	1.7	1.7	0.9	2.8	1,124	110	0.56	1.8
Vegetables	1.4	1.3	0.8	2.1	1,594	210	0.46	1.8
Meat & eggs	2.9	2.3	1.1	5.0	3,024	146	0.56	5.9
Fish	3.1	2.7	1.8	5.0	111	166	0.66	0.2
Processed food	1.5	1.1	0.5	2.7	66,459	271	0.46	74.4
Fruit	1.7	1.5	0.7	3.0	1,941	256	0.44	1.8
Vegetables	1.4	1.1	0.6	2.2	4,102	245	0.43	3.8
Meat	1.7	1.5	0.9	2.8	8,436	300	0.53	9.1
Fish	2.3	1.8	0.8	4.0	2,119	218	0.48	2.3
Bread & cereals	1.4	1.2	0.6	2.5	10,399	271	0.48	9.5
Dairy products & fat	1.1	0.9	0.4	2.0	10,167	407	0.46	16.4
Beverages	1.8	1.0	0.3	4.2	12,303	236	0.42	18.0
Other food products	1.4	1.1	0.5	2.6	16,991	226	0.46	13.5
Tobacco								
Non-energy industrial goods	2.3	1.5	0.5	4.5	24,509	87	0.27	15.5
Non-durables	2.3	1.5	0.5	4.5	23,752	87	0.27	14.9
Semi-durables	3.5	1.9	0.7	8.7	674	105	0.22	0.4
Durables	13.9	10.0	5.0	23.3	83	47	0.10	0.2
Total HICP	1.7	1.3	0.5	3.0	96,551	190	0.43	100

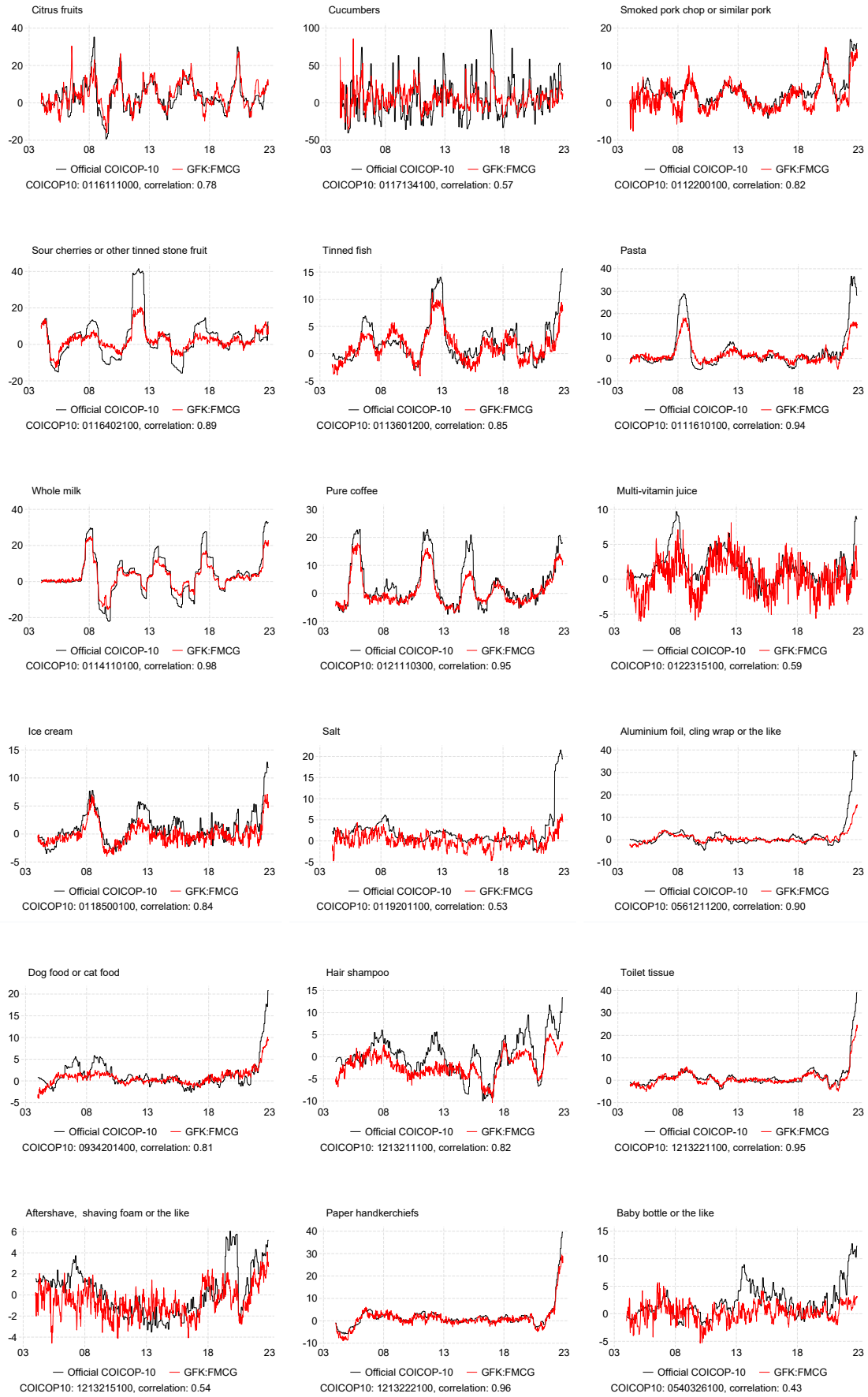
Notes: This table reports the mean, the median, and the 10th and 90th percentile of the daily scanner prices. Column #ID shows the average number of unique products per month, #T the average lifetime of products in days, #discount the share of products bought in discount shops and #expenses the share of consumption purchases for each product category.

Figure 2.6: German HICP inflation and high-frequency counterparts: energy and package holidays



Notes: The figure shows YoY inflation rates (% change) for HICP subcomponents aggregated using all of the corresponding COICOP 10-digit level series for which we have high-frequency data available. "HICP" refers to the aggregates using official COICOP-10 series, "WOB" refers to the data from the weekly oil bulletin and from commodity prices, and "Amadeus" refers to the transaction data for package holidays. Note that the official inflation rate for package holidays has been adjusted for a large jump in 2015 caused by a statistical methodology change. ρ reports the correlation coefficient between both series.

Figure 2.7: Inflation series of selected COICOP-10 items and weekly scanner-based counterparts



2.A.3 Construction of transaction-based price indicators for package holidays

The German HICP subindex on package holidays mainly consists of international flight holidays. Importantly, prices for package holidays enter the HICP on the day of travel rather than the day of booking, contrary to other products such as food, where prices are recorded on the day the price is recorded in-store. Therefore, in the case of package holidays, we can exploit booking advances for forecasting purposes, since we are already able to observe the bulk of bookings for a future package holiday in a given month well in advance.

We aggregate the AMADEUS micro dataset on package holidays in various stages. First, we omit all bookings in the dataset that have been canceled by the traveler or the travel agencies and select only those holiday regions which have been entered into the official price index up to 2022 (Turkey, Spain including Balearic Islands and Canary Islands, Greece, Dominican Republic and Egypt).[31] Second, since micro prices refer to the overall price of a booking, we perform a quantity adjustment and define the price by traveler and travel days:

$$p_{i,t}^{raw,amadeus} = sales\_exp_{i,t}^{amadeus} \frac{1}{N_{i,t}} \frac{1}{D_{i,t}} \tag{2.8}$$

In addition, we compute the advance booking in days as the difference between the day of booking and the day of travel. To correct for outliers, we omit prices which are smaller than 20 euro and larger than 600 euro per traveler and day and with advance booking of more than one year. Third, similar to official price statistics, we categorize the bookings by advance booking to differentiate between different pricing schemes of early (last-minute) bookings, regular and early bookings. We opt for five different booking categories (bookings up to 14 days before departure, bookings between 15-30, 31-90, 91-180 and 181-270 days, as well as more than 270 days before departure), whereas last-minute bookings convey the most helpful signal to nowcast our target series. Finally, to meet the structure of the weekly nowcasts, we aggregate daily prices into weeks 1 to 4, but in addition, each week now consists of price indices categorized by five different categories of booking advances.

[31] The German price index for package holidays includes additional price representatives such as domestic package holidays, cruises and city trips. However, the bulk of the index represents those five regions above. See Henn et al. (2019) for a description of the underlying HICP methodology.

2.B Supplementary information on the econometric methodology

2.B.1 Nowcasting item-level inflation

To set the scene, let us assume that the latest official data on inflation of a COICOP-10 item c has been released for a given month t. For convenience of notation, however, we stop referring to the item c. Next, conditional on high-frequency data available up to period t and a pre-sample observation π_0^M, our baseline U-MIDAS (2.6), neglecting seasonal dummies and autoregressive lags greater than one, can be estimated via OLS following the matrix representation

$$\begin{bmatrix} \pi_1^M \\ \pi_2^M \\ \vdots \\ \pi_t^M \end{bmatrix} = \begin{bmatrix} 1 & \pi_0^M & x_1^{(m)} & x_{1-1/m}^{(m)} & x_{1-2/m}^{(m)} & x_{1-3/m}^{(m)} \\ 1 & \pi_1^M & x_2^{(m)} & x_{2-1/m}^{(m)} & x_{2-2/m}^{(m)} & x_{2-3/m}^{(m)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \\ 1 & \pi_{t-1}^M & \underbrace{x_t^{(m)}}_{\text{4th week}} & \underbrace{x_{t-1/m}^{(m)}}_{\text{3rd week}} & \underbrace{x_{t-2/m}^{(m)}}_{\text{2nd week}} & \underbrace{x_{t-3/m}^{(m)}}_{\text{1st week}} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \rho_1 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix} \tag{2.9}$$

Note that the matrix representation (2.9) makes explicit the transformation of the high-frequency predictor $x_{t-k/m}^{(m)}$ into m low-frequency vectors $(x_{1-k/m}^{(m)}, \ldots, x_{t-k/m}^{(m)})'$, for $k = 0, \ldots, m-1$. Hence, model (2.6) is estimated in the low-frequency dimension, but our nowcasts can be updated each time high-frequency increments become available after t, whereas random walk forecasts of the most recent high-frequency information are used to deal with the "ragged-edge problem".

Finally, note that $B(L^{1/m}; \theta) = \sum_{k=0}^{K} B(k; \theta) L^{k/m}$ is a polynomial of length $(K+1)$ in the $L^{1/m}$ operator with $L^{k/m} x_t^{(m)} = x_{t-k/m}^{(m)}$. Despite we assume $K = m-1$ for simplicity, leading to a static model in the high-frequency component for $h = 0$, one might include high-frequency distributed lags $K > m-1$ that span over past low-frequency periods (see robustness section 2.6).

2.B.2 Nowcasting product group-specific inflation

To nowcast product groups covered by the GFK:FMCG dataset such as unprocessed food, processed food and non-energy industrial goods, let us assume a set of monthly aggregated inflation indicators $\mathbf{x}_t = (x_{1t}, \ldots, x_{qt})'$ such that $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_t)'$, where q denotes an abundant number of COICOP-10 series belonging to a given product group. Hence, conditional on official inflation data available at t, we model our higher-level product group $\pi^{\mathbf{M}} = (\pi_{g,1}^M, \ldots, \pi_{g,t}^M)'$ as a function of $\mathbf{X} = (\iota, \mathbf{x})$ using standard shrinkage methods such as LASSO, ridge and elastic net regression, where ι is a t-dimensional vector of ones. The hybrid elastic net estimator solves the following penalized least squares problem:

$$\hat{\beta} = \min_{\hat{\beta}} ||\pi^{\mathbf{M}} - \mathbf{X}\beta||^2 + \lambda \left(\alpha |\beta|_1 + \frac{(1-\alpha)}{2} ||\beta||^2 \right), \tag{2.10}$$

where $\alpha \in (0, 1]$ is a weight parameter that interpolates between LASSO ($\alpha = 1$) and ridge regression (as $\alpha \to 0$) while the regularization parameter λ controls the amount of shrinkage in β.[32] Hence, the idea is to shrink coefficients $b_{g,c}$ to or towards zero if the c-th COICOP-10 series is not relevant. Finally, we construct the monthly aggregated estimate of $\mathbf{x}$ using the available contemporaneous weekly information at the time of the nowcast and compute it based on the estimated parameters $\hat{b}_g$.

As a second class of models, we construct a nowcast for $\pi^M_{g,t+h}$ using the sg-LASSO-MIDAS framework (see Babii et al., 2021) that handles high-dimensional mixed-frequency prediction problems. Let the matrix of covariates now be defined as:

$$\mathbf{X} = (\iota, \mathbf{X_1^{(m)}} W, \ldots, \mathbf{X_q^{(m)}} W), \tag{2.11}$$

where $\mathbf{X_j^{(m)}} = (X^{(m)}_{c1}, \ldots, X^{(m)}_{ct})'$ is a $t \times m$ matrix of the c-th high-frequency covariate and W denotes a predetermined $m \times L$ matrix of weights based on Legendre polynomials of degree L that aggregate over the high-frequency lags. Then, the sg-LASSO estimator solves the penalized least squares problem:

$$\hat{\beta} = \min_{\hat{\beta}} ||\pi^{\mathbf{M}} - \mathbf{X}\beta||^2 + 2\lambda \left(\alpha |\beta|_1 + (1-\alpha) ||\beta||_{2,1} \right), \tag{2.12}$$

where $||\beta||_{2,1} = \sum_{G \in \mathcal{G}} |\beta_G|_2$ is the group LASSO norm for a group structure $\mathcal{G}$ that hereby constitutes all high-frequency lags of a single covariate. Thus, in this case, $\alpha \in [0, 1]$ determines the relative importance of LASSO sparsity and the group structure.[33] This implies that sg-LASSO promotes sparsity between and within COICOP-10 items, allowing us not only to select the relevant COICOP-10 series but also the appropriate lag structure of each item.

2.B.3 Modeling the impact of policy measures in the evaluation period

To keep our nowcasting models on a competitive level with expert opinions, we extend their information set by *a priori* knowledge on three non-standard policy measures implemented by the German government which affected consumer prices. All these measures were announced well before taking effect and were hence known to professional forecasters when they were nowcasting for the corresponding month.

First, to mitigate the negative economic effects of the COVID-19 pandemic, the German government announced on 4 June 2020 that the VAT rate would temporarily be cut in

[32] Both tuning parameters λ and α are determined via expanding cross-validation.

[33] Note that $\alpha = 0$ leads to the group LASSO estimator, which is reminiscent of the elastic net regressor.

the period July to December 2020. Specifically, it decreased the regular VAT rate (which applies to about 65% of prices collected in the HICP) from 19% to 16%, and the reduced rate that mainly applies to food (excluding beverages), newspapers and books from 7% to 5%. We implement this *a priori* information in our nowcasting models as follows. First, all COICOP 10-digit items are classified according to their VAT category (regular, reduced, or VAT-free). Next, while the dates and the extent of the VAT changes were announced beforehand and were thus known to forecasters, the pass-through to actual prices was not (see, for example, Deutsche Bundesbank, 2020). We thus include in our forecasts an ex-ante pass-through of 50% applied uniformly to all products, which mimics a Bayesian forecaster with a flat prior and symmetric loss function. This ex-ante VAT impact is subtracted from the target variable prior to model estimation and then added to the model's forecast. To account for the VAT changes ex-post, we fit a dummy variable that takes the value -1 in July 2020 and +1 in January 2021.

Second, we consider a reduction in the price of public transport tickets, which was passed by the German government on 21 May 2022 as a response to soaring energy prices after Russia's invasion of Ukraine.Known as the 9 euro ticket, it allowed consumers to travel by local and regional public transport for 9 euros per month from June to August 2022. It reduced public transport prices considerably and in a foreseeable manner. We thus assume that forecasters were able to predict the corresponding changes in the affected price indices in June and September 2022.[34]

Third, the German government announced on 19 November 2022 a one-off emergency aid package for natural gas and heating taking effect in December 2022. The government assumed the December installment of the households' contracts with their natural gas and district heating suppliers. Due to the complex and heterogeneous contract designs in Germany, it was unclear by how much this measure would affect the relevant HICP series. Similarly to our strategy for the temporary VAT cut, we apply a simple assumption that a professional forecaster may have followed in real time. Specifically, we distinguish between homeowners and tenants. According to the latest EU-SILC survey, 46.7% of German households own their house or apartment. These households typically have a contract with a supplier of gas or district heating and thus directly benefited from the "December aid". The remaining households are tenants who typically pay their household energy via the landlord, from whom they receive an annual energy bill. Hence, they typically did not benefit (in December 2022) from the "one-off aid". We thus assume that in December 2022, the price for natural gas supply (0452103000) and district heating (0455002200) was zero for homeowners and unreduced for tenants. This implies an overall reduction by 46.7% for the December 2022 prices of natural gas supply and district heating.

[34]The affected COICOP-10 categories are "Train journey, short-distance" (0731111100), "Transport association, single or day ticket for adults" (0735011000), "Transport association, season ticket for apprentices" (0735013100), and "Transport association, monthly ticket for adults" (0735015000).

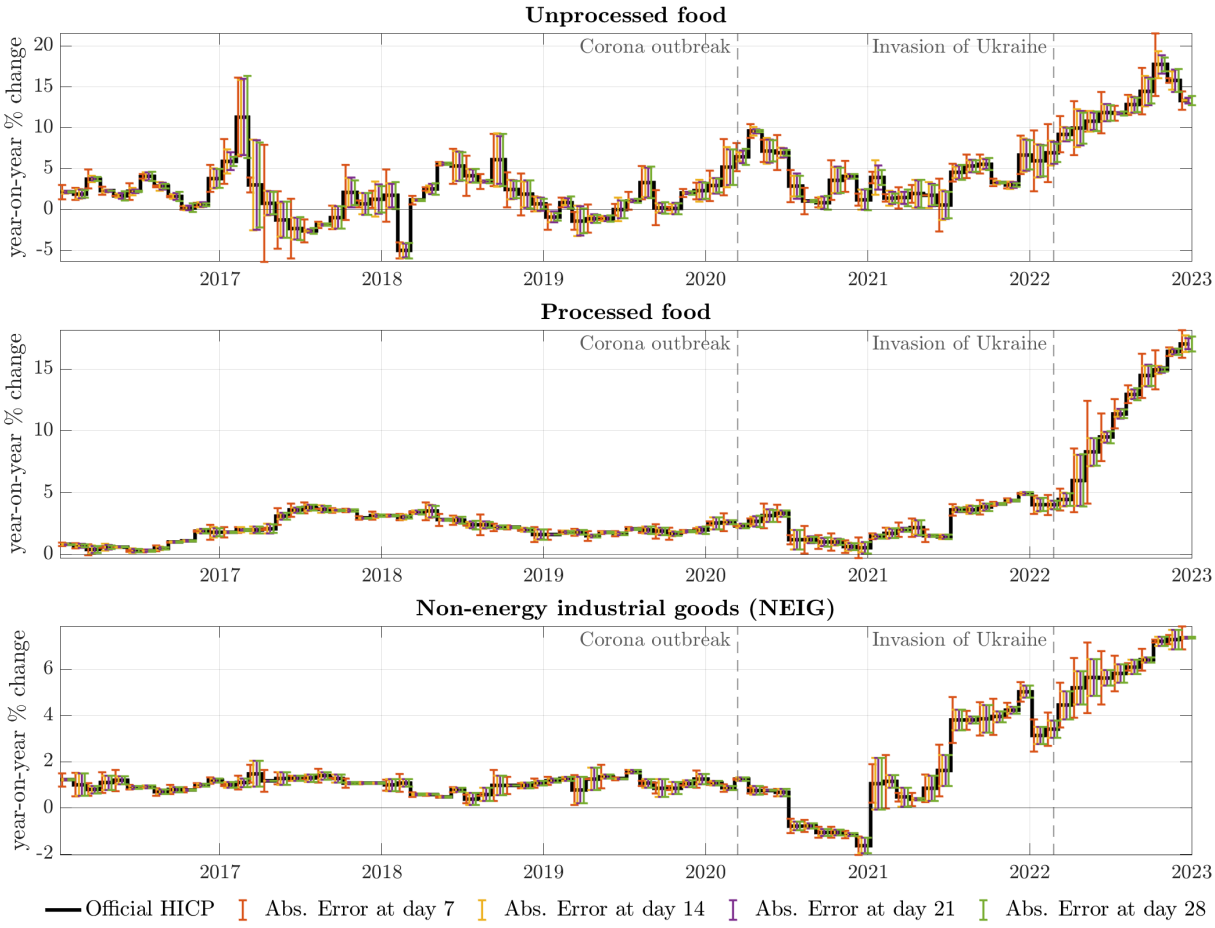
2.C Supplementary results

2.C.1 Nowcasting of product-level, group-specific and headline inflation

Table 2.8: RMSE for FMCG product-level inflation: OLS-match relative to the SD-AR benchmark

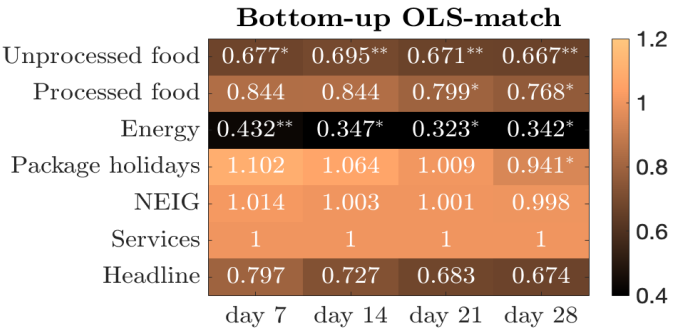
Sources: GfK household panel; own calculations.
Notes: The figure shows heatmaps of RMSE values for the OLS match model relative to the SD-AR benchmark at nowcasting days 7, 14, 21 and 28 for the best-performing COICOP-10 items within selected FMCG product groups. Results for the Diebold and Mariano (1995) test in the event of outperformance relative to the SD-AR model are indicated by the symbols * (5% level) and ** (1% level).

Figure 2.8: Absolute forecast errors over time: tracking the nowcasting performance of GFK:FMCG product groups



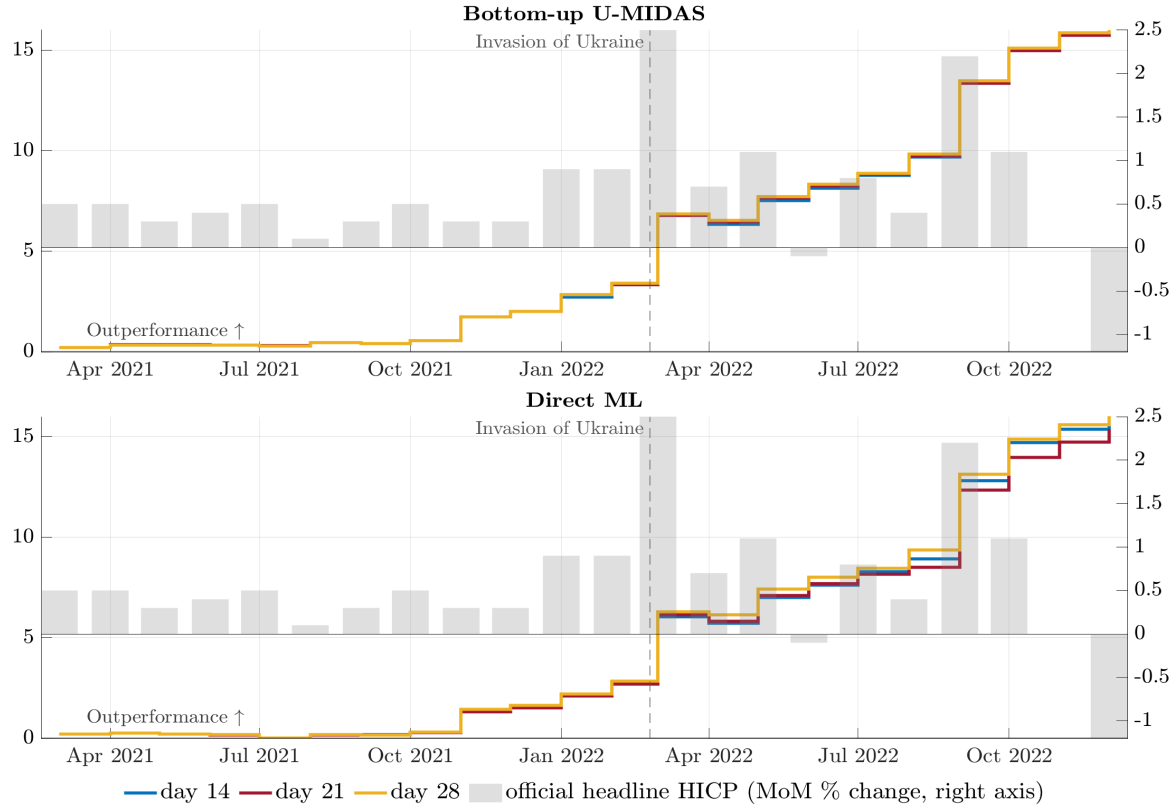
Sources: GfK household panel; own calculations.
Notes: The figure shows the evolution over time of the official inflation rates and the absolute forecast errors of the best-performing models on days 7, 14, 21 and 28.

Table 2.9: RMSE of headline inflation and its components: bottom-up OLS match approach relative to the benchmark approach



Sources: GfK household panel; European Commission's Weekly Oil Bulletin; AMADEUS; own calculations.
Notes: The figure shows heatmaps of RMSEs for nowcasts based on the bottom-up OLS match approach with aggregation via HICP weights relative to the benchmark approach, which is a bottom-up nowcast based on SD-AR models fitted at the COICOP-10 level. Results for the Diebold and Mariano (1995) test in the event of outperformance relative to the benchmark are indicated by the symbols * (5% level) and ** (1% level).

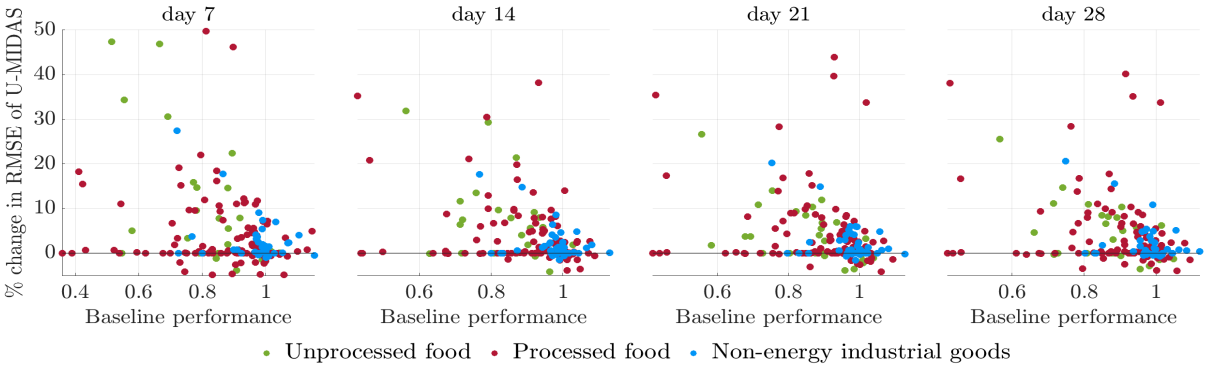
Figure 2.9: Cumulative sum of the squared forecast error differentials: models versus Consensus market expectations



Sources: GfK household panel; European Commission's Weekly Oil Bulletin ; AMADEUS; Consensus survey; own calculations.
Notes: The figure shows, on the left axis, the cumulative sum of the squared forecast error differential of the bottom-up U-MIDAS approach (top panel) and the direct machine learning in relation (bottom panel), respectively, in comparison to Consensus market expectations on days 14, 21 and 28. The gray bars (right axis) represent official month-over-month percentage changes in headline inflation.

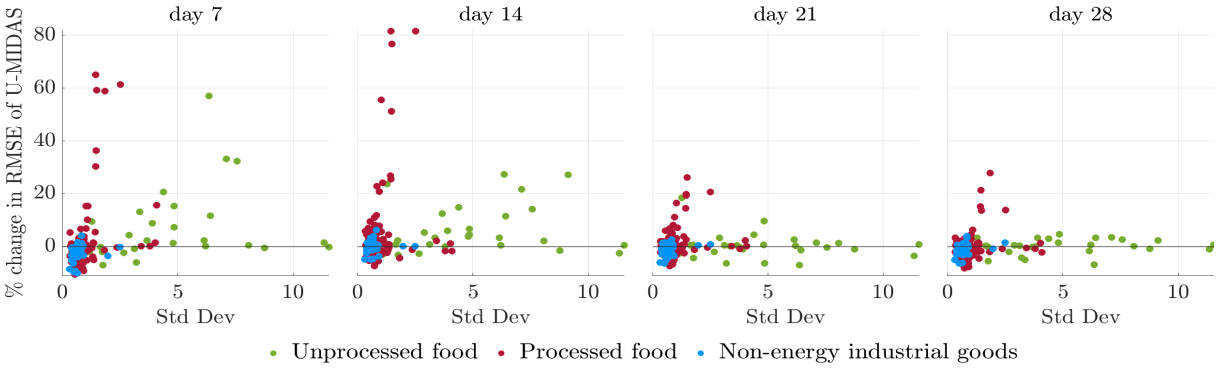
2.C.2 Robustness checks

Figure 2.10: RMSE change of implementing the best compiled COICOP-10 price indices



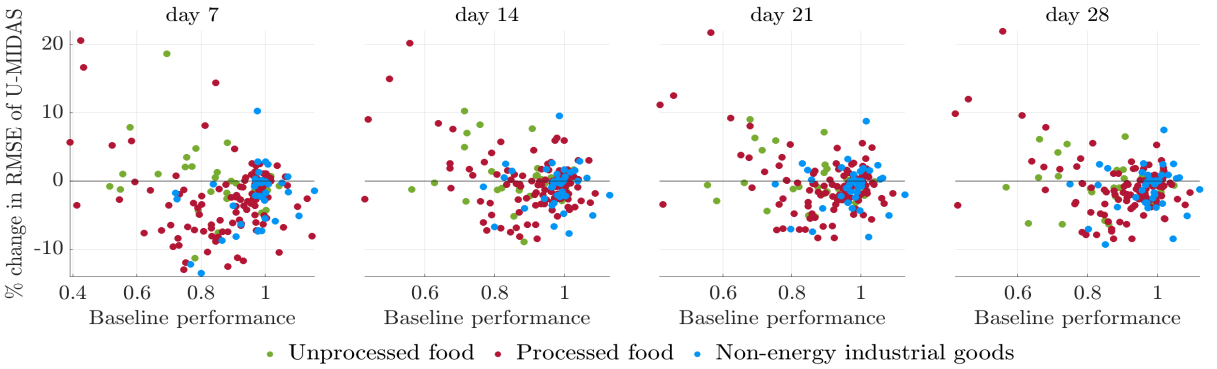
Sources: GfK household panel; own calculations.
Notes: For each GFK:FMCG COICOP-10 item, the figure shows the percentage change in RMSE of implementing the best-compiled price index (based on month-over-month inflation correlations), compared to the baseline U-MIDAS (2.6) as a function of its relative performance to SD-AR.

Figure 2.11: RMSE change of a moving average smoother as a function of the volatility level

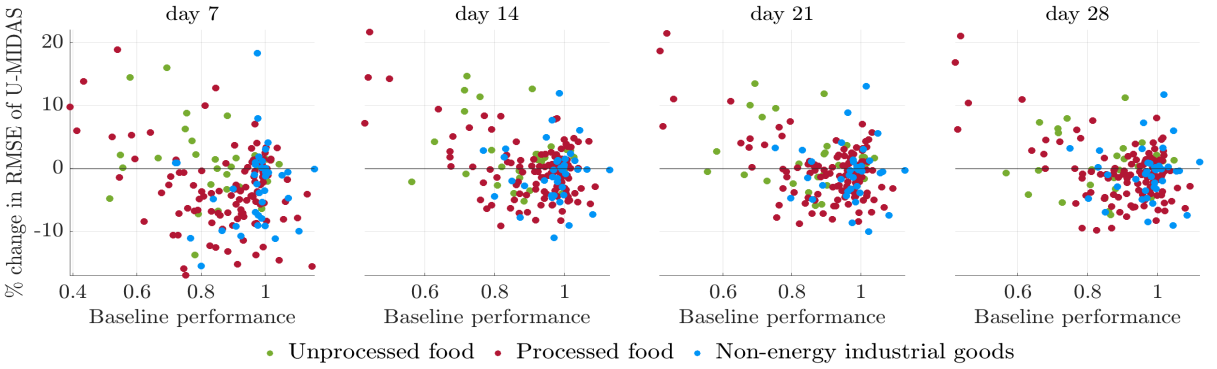


Sources: GfK household panel; own calculations.
Notes: For each GFK:FMCG COICOP-10 item, the figure shows the percentage change in RMSE of applying a four-week moving average smoother, compared to the baseline U-MIDAS (2.6) as a function of the weekly volatility levels (standard deviation of GFK:FMCG month-over-month inflation rates).

Figure 2.12: RMSE change of additional high-frequency distributed lags



(a) Contemporaneous + high-frequency inflation rates at $t-1$



(b) Contemporaneous + high-frequency inflation rates at $t-1$ and $t-2$

Sources: GfK household panel; own calculations.
Notes: For each GFK:FMCG COICOP-10 item, the figure shows the percentage change in RMSE of adding past high-frequency lags of GFK:FMCG indicators compared to the baseline static U-MIDAS (2.6) as a function of its relative performance to SD-AR.

Chapter 3

Bond Portfolio Optimization in Turbulent Times: a Dynamic Nelson-Siegel Approach with Wishart Stochastic Volatility

Abstract

Modeling and forecasting the time-varying volatility of bond yields plays a prominent role in many finance applications. However, amid periods of financial turmoil, managing interest rate risk on a daily basis poses significant challenges due to sudden shifts and extreme realizations in bond yields, often resulting in implausible density forecasts. This study aims to alleviate forecasting uncertainty and address structural instability in volatile bond markets by examining the predictive capabilities of flexible yield curve models featuring time-varying VAR parameters and Wishart stochastic volatility within a Bayesian MCMC scheme. A bond portfolio optimization and Value-at-Risk forecasting application to daily US Treasury yields underscore the potential benefits of modeling frameworks with factor Wishart stochastic volatility. The results indicate that a flexible yield curve model might provide statistical gains in density forecasting. In particular, Wishart stochastic volatility is economically justified due to its outperformance in terms of portfolio allocation and risk management, particularly during turbulent times such as the Great Recession and the COVID-19 pandemic.

Keywords: dynamic Nelson-Siegel model, Wishart stochastic volatility, Bayesian inference, bond portfolio optimization, Value-at-Risk forecasting.

JEL classification: C58, C11, E43, E47.

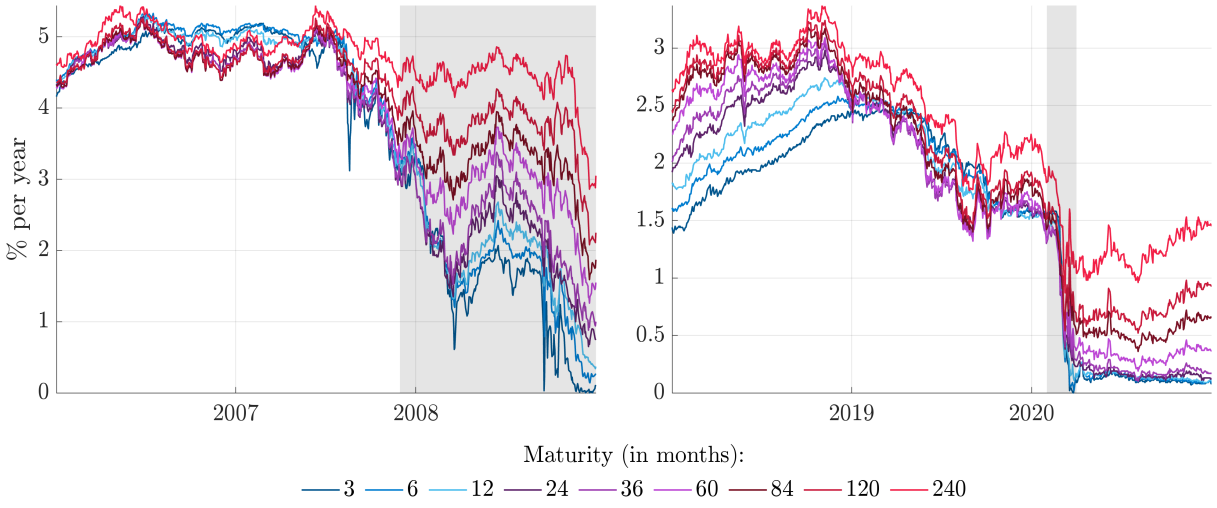
This study is a single-authored paper.

3.1 Introduction

Much research in financial economics has been devoted to yield curve modeling and forecasting using risk factors that incorporate the evidence of time-varying volatility in yields data. Interest rate point and density forecasting permeate many important tasks in finance, including bond asset pricing, portfolio allocation and risk management. In macroeconomics, the yield curve serves as a crucial indicator of economic conditions, furnishing policymakers with valuable insights into the state of the economy.

Nonetheless, Figure 3.1 illustrates how the US term structure of interest rates is susceptible to sudden and large changes in turbulent times such as the Great Recession and the COVID-19 pandemic. From an econometric perspective, extreme realizations (of either sign) and sudden changes in bond yields amid recessions can lead to strong effects on parameter estimates and implausible density forecasts (either too wide or too narrow) in constant volatility models. Amidst sharp volatility rises and instability in the dynamics of the US yield curve, the management of interest rate risk becomes notably challenging.

Figure 3.1: Daily US Treasury yields at constant maturities



Notes: The figure reports yields from the 3- up to the 240-month maturity. Shaded areas indicate US recession periods.

This study focuses on the economic value of yield curve density forecasts derived from Wishart stochastic volatility structures, particularly in the context of portfolio allocation and Value-at-Risk (VaR) forecasting during periods of turbulence in bond markets. Building upon the well-known dynamic Nelson-Siegel (DNS) model, this paper raises the following questions: Does a flexible factor structure with time-varying parameters (TVP) and Wishart stochastic volatility deliver (i) improved bond yield forecasts in moments of substantial financial distress? (ii) superior portfolio performance? (iii) better VaR forecasts? The answer to these questions contributes to a growing literature on yield curve forecasting and volatility modeling, whereas the proposed modeling framework is expected

to fully capture structural instability and reduce forecasting uncertainty in volatile bond markets. The study is among the first attempts to model yield curve dynamics with Wishart stochastic volatility. Additionally, it highlights the importance of multivariate volatility modeling in the context of bond portfolio optimization and risk management.

It is well-documented that modeling yield volatility movements can be computationally challenging. As with many financial assets, Figure 3.1 shows that yields at different maturities do not evolve independently over the business cycle, which suggests a common factor structure in large yields data. The DNS model suggested by Diebold and Li (2006) builds upon such factor structure by fitting all possible shapes of the yield curve using only three unobserved factors associated with the level, slope and curvature of the term structure. Although the baseline DNS framework proved to have surpassing predictive performance it assumes constant volatility and thus neglects the appropriate treatment of conditional yield volatility. In this sense, studies have provided evidence that its empirical performance can be improved by introducing stochastic volatility in the underlying yield curve factors (see, e.g., Bianchi et al., 2009; Hautsch and Yang, 2012; Laurini and Caldeira, 2016).

A recent development in financial econometrics is the application of Wishart processes to model multivariate stochastic volatility (MSV). Wishart processes naturally accommodate the essential statistical properties and flexibility to model nontrivial dynamics in covariance matrices (see, e.g., Gouriéroux et al., 2009). Philipov and Glickman (2006b) and Asai and McAleer (2009) propose a general Wishart inverse covariance (WIC) framework to assess MSV in observed financial returns. Thereafter WIC models have enjoyed growing popularity due to their ability to capture stylized facts observed in financial markets such as volatility clustering and volatility spillover effects in multivariate settings (see, among many others, Ku et al., 2014; Windle et al., 2014; Gribisch, 2016; Gorgi et al., 2018; Wu et al., 2018). These papers provide strong evidence that MSV driven by Wishart processes can be implemented with high-dimensional data carrying a factor structure, and thus making them feasible for practical applications in finance and macroeconomics.

Some minor efforts have already been devoted to incorporating Wishart stochastic volatility into the DNS framework. Byrne et al. (2017) propose a Wishart discount process based on exponential smoothing of volatility (see Koop and Korobilis, 2013) to compare DNS specifications with monthly macro-finance information. Kleppe et al. (2021) builds upon the Wishart process introduced by Uhlig (1997) to evaluate density forecasts of daily crude oil future contracts in a VaR forecasting application. Building on this strand of research, dynamic covariances of yield curve factors will be here driven by the general WIC process (suggested by Philipov and Glickman, 2006b; Asai and McAleer, 2009) and estimated using a one-step Bayesian MCMC scheme. On the empirical side, a key novel feature is the qualitative assessment of density forecasts of daily US Treasury yields in a

portfolio analysis during periods of financial distress.

The results suggest some main conclusions. First, random walk point forecasts represent a hard-to-beat benchmark, however, factor Wishart stochastic volatility does provide statistical gains in density forecasting. Second, optimal portfolios based on Wishart modeling strategies typically deliver higher excess returns and Sharpe ratios compared to benchmark models, especially for aggressive investment scenarios. Third, Wishart strategies fare best overall in terms of VaR forecasting, leading to better estimates for interest rate risk during highly volatile periods. In summary, there is compelling evidence that Wishart stochastic volatility and TVP-VAR (to a smaller degree) modeling features pay off in terms of portfolio allocation and risk management.

The outline of the paper is organized as follows. Section 3.2 illustrates the baseline DNS framework and introduces the proposed Wishart stochastic volatility specification. Section 3.3 describes the Bayesian simulation-based estimation scheme. Section 3.4 describes the data and empirical application. Specifically, the empirics section consists of a real-time out-of-sample forecasting exercise followed by a bond portfolio optimization and a VaR exercise. Finally, Section 3.5 concludes and briefly discusses some future research directions.

3.2 Yield Curve Modeling

3.2.1 The baseline dynamic Nelson-Siegel model

Diebold and Li (2006) suggest a parsimonious structure to properly fit the different shapes of the yield curve over time, the well-known DNS framework. It has been very popular in the financial literature because it combines flexibility and simplicity with surprising forecasting performance. To sum up, the DNS model specifies the evolution of yield curves by reducing a large set of observed yields to a lower-dimensional factor structure:

$$y_t = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \tag{3.1}$$

where y_t is the p-vector of continuously compounded yields to maturity at time t of \$1 receivable τ periods ahead, so that $y_t = [y_t(\tau_1), y_t(\tau_2), ..., y_t(\tau_p)]'$, with τ_1 being the shortest maturity considered and τ_p the longest. From a time series perspective, the factor loadings $\Lambda = (1, 1-\mathrm{e}^{-\lambda\tau}/\lambda\tau, 1-\mathrm{e}^{-\lambda\tau}/\lambda\tau-\mathrm{e}^{-\lambda\tau})$ are parameters, whereas λ controls the exponential decay rate of the curve, i.e., the rate at which factor loadings decay to zero.

The DNS model also carries appealing interpretation features: the time-varying coefficients can be interpreted as the yield curve Level (L_t), Slope (S_t) and Curvature (C_t) factors. In addition, the way each yield factor loads on observed yields y_t via Λ allows

us to interpret them respectively as long, short and medium-term factors. For instance, the slope factor S_t approximates the long-short spread and captures yield curve changes with larger impact on short-term maturities because its loading function starts at 1 but quickly decays to zero. All in all, yield curve factors govern the dynamics of y_t for any τ, and thus summarize broad information on bond price.

Diebold et al. (2006) frame the DNS factor structure into a dynamic state-space model (SSM), where L_t, S_t and C_t are treated as latent state variables and follow a VAR(1) process. The baseline DNS model can be summarized by the matrix notation:

$$y_t = \Lambda\alpha_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}_p(0, \Omega) \tag{3.2}$$

$$\alpha_t = \mu + \Upsilon\alpha_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}_k(0, \Sigma) \tag{3.3}$$

for $t = 1, ..., T$, where idiosyncratic shocks ϵ_t capture y_t movements not driven by yield curve factors and are assumed to be independent of factor shocks η_t. Note that α_t is the k-vector of latent factors, i.e., $\alpha_t = (L_t, S_t, C_t)'$, the factor loading matrix Λ is $p \times k$ while VAR coefficient matrices μ and Υ comprise a k-vector and $k \times k$ matrix, respectively. Finally, it is worth noting that the $p \times p$ measurement covariance Ω and $k \times k$ factor covariance Σ are time-invariant for all maturity spectrum and underlying yield factors.

The baseline DNS by itself captures a substantial part of the information in the yield curve, however, important dynamics concerning parameter instability and yields volatility are neglected (see Bianchi et al., 2009; Hautsch and Yang, 2012; Laurini and Caldeira, 2016; Byrne et al., 2017). It fails to capture sudden structural changes that occur by shifting interactions between yield curve factors in times of financial turmoil. Recent literature on macro-finance volatility has found that the combination of time-varying parameters and stochastic volatility often improves the accuracy of point and density forecasts when compared to constant-coefficient counterparts (see, among others, Clark and Ravazzolo, 2015; Chan and Eisenstat, 2018; Carriero et al., 2019a).

That said, to account for both model uncertainty and structural instability first, the DNS model is extended to allow for time-varying parameters in the VAR(1) process that describes the evolution of yield factors over time[1]. More precisely, define $B_t = (\mu_t'; \Upsilon_t')$ as the $(k + 1) \times k$ matrix of TVP-VAR coefficients and $\beta_t = \text{vec}(B_t)$ such that factor dynamics (3.3) develop into

$$\alpha_t = \mu_t + \Upsilon_t\alpha_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}_k(0, \Sigma) \tag{3.4}$$

$$\beta_t = \beta_{t-1} + e_t, \qquad e_t \sim \mathcal{N}_m(0, Q) \tag{3.5}$$

[1] As emphasized by Byrne et al. (2017), the null hypothesis that VAR coefficients are constant over time is consistently rejected, and hence employing a TVP-VAR structure to the factor equation (3.3) is appropriate. The study also finds that the competing TVP-VAR forecasting model consistently outperforms the constant parameter VAR.

where Q is the $m \times m$ covariance matrix of β_t with $m = k(k+1)$. It is worth emphasizing that relaxing the constant VAR parameter assumption is appealing because yield curve dynamics can suddenly change during unstable periods.

Secondly, since the ultimate goal here is to assess the statistical and economic value of yield curve forecasts amid recessions, it seems a natural choice to combine the DNS factor structure with Wishart stochastic volatility to fast-track structural changes in the (co)variance of yield factors. Diebold and Rudebusch (2013) indeed acknowledge a preference for modeling stochastic volatility in the underlying yield factors since the factor structure characterizes the deep workings of the system[2]. Furthermore, a factor volatility structure preserves the reduced-form modeling while providing a nice interpretation of factor volatilities as capturing the uncertainty surrounding changes in short-, medium- and long-term maturities.

3.2.2 The DNS with Wishart stochastic volatility

Based on the factor structure exhibited by yields data and the increasing literature on MSV modeling via Wishart processes, this paper applies a *Wishart stochastic volatility* (WSV) structure – as in Philipov and Glickman (2006a) and Asai and McAleer (2009) – for factor covariances Σ_t. The proposed DNS-WSV model can be cast in a nonlinear and non-Gaussian SSM:

$$y_t = \Lambda\alpha_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}_p(0, \Omega) \tag{3.6}$$

$$\alpha_t = \mu_t + \Upsilon_t\,\alpha_{t-1} + \eta_t, \qquad \eta_t \sim \mathcal{N}_k(0, \Sigma_t) \tag{3.7}$$

$$\Sigma_t^{-1} \sim \mathcal{W}_k(\nu, S_{t-1}), \tag{3.8}$$

$$S_{t-1} = \frac{1}{\nu}(A^{\frac{1}{2}})(\Sigma_{t-1}^{-1})^d(A^{\frac{1}{2}})', \tag{3.9}$$

where TVP-VAR coefficients evolve according to (3.5). Note that Σ_t is now time-varying and driven by the Wishart process (3.8)-(3.9) with ν degrees of freedom and scale matrix S_{t-1}. Hence dynamic factor volatilities are non-Gaussian and enter the factor equation (3.7) nonlinearly.

Parameter A is a positive definite symmetric matrix that carries information about intertemporal sensitivity relationships and determines how the elements of the contemporaneous precision matrix Σ_t^{-1} relate to the previous precision Σ_{t-1}^{-1}. As a technical remark, note that $A^{\frac{1}{2}}$ represents the lower Cholesky factor of A so that $A = A^{\frac{1}{2}}A^{\frac{1}{2}'}$. Parameter d determines the overall strength of these intertemporal relationships and accounts for persistence in the volatility process[3].

[2]Shin and Zhong (2017) show that modeling stochastic volatility in the factor equation better captures the time-varying volatility in yields data compared to stochastic volatility in the measurement equation.

[3]The power operator attributed to d is assumed to work element-wise.

It is worth emphasizing that the WSV modeling strategy enriches dynamic interactions among yield curve factors by allowing for cross-dependencies between volatility factors, i.e., it captures not only persistence in volatilities (volatility clustering) but also co-movements (correlation dynamics) in volatility factors. Hence the recent success of WSV modeling in a wide range of financial applications, such as risk management and asset pricing. In addition, the Wishart process does not suffer from high parameterization and allows for forecasting in a straightforward manner[4]. Nevertheless, the nonlinear non-Gaussian nature of (3.8)-(3.9) implies that Bayesian estimation becomes a bit difficult because conditional posteriors of Wishart parameters A, ν and d are not available in a closed-form solution. Section 3.3 illustrates the step-by-step Monte Carlo Markov Chain (MCMC) procedure that overcomes the technical challenges imposed by such a Wishart process.

Although the proposed factor WSV specification has been successfully applied to capture MSV in equity returns, it has not yet been contemplated by the DNS literature. Notwithstanding, let us consider an alternative Wishart structure for Σ_t: the Wishart matrix discount process employed by Byrne et al. (2017), which removes the need for a posterior simulation algorithm. The process approximates an Exponentially Weighted Moving Average (EWMA) filter (see Riskmetrics, 1996) and takes the form:

$$\Sigma_t^{-1} \sim \mathcal{W}_k(\nu_t, S_t), \tag{3.10}$$

$$\nu_t = \delta\nu_{t-1} + 1, \tag{3.11}$$

$$S_t = \delta S_{t-1} + f(\eta_t\eta_t'), \tag{3.12}$$

where δ is a discount factor that indicates the amount of time-variation expected in factor volatilities Σ_t and $f(\eta_t\eta_t')$ denotes a specific function of the squared residuals in the factor equation (3.7)[5]. Applications with EWMA estimators (for instance, Koop and Korobilis, 2013) recommend to select δ in the region of $(0.94, 0.98)$. Here the decay factor is set to $\delta = 0.95$ as in Byrne et al. (2017) so that volatility information 10 periods ago receives around 63% as much weight as the last period information.

[4]The reader is referred to Gribisch (2016) for a detailed description of WSV properties and the corresponding influence of Wishart parameters A, ν and d on the dynamic characteristics of volatilities Σ_t.

[5]The online appendix provided by Byrne et al. (2017) describes the shape of $f(\eta_t\eta_t')$ and details on the Bayesian Kalman filter for approximations with decay/forgetting factors.

3.3 Bayesian Posterior Analysis

3.3.1 MCMC algorithm

Statistical inference in simple state-space models (SSM) such as the baseline DNS can be carried out by standard Bayesian or classical likelihood procedures[6]. Nevertheless, the estimation of yield curve models with Wishart stochastic volatility is rather a challenging task because the associated high-dimensional likelihood integration problem is analytically intractable. Its evaluation is nontrivial since it requires integrating out the unobserved multivariate states $(\alpha_{1:T}, \beta_{1:T}, \Sigma_{1:T})$. Bayesian MCMC posterior analysis is particularly attractive for the WSV model. More precisely, Gibbs sampling is used to simulate from the joint posterior of parameters θ augmented by the set of latent states, whereas Metropolis-hastings (MH) steps are implemented to sequentially draw from conditional distributions that are not available in closed-form solution. Under some mild regularity conditions, the simulated Markov chain converges to the joint posterior distribution of model parameters:

$$p(\theta, \alpha_{1:T}, \beta_{1:T}, \Sigma_{1:T}|y_{1:T}) \propto \prod_{t=1}^{T} f_\theta(y_t|\alpha_t)\, f_\theta(\alpha_t|\alpha_{1:t-1}, \beta_t, \Sigma_t)\, f_\theta(\beta_t|\beta_{1:t-1})\, f_\theta(\Sigma_t|\Sigma_{1:t-1})\, p(\theta), \tag{3.13}$$

where $f_\theta(y_t|\alpha_t)$, $f_\theta(\alpha_t|\alpha_{1:t-1}, \beta_t, \Sigma_t)$ and $f_\theta(\beta_t|\beta_{1:t-1})$ are the measurement, factor and TVP-VAR conditional densities at time t, respectively, defined by Eqs. (3.6), (3.7) and (3.5), while $f_\theta(\Sigma_t|\Sigma_{1:t-1})$ is the conditional density associated with factor volatilities driven by the Wishart process (3.8)-(3.9). The joint prior density is denoted by $p(\theta)$ and will be discussed in Section 3.3.2 while the exact expression of the joint posterior (3.13) is provided in Appendix 3.B.

It is worth noting that the tractable linkage between linear and nonlinear states in the factor WSV setting is a key feature that allows us to express the complete factor structure as a conditionally linear and Gaussian SSM. In other words, conditional on nonlinear non-Gaussian factor covariances $\Sigma_{1:T}$, the SSM for $y_{1:T}$ with states $\alpha_{1:T}$ and $\beta_{1:T}$ is linear-Gaussian such that linear states can be immediately simulated from standard Kalman-filter algebra like the Carter and Kohn (1994) simulation smoother. However, the conditional posterior for $\Sigma_{1:T}$ is not available in closed-form and requires a simulation smoother to nonlinear non-Gaussian SSM (see Carlin et al., 1992). This can be considered as a drawback of the WSV model compared to alternative Wishart structures such as the EWMA estimator and Uhlig-like models (see Windle et al., 2014; Kleppe et al., 2021), whereas the conditional posterior for Wishart factor covariances is available in analytical formulae such that computational shortcuts can be derived.

One major concern regarding MCMC efficiency within the WSV framework is that

[6]For a detailed discussion on estimation of state-space models, see Durbin and Koopman (2012).

of slow convergence to stationary conditions and slow mixing of the simulated Markov chain when applied to highly persistent latent variables such as factor covariances $\Sigma_{1:T}$ (Gribisch, 2016)[7]. This implies that samples will typically be autocorrelated within a Markov chain, which increases uncertainty in posterior estimates. Mitigation of this issue is here twofold. First, a relatively high number of Gibbs replications are generated to ensure low numerical standard errors. Second, we only keep every tenth draw after burn-in to guarantee relatively independent posterior thinned samples.

3.3.2 Parameter restrictions and prior choices

The factor WSV model (3.6)-(3.9) consists of multivariate latent states $(\alpha_{1:T}, \beta_{1:T}, \Sigma_{1:T})$ with given initial conditions $(\alpha_0, \beta_0, \Sigma_0)$ and unknown parameters $(\lambda, \Omega, Q, A, \nu, d)$. In this empirical application the following parameter restrictions will be deliberately imposed. First, the decay parameter λ in the measurement equation (3.1) will be treated as known within the MCMC procedure, which is a standard assumption in the literature[8]. Faster MCMC convergence has been achieved by setting λ at the optimal value from fitting the baseline DNS using the one-step SSM approach suggested by Diebold et al. (2006). At second instance, the measurement and TVP-VAR covariances Ω and Q are assumed to be diagonal matrices, which reduces estimation uncertainty in applications with a large number p of observed yields. Next, the inequality $\nu > k$ ensures that Σ_t^{-1} is invertible while $d \in [0, 1)$ is a suitable assumption to financial data applications because it guarantees weak stationarity of the logarithmic determinant of the Wishart process and rule out stochastic processes for Σ_t^{-1} which alternate between powers of inverses (Gribisch, 2016). Finally, it is appealing to set the persistence parameter d to a fixed value since it eliminates a non-desirable sensitivity between the draws of A and d in the Markov chain (see, Philipov and Glickman, 2006b). The choice of $d = 0.99$ revealed to achieve the most steady and speedy MCMC convergence with a relatively small number of burn-in draws (around $5,000$), which implies a strong persistence of factor volatilities Σ_t. Under this set of restrictions the parameter set is given by $\theta = (\Omega, Q, A, \nu)$ so that WSV becomes more parsimonious yet maintaining its flexibility to capture structural instabilities in recession periods and generate good density forecasts.

Starting values for the latent states $(\alpha_{1:T}, \beta_{1:T}, \Sigma_{1:T})$ are set using a training sample $y_{-T_0:-1}$ with sample size T_0. The corresponding initial distribution for $(\alpha_0, \beta_0, \Sigma_0)$ is assumed to be Gaussian with parameters set by training sample estimates as well. These

[7] Multi-move samplers to simulate $\Sigma_{1:T}$ at once could also mitigate the problem of autocorrelation and slow mixing in Markov chains. However, multi-move samplers such as the efficient importance sampling (EIS) technique (see, e.g., Richard and Zhang, 2007; Liesenfeld and Richard, 2008) and the Particle Gibbs method of Andrieu et al. (2010) revealed to be practically infeasible within WSV-like settings (see Gribisch, 2016). Hence, in this paper the single-move sampler of Carlin et al. (1992) to nonlinear non-Gaussian SSM is implemented to sample $\Sigma_{1:T}$ for all periods t separately.

[8] For instance, Bianchi et al. (2009) and Byrne et al. (2017) set $\lambda = 0.0609$ according to Diebold and Li (2006) such that C_t has the largest impact on the 30-month Treasury maturity.

estimates are obtained by applying the baseline DNS model (3.2)-(3.3) using the two-step approach via OLS suggested by Diebold and Li (2006). To be precise, OLS estimates $\alpha_0 = (\Lambda_0'\Lambda_0)^{-1}\Lambda_0' y_{T_0}$ are first obtained by computing Λ_0 with $\lambda = 0.0609$. From there, let us define a $T_0 \times (k+1)$ matrix $X_0 = (l_{T_0}, \alpha_{-T_0+1:0})$ where l_{T_0} denotes the T_0-vector of ones and $\alpha_{-T_0+1:0}$ is the $T_0 \times k$ matrix of yield curve factors computed via OLS for $T_0+1, T_0+2, \ldots, 0$. Next, the initial condition for TVP-VAR coefficients is set using $\beta_0 = (X_0'X_0)^{-1}X_0'\alpha_{-T_0:-1}$ with covariance $P_0^\beta = h_0 \otimes (X_0'X_0)^{-1}$, where $h_0 = \frac{(\alpha_{-T_0+1:0} - X_0\beta_0)'(\alpha_{-T_0+1:0} - X_0\beta_0)}{T_0 - K}$. Starting values for $\beta_{1:T}$ are assumed to be fixed for all t such that $\beta_{1:T} = \beta_0$. Lastly, the initial covariance of α_0 is set to $P_0^\alpha = h_0$ as well as $\Sigma_0 = h_0$ and all starting values $\Sigma_{1:T} = h_0$.

Prior choices will be fairly loose and based on training sample estimates so that data information prevails over any subjective prior information and estimation results are not significantly impacted. That said, the joint prior distribution chosen for the model parameters (Ω, Q, A, ν) follows:

- The measurement covariance Ω is assumed to follow an inverse Wishart distribution with $\Omega \sim i\mathcal{W}(\underline{\nu}_\epsilon, \underline{S}_\epsilon)$, where $\underline{\nu}_\epsilon = p + 1$ and $\underline{S}_\epsilon = \text{diag}\left\{\frac{(y_{T_0} - \alpha_0\Lambda_0')'(y_{T_0} - \alpha_0\Lambda_0)}{T_0 - 1}\right\}$.

- The covariance matrix of TVP-VAR coefficients Q is assumed to follow an inverse Wishart distribution with $Q \sim i\mathcal{W}(\underline{\nu}_e, \underline{S}_e)$, where $\underline{\nu}_e = T_0$ and $\underline{S}_e = \text{diag}\{q_0 \times P_0^\beta \times T_0\}$. It is worth emphasizing that the scaling factor q_0 is extremely relevant to determine the amount of time-variation allowed for TVP-VAR parameters β_t and largely impacts overall MCMC efficiency. Here the calibrated value $q_0 = 10^{-3}$ captures a smooth evolution of TVP-VAR parameters over recession periods (see Figure 3.6) while maintaining MCMC efficiency.

- The inverse of the intertemporal sensitivity matrix A^{-1} is assumed to follow an uninformative Wishart distribution with $A^{-1} \sim \mathcal{W}(\underline{\nu}_A, \underline{S}_A)$, where $\underline{\nu}_A = k + 1$ and $\underline{S}_A = I_k$.

- The d.o.f. parameter ν is specified according to a Gamma distribution with $\nu \sim \mathcal{G}(\underline{\alpha}_\nu, \underline{\beta}_\nu)$. Prior hyperparameters are centered at the mean of initial Gibbs draws (after convergence) obtained under an uninformative prior so that faster MCMC convergence and improved numerical efficiency can be achieved in the out-of-sample forecasting exercise.

3.3.3 Gibbs sampling

After initialization, the Gibbs sampler designed for the proposed WSV model, using $S = S_0 + S_1$ replications with S_0 burn-in draws, can be summarized in the following updating steps:

1. Sample $\alpha_{1:T}$ from the conditional posterior distribution $f_\theta(\alpha_{1:T}|\cdot)$ using the Carter and Kohn (1994) simulation smoother.

2. Sample Ω from the conditional posterior distribution $p(\Omega|\cdot)$.

3. Sample $\beta_{1:T}$ from the conditional posterior distribution $f_\theta(\beta_{1:T}|\cdot)$ using the Carter and Kohn (1994) simulation smoother. Stability conditions of β_t are ensured at each point $t = 1, \ldots, T$[9].

4. Sample A^{-1} from the conditional posterior distribution $p(A^{-1}|\cdot)$ using an independence chain MH algorithm step.

5. Sample ν from the conditional posterior distribution $p(\nu|\cdot)$ using a random walk MH algorithm step.

6. Compute the acceptance ratio $\alpha^*(A^{-1}, \nu)$ and decide on the acceptance-rejection, not treating A^{-1} and ν independently.

7. Sample $\Sigma_{1:T-1}^{-1}$ from the conditional posterior distribution $f_\theta(\Sigma_{1:T-1}^{-1}|\cdot)$ using the MH-based simulation smoother suggested by Carlin et al. (1992). It is worth noting that Σ_T^{-1} can be directly sampled from a closed-form conditional posterior $f_\theta(\Sigma_T^{-1}|\cdot)$.

After burn-in cycles S_0 are discarded, a sample average of the S_1 Gibbs replications are used to compute Bayesian point estimates of the joint posterior (3.13) and assess parameter uncertainty. Conditional posterior densities of parameters θ are illustrated in Appendix 3.C while details on the smoother algorithm that simulates dynamic factor covariances $\Sigma_{1:T}$ are presented in Appendix 3.D. The Bayesian MCMC scheme is implemented on a 2016 MacBook Pro with a 2.6GHz Intel Core i7 processor and 16GB RAM using MATLAB (version R2020b). One loop of the Gibbs sampler only takes 0.25 seconds so that the proposed MCMC algorithm is fast and easy to implement.

3.4 Empirics

3.4.1 Data

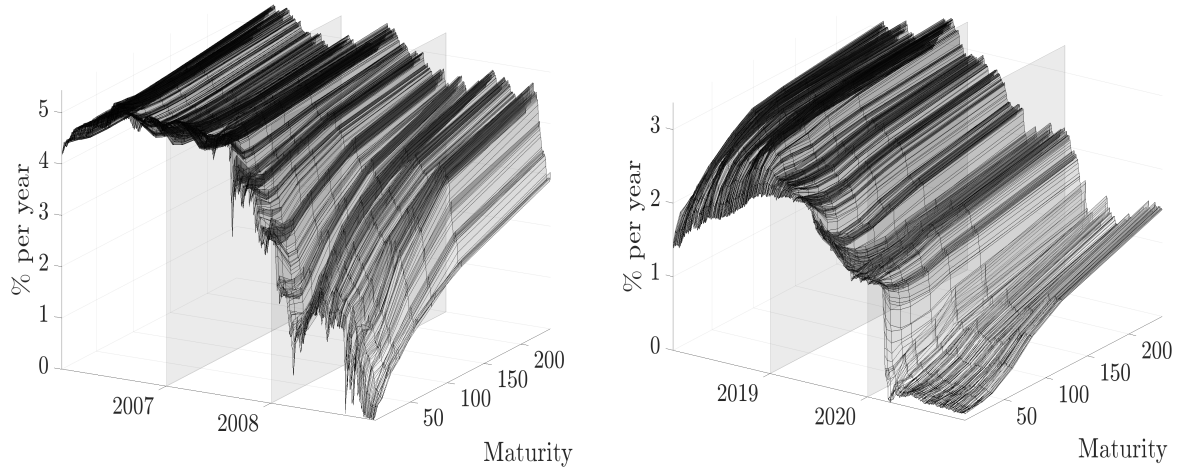
The data set consists of daily US zero-coupon Treasury yields at the following $p = 9$ constant maturities: $\tau = 3, 6, 12, 24, 36, 60, 84, 120$ and 240 months. To increase the robustness of findings, the empirical application is based on two sample periods of substantial financial distress: (i) a Great Recession period ranging from Jan 3, 2006 until Dec 31, 2008 and (ii) a COVID-19 Recession period ranging from Jan 2, 2018 until Dec

[9]Each Gibbs iteration considers a maximum number of 100 sampling attempts to ensure stability of the VAR(1) process in the factor equation (3.7) for each $t = 1, \ldots, T$. Otherwise, previous draws $\beta_{1:T}$ are kept for the next Gibbs round.

31, 2020. The predictive analysis using a rolling-window scheme is carried out over the $N = 251$ business days of 2008 and 2020, which both characterize evaluation periods with high financial uncertainty. The estimation period of the initial rolling-window spans over the 251 and 250 observations of 2007 and 2019, respectively, while previous information from 2006 and 2018 comprise the training samples $y_{-T_0:-1}$ to initialize parameters and latent states in each yield curve model. Figure 3.2 displays the three-dimensional time series plot of the US term structure of interest rates and provides an illustration of the recursive estimation window for $t = 1 : T$. The yields data on US Treasuries is retrieved from Datastream (Thomson Reuters database).

The time series in Figures 3.1-3.2 reveal a strong co-movement of yields at different maturities, especially for the COVID-19 data set, which suggests a common factor structure in yields data. However, some interesting differences can be highlighted between both predictive samples. The substantial time variation in term structure shapes and the high volatility at short-maturity rates observed throughout the financial events of 2008 do not apply to the COVID-19 period due to the nature of a pandemic shock, except for the sudden decline in rates encompassing the moment at which the WHO declares the novel coronavirus outbreak a global pandemic. Moreover, in contrast to the upward-sloping and concave term structure shown in 2008, a flattening yield curve has become more evident in recent times following Quantitative Easing (QE) programs that turn the Federal Reserve into an active participant at the long end of the term structure. Hence, these predictive data sets provide appealing and distinguishing features to perform portfolio allocation and VaR forecasting.

Figure 3.2: Daily US term structure of interest rates at constant maturities



(a) Great Recession sample period (b) COVID-19 sample period

Notes: The figure reports the 3D yield curve along the 3- up to the 240-month maturity. The vertical rectangles respectively divide the datasets into training sample, in-sample, and out-of-sample periods.

Descriptive statistics for yields at each data set are given in Table 3.4 of Appendix 3.A while Figure 3.4 provides a better inspection of the sample mean and standard deviation at different maturities. As expected, time-averaged yields increase with maturity, consistent with the existence of liquidity preferences and a higher risk premium at the long end of the curve. In contrast, the year before 2008 events displayed a slightly inverted yield curve, already signaling a potential economic downturn ahead. Furthermore, the upper panel of Figure 3.4 reveals that the mean of yields suffers a substantial drop in predictive samples compared to estimation data sets, which shows that recession times trigger a negative response of yields across all maturities anticipating Fed's efforts to stimulate the economy. The negative shift in interest rates during both predictive samples is such that the short end of the curve displays minimum rates close to the zero lower bound.

As for the volatility in the average term structure displayed in the lower panel of Figure 3.4, short-maturity rates are usually more volatile than yields at the long end since the latter can be viewed as a simple average of expected short-rates in the future. However, such a pattern does not apply to the COVID-19 estimation sample, whereas volatility increases with maturity. In addition, volatility associated with short-term interest rates is higher in the predictive samples due to higher uncertainty about the Fed's decisions following the huge shocks that hit the US economy over 2020. Finally, volatility associated with yield curve factors is higher in the Great Recession predictive sample and COVID-19 estimation data set.

Table 3.4 also reports negative skewness and excess kurtosis for yields across most maturities and data sets – typical of financial data. The time series of yields and yield curve factors feature high autocorrelation as well, which declines with the lag order and is smaller in predictive data sets compared to estimation samples. All in all, these systematic patterns help motivate the use of yield curve forecasting models in practice, especially in recession periods with sharp volatility changes.

3.4.2 Estimation results

Bayesian MCMC estimation of the WSV model proposed in Section 3.2.2 is based on 20,000 Gibbs iterations discarding the first $15,000$ draws as burn-in. Therefore, saving every tenth draw after burn-in gives us a posterior thinned sample composed of 500 replications. Table 3.5 presents posterior estimates and MCMC diagnostics of key model parameters $A^{\frac{1}{2}}$ and ν for the final rolling-windows used in the forecasting exercise. Estimation results are fairly stable across both sample periods. Together with a persistence parameter of $d = 0.99$, posterior mean estimates slightly greater than one for diagonal elements in $A^{\frac{1}{2}}$ indicate the presence of strong volatility clustering. On the other hand, posterior mean estimates close to zero for non-diagonal elements in $A^{\frac{1}{2}}$ suggest non-significant co-movements across yield factors.

Figure 3.5 in Appendix 3.E.1 depicts the smoothed mean trajectories followed by yield curve factors α_t and the factor volatilities Σ_t over both predictive samples. It becomes clear that large volatility clusters build up from the 3rd quarter of 2008 and the 1st quarter of 2020 while staying relatively low in the remaining quarters, thus characterizing samples with different volatility regimes. The beginning of both high volatility regimes exactly coincides with the most dramatic events of 2008 - the failure of Lehman Brothers - and the moment COVID-19 was declared a global pandemic.

These rare events translated into sharp declines in the yield curve level and substantial movements in the slope and curvature factors, making the bond market riskier. The slope factor significantly decreased at the time both shocks hit the US economy, which indicates a stronger upward-sloping term structure due to higher risk premia. The curvature factor exhibits a higher volatility rate across yield factors, and thus medium-term maturities are relatively more impacted in times of financial distress. This seems reasonable since the curvature factor is mostly related to business cycle movements and carries a higher relative shock variance (see Diebold et al., 2006). Moreover, higher volatility levels were captured during the financial events of 2008 compared to the pandemic period. Ultimately, little evidence of spillover effects across yield factors has been detected since factor correlations are consistently around zero although some co-movements intensify on a small scale when overall volatility builds up. In summary, observed yields are fairly informative about yield curve factors whereas the flexible WSV framework can track the dynamics of factor volatilities quite well, although factor correlations reveal to be insignificant and constant over time.

To address the numerical accuracy of the MCMC algorithm, low numerical standard errors reported in Table 3.5 indicate a very small variation induced by the simulation-based scheme while variation induced by the data is much larger, even though posterior uncertainty measured via standard deviation is also very low across all parameters. Furthermore, some convergence information about the Gibbs chains is provided by the CD diagnostic suggested by Geweke (1992), whereas the long-run variance estimate robust to serial correlation follows Newey and West (1987) with $L = 20$ lags. Since the CD statistics are all smaller than 1.96 in absolute values[10], the chosen number of Gibbs iterations is sufficiently large and results do not indicate convergence issues toward the stationary distribution.

To evaluate the sampling efficiency of the MCMC procedure, the effective sample size (ESS) measure is computed. It is defined as $\text{ESS} = S_1 \left[1 + 2\sum_{l=1}^{L} \hat{\rho}_l\right]^{-1}$, where $\sum_{l=1}^{L} \hat{\rho}_l$ represents the L monotone sample autocorrelations within the Markov chain of a parameter (Geyer, 1992)[11]. The interpretation is that the $S_1 = 500$ MCMC posterior draws for

[10] This includes the covariances Ω and Q, which are not reported in Table 3.5 but can be provided by request.

[11] Here ESS measures are directly obtained from the Newey and West (1987) estimator with $L = 20$ lags computed for the CD convergence statistics.

a parameter lead to the same numerical precision as a hypothetical iid sample from the posterior of size ESS so that large ESS values close to S_1 are to be preferred (Kleppe et al., 2021). The reported ESS values for elements in $A^{\frac{1}{2}}$ range from 17 to 41 while ν displays smaller ESS values between 7 and 9. This indicates a reasonable sampling efficiency of the MCMC procedure for $A^{\frac{1}{2}}$ parameters but slow mixing for the degrees of freedom ν due to significant autocorrelation of up to 20 lags within the chains[12]. Hence a thinning strategy of MCMC posterior samples helps to reduce within-chain autocorrelation.

Finally, to highlight the importance of the TVP feature to capture parameter instability, Fig. 3.6 illustrates the smooth time evolution of TVP-VAR parameters estimated at the last rolling-window of each recession sample. Coefficients associated with the slope factor exhibit a substantial time variation with shifting dynamics that might indicate sudden structural changes, which is not surprising since the long-short spread is commonly associated with economic activity dynamics in the literature. Parameters concerning the curvature factor also display a time-varying pattern to some degree while structural instability is less likely to occur for VAR parameters of the level factor.

3.4.3 Forecasting exercise

Forecasting with the DNS-WSV model

To construct point and density forecasts for future yields, the Gibbs sampler outlined in Section 3.3.3 provides a straightforward solution. The one-step-ahead predictive density for yields y_{t+1} is given by

$$p(y_{t+1}|y_{1:t}) = \int f_\theta(y_{t+1}|\alpha_{t+1})\, f_\theta(\alpha_{t+1}|\alpha_t, \beta_{t+1}, \Sigma_{t+1})\, f_\theta(\beta_{t+1}|\beta_t)\, f_\theta(\Sigma_{t+1}|\Sigma_t) \times p(\theta, \alpha_{1:t}, \beta_{1:t}, \Sigma_{1:t}|y_{1:t})\, d\alpha_{1:t+1}\, d\beta_{1:t+1}\, d\Sigma_{1:t+1}\, d\theta. \quad (3.14)$$

Hence, conditional on data up to time t, uncertainty in the forecasts comes from parameter uncertainty, uncertainty about current states, and uncertainty about future shocks. The predictive density (3.14) can be easily approximated by direct MC integration, but it is worth noting that such approximations require a full Gibbs sampling update of the full posterior of state variables and parameters at each evaluation sample with data up to $t = \{T, T+1, \ldots, T+N-1\}$. Computation of log-predictive density scores, $\text{LPDS} = \sum_{t=T}^{T+N-1} \log p(y_{t+1}|y_{1:t})$, motivated in Geweke and Amisano (2010) follow straightforwardly using the MC approximations to $p(y_{t+1}|y_{1:t})$ and indicates how well a model performs in density forecasting.

[12]It is also worth emphasizing that the MH candidate distributions described in Appendix 3.C deliver acceptance rates for the sampling of $A^{\frac{1}{2}}$ and ν that vary between 0.2 and 0.25. The mean MH acceptance rate for smoothing of factor volatilities Σ_t amounts to 0.1, which indicates a fairly good performance of the single-move algorithm to sample Σ_t taking into account the fact that single-move algorithms typically deliver a low acceptance probability.

Under quadratic loss, the point and variance forecasts of y_{t+1} for a given θ follow

$$\begin{aligned} \bar{y}_{t+1} &= \mathrm{E}(y_{t+1}|y_{1:t},\theta) \quad = \Lambda \mathrm{E}(\alpha_{t+1}|y_{1:t},\theta), && (3.15)\\ \bar{F}_{t+1} &= \mathrm{Var}(y_{t+1}|y_{1:t},\theta) = \Lambda \mathrm{Var}(\alpha_{t+1}|y_{1:t},\theta)\Lambda' + \Omega, && (3.16)\\ \bar{P}_{t+1} &= \mathrm{Var}(\alpha_{t+1}|y_{1:t},\theta) = \mathrm{E}(\Upsilon_t \bar{P}_t \Upsilon_t' + \Sigma_t|y_{1:t},\theta) && (3.17) \end{aligned}$$

where the predictive moments $\mathrm{E}(y_{t+1}|y_{1:t},\theta)$ and $\mathrm{Var}(y_{t+1}|y_{1:t},\theta)$ are approximated by direct MC integration while posterior mean estimates are used as parameter values for θ. Note that the factor structure allows us to decompose the forecast uncertainty (3.16) into a first component that captures estimation uncertainty about yield curve factors (common market risk factors) while Ω accounts for uncertainty of future unforeseen shocks in the term structure (idiosyncratic risk factors)[13]. For alternative non-Bayesian specifications, one can use the fact that (3.2)-(3.3) implies a conditional normal distribution for y_{t+1} given information up to period t such that $y_{t+1} \sim \mathcal{N}(\bar{y}_{t+1}, \bar{F}_{t+1})$. Out-of-sample validation tests can be performed based on the predictive Pearson residuals at any maturity τ_i

$$\xi_{it+1} = \left[y_{t+1}(\tau_i) - \mathrm{E}(y_{t+1}(\tau_i)|y_{1:t},\theta)\right] / \mathrm{Var}(y_{t+1}(\tau_i)|y_{1:t},\theta)^{\frac{1}{2}}, \quad i = 1,\ldots,p \tag{3.18}$$

where point and variance forecasts follow from (3.15)-(3.16). If the predictive model is valid, then ξ_{it+1} has mean zero and unit variance, and ξ_{it+1} as well as ξ_{it+1}^2 are serially uncorrelated.

Forecasting performance

To investigate the value added of each modeling feature when yield curve forecasting and portfolio analysis are of major interest, let us consider a variety of DNS models. In a forecasting horserace, the out-of-sample forecasting performance of the proposed DNS-WSV model will be compared to the following alternative specifications:

- The popular benchmark in the forecasting literature – the naive random walk (RW). Independent random walks without drift components are specified for each yield-to-maturity:

$$y_t(\tau_i) = y_{t-1}(\tau_i) + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}_p(0, \sigma_i^2), \qquad \text{with } i = 1,\ldots,p. \tag{3.19}$$

- The baseline DNS model with time-invariant parameters using the two-step estimation approach via OLS suggested by Diebold and Li (2006).

[13] See Durbin and Koopman (2012) for further details on the derivation of (3.15)-(3.16).

- The DNS with simple random walk stochastic volatility (SV) given by

$$
\begin{aligned}
y_t &= \Lambda\alpha_t + \epsilon_t, && \epsilon_t \sim \mathcal{N}(0, \Omega) \\
\alpha_t &= \mu + \Upsilon\alpha_{t-1} + e^{\frac{1}{2}h_t}\,\eta_t, && \eta_t \sim \mathcal{N}_k(0, 1) \qquad (3.20) \\
h_t &= h_{t-1} + u_t, && u_t \overset{i.i.d.}{\sim} \mathcal{N}_k(0, \sigma_h^2) \qquad (3.21)
\end{aligned}
$$

  where h_t denotes a k-vector of factor log-volatilities that evolve according to independent random walk processes, whereas u_t is a random variable with $\sigma_h^2 = \text{diag}\{\sigma_{h,1}^2, \ldots, \sigma_{h,k}^2\}$[14]. This model allows us to assess any improvement by considering time-varying correlations.

- The DNS with EWMA estimator discussed in Section 3.2.2, which provides a good benchmark of Wishart factor covariances with posterior analytical formulae.

Hence, the proposed WSV framework will be tested against the following competitors: RW, baseline DNS, SV, and EWMA. Moreover, the relevance of parameter instability will be incorporated via TVP-VAR coefficients following Eq. (3.5) in the proposed WSV, SV, and EWMA structures. Bayesian estimation of SV and EWMA factor volatilities follow a straightforward simplification of the MCMC procedure described in Section 3.3, whereas SV uses the auxiliary mixture sampler proposed by Kim et al. (1998) and EWMA applies standard Kalman filtering to the forgetting factor structure (see Koop and Korobilis, 2013)[15]. These computational shortcuts available for SV and EWMA imply a much faster MCMC convergence so that 1,500 Gibbs iterations with 500 burn-in draws (and saving every 2nd draw after burn-in) are sufficient for each forecasting window.

Table 3.1 reports the one-step-ahead point forecast precision of the competing models using the root mean squared error (RMSE) measure relative to the RW benchmark. Lower RMSE values indicate better performance in general, for a particular maturity τ_i, while values lower than one indicate outperformance relative to RW forecasts. It becomes clear that RW is a tough competitor since it consistently outperforms across the entire maturity spectrum. This result is indeed well documented in the literature (see, among recent papers, Kleppe et al., 2021; Carriero et al., 2021).

Preferred RMSEs excluding RW are also displayed in bold while grey-shaded cells provide information on rejection of the null hypothesis of equal predictive accuracy relative to such preferred model using the Giacomini and White (2006) test that handles comparison of nested models estimated under a Bayesian framework. Hence whiter columns in Table 3.1 are associated with higher overall performance. In many cases, Wishart stochastic

[14] Bianchi et al. (2009), Hautsch and Yang (2012) and Shin and Zhong (2017) apply the SV extension to the DNS model but allowing factor log-volatilities h_t to evolve according to stationary AR(1) processes.

[15] More precisely, Gibbs steps (4)-(7) illustrated in 3.3.3 are replaced by the respective filtering and smoothing algorithms for SV and EWMA.

volatility models deliver fairly good predictive performance, especially at the short end of the term structure and to some extent at longer maturities. The introduction of TVP-VAR parameters appears to significantly increase estimation imprecision for the Great Recession period, leading to less accurate point forecasts. On the other hand, there is little evidence of such underperformance of TVP-VAR specifications during the COVID-19 period.

To sum up, it is hard to beat RW point forecasts on a daily forecasting exercise, even though these differences are not significant in some cases and become smaller over the COVID-19 period. More precisely, RW forecasts outperform preferred yield curve models with average gains up to 23.2% during the financial crisis of 2008, whereas overall performance increases with maturity. On the other hand, relative RW average gains equal 10.7% throughout the pandemic with yield curve models performing better at the short end of the curve.

Table 3.1: RMSE and LPDS: competing models relative to the RW benchmark

	Panel A: Great Recession predictive sample							**Panel B: COVID-19 predictive sample**						
					Time-Varying Parameters							Time-Varying Parameters		
τ (months)	RW*	DNS	SV	WSV	EWMA	SV	WSV	RW*	DNS	SV	WSV	EWMA	SV	WSV
3	0.134	1.732	2.610	2.603	**1.634**	2.612	2.606	0.035	**1.010**	1.119	1.092	1.061	1.251	1.257
6	0.108	1.218	1.079	**1.061**	1.243	1.124	1.137	0.033	1.043	**1.001**	1.019	1.056	1.057	1.093
12	0.093	1.651	1.308	**1.296**	1.631	1.346	1.369	0.030	1.544	1.336	1.357	**1.309**	1.325	1.351
24	0.106	**1.172**	1.516	1.502	1.460	1.590	1.589	0.033	1.432	1.036	1.039	**0.999**	1.073	1.104
36	0.108	**1.309**	1.599	1.597	1.882	1.734	1.736	0.037	1.168	**1.101**	1.106	1.156	1.269	1.315
60	0.106	1.199	**1.034**	1.043	1.910	1.141	1.137	0.043	**1.162**	1.461	1.470	1.705	1.669	1.713
84	0.098	1.401	**1.240**	1.261	2.338	1.356	1.341	0.049	1.088	1.077	1.071	1.202	**1.069**	1.084
120	0.088	1.233	**1.064**	1.080	2.106	1.093	1.094	0.055	1.327	**1.119**	**1.119**	1.597	1.182	1.206
240	0.076	**1.275**	2.075	2.080	1.854	2.435	2.560	0.063	1.246	2.042	2.047	**1.198**	2.009	1.996
LPDS	1,825	**1,887**	1,747	1,717	1,384	1,672	1,683	3,852	4,629	4,252	4,227	**4,813**	4,200	4,202

Notes: The table reports the RMSE of one-step ahead forecasts for each competing model relative to the random walk (RW) using an evaluation sample of $N = 251$. * Absolute RMSE values are provided for the RW model. Preferred RMSE values (other than RW) for each maturity are given in **bold** while results for the Giacomini and White (2006) test of equal predictive accuracy relative to such preferred model are indicated by grey-shaded cells (significance at the 5% level). Absolute LPDS values for each model are provided in the last row with preferred values in **bold**.

As for density forecasts, Table 3.1 reports the LPDS for both predictive samples. In general, the baseline DNS model provides the most accurate density forecasts over the Great Recession while EWMA-TVP outperforms during the pandemic. Figure 3.7 in Appendix 3.E.2 exhibits the period-wise LPDS, and thus provides a closer inspection of the relative likelihood associated with every density forecast. The picture reveals that an identical accuracy is achieved by SV and WSV specifications throughout both predictive samples, as well as their TVP versions. These models also deliver log scores that are somewhat more stable and less sensitive to the extreme financial events that deteriorate overall performance on impact. In particular, SV and WSV better pick up the contemporaneous effects of these shocks on yields volatility, but after a few days, the constant volatility DNS has picked up enough of a rise in volatility that it resumes yielding better relative predictive scores[16].

[16]This systematic pattern is better illustrated in Figure 3.8 whereas estimated volatilities of SV and WSV specifications,

It is also worth noting that DNS factor structures offer a consistent and sizable advantage over RW's density accuracy since the latter falls behind most quarters, especially after the pandemic hits the US economy. Hence the superiority of RW in point forecasting does not replicate here. Moreover, including Wishart stochastic volatility via EWMA-TVP slightly improves the accuracy of density forecasts relative to the constant volatility DNS during Covid. The bond portfolio optimization and VaR forecasting application coming next aims to practically assess whether the aforementioned statistical gains in density accuracy lead to sizable economic gains.

Validation tests

Figure 3.9 in Appendix 3.E.3 summarizes all information regarding out-of-sample diagnostic checks based on the predictive Pearson residuals ξ_{it+1}. It is clear that DNS specifications with stochastic volatility (with TVP to a larger extent) tend to overpredict future yields. This pattern happens on a larger scale over the Great Recession whereas residuals ξ_{it+1} are generally close to zero throughout COVID-19 for most models, except EWMA-TVP which produces substantial differences from the zero mark for medium-maturity rates. Turning to the standard deviations of ξ_{it+1} under a correctly specified model, the proposed models capture quite well yields variation in the 2008 crisis. However, in the pandemic standard deviations are relatively inferior to the benchmark value of 1, especially at short- and medium-term maturities for SV and WSV specifications. Hence, these models overpredict yields volatility during the COVID-19 crisis.

Lastly, p-values of the Ljung-Box test with 10 lags suggest significant autocorrelations left in ξ_{it+1} and ξ^2_{it+1} during the pandemic and thus confirm the aforementioned pattern in standard deviations. To be precise, DNS factor models fail to fully capture the observed dynamics in US Treasury yields and the associated volatility in the 2020 crisis period. Nonetheless, in 2008 most DNS models with stochastic volatility efficiently accounted for yields dynamics in levels and volatility - mainly long-term maturities predicted by SV and WSV structures. Similar patterns have been documented in Kleppe et al. (2021) for the 2008 crisis period.

3.4.4 Bond portfolio optimization

Bond log-returns and portfolio problem

The DNS framework is designed to model bond yields at selected maturities, bond portfolio optimization, however, requires estimates of expected bond returns at each maturity, as well as their forecasted covariance matrix. Caldeira et al. (2016) describe how to obtain expressions for the expected return of bonds and their covariance matrix based on point

including their TVP counterparts, react quickly and profoundly to the huge shocks that affect yields predictability.

and variance forecasts of yields. Using the mathematical expression for the discount curve $P_t(\tau_i) = \exp(-\tau_i \cdot y_t(\tau_i))$ and the log-return expression, the realized return of holding a \$1 bond from t to $t+1$, while its maturity decreases from τ_i to τ_{i-1}, can be computed as follows

$$\begin{aligned} r_{t+1}(\tau_i) &= \log\left(\frac{P_{t+1}(\tau_{i-1})}{P_t(\tau_i)}\right) = \log P_{t+1}(\tau_{i-1}) - \log P_t(\tau_i) \\ &= -\tau_{i-1} \cdot y_{t+1}(\tau_{i-1}) + \tau_i \cdot y_t(\tau_i), \end{aligned} \tag{3.22}$$

where $\tau_{i-1} = (\tau_1 - 1, \tau_2 - 1, \ldots, \tau_p - 1)'$. Note that (3.22) represents the change in market prices of constant maturity bonds.

From there, let us use the conditional Gaussian structure of the DNS model to reasonably assume that the p-vector of one-step-ahead forecasts of bond log-returns $\hat{R}_{t+1} = (\hat{r}_{t+1}(\tau_1), \ldots, \hat{r}_{t+1}(\tau_p))'$ follows a Normal distribution with $\mathcal{N}(\bar{R}_{t+1}, \bar{\sigma}^2_{R,t+1})$ where

$$\bar{R}_{t+1} = -\tau_{i-1} \circ \bar{y}_{t+1}(\tau_{i-1}) + \tau_i \circ y_t(\tau_i). \tag{3.23}$$

The point forecast $\bar{y}_{t+1}(\tau_{i-1})$ is available from (3.15) (with a small adjustment to the p-vector τ_{i-1} of outstanding bond maturities at $t+1$) while $y_t(\tau_i)$ is known at t. The conditional covariance of the expected log-returns, which only depends on the stochastic part of $\bar{R}_{t+1}$, simply follows

$$\begin{aligned} \bar{\sigma}^2_{R,t+1} &= \mathrm{E}\left[(-\tau_{i-1} \circ \bar{y}_{t+1}(\tau_{i-1}))(-\tau_{i-1} \circ \bar{y}_{t+1}(\tau_{i-1}))'\right] \\ &= \tau_{i-1}\tau'_{i-1} \circ \bar{F}_{t+1}(\tau_{i-1}), \end{aligned} \tag{3.24}$$

where $\bar{F}_{t+1}(\tau_{i-1})$ is the variance forecast in (3.16) adjusted to maturities τ_{i-1}. Since holding a τ_i-maturity bond at t is the same as holding a τ_{i-1}-maturity bond after one period, both $\bar{y}_{t+1}(\tau_{i-1})$ and $\bar{F}_{t+1}(\tau_{i-1})$ must be adjusted to maturities τ_{i-1} because point and variance forecasts (3.15)-(3.16) are done using the selected maturity spectrum $\tau_1, \tau_2, \ldots, \tau_p$[17].

It is worth noting that the conditional covariance in (3.24) reflects the price volatility of selected maturity bonds, which for zero-coupon bonds can be read as duration times yields volatility. Yields on short-term bonds are typically more volatile than longer-term yields[18], nonetheless, this difference is generally smaller than the duration gap. Therefore,

[17] A simple approach that might be helpful to overcome this issue related to bond aging is the interpolation of forecasted yields $\bar{y}_{t+1}$ in (3.15) for the maturity spectrum $\tau_1 - 1, \tau_2 - 1, \ldots, \tau_p - 1$. Similarly, the adjustment of the covariance matrix $\bar{F}_{t+1}(\tau_{i-1})$ to maturities $\tau_{i-1}\tau'_{i-1}$ is done via interpolation, where original forecasts (3.16) relate to the maturity spectrum $\tau_i\tau'_i$. Moreover, the Schur product theorem ensures that the element-wise multiplication between positive elements $\tau_{i-1}\tau'_{i-1}$ and $\bar{F}_{t+1}(\tau_{i-1})$ is also positive-definite $\forall t$. Note that the smaller the frequency of the data (e.g., monthly yields) and the longer the forecast horizon, the more important such adjustment turns out to be because forecasted yields get closer to time-to-maturity.

[18] Monetary policy mainly operates in the short end of the yield curve whereas near future expectations about the business cycle are more volatile than long-term expectations. On the other hand, long-term bond yields reflect inflation expectations averaged over their time to maturity plus a maturity premium that comprises future uncertainty about the state of the economy; both factors do not fluctuate as radically as short-horizon expectations. All in all, short yield-to-maturity

long-term bondholders face a higher price volatility since they are more exposed to the duration factor - they must endure inflation uncertainty and interest-rate risk for a longer time horizon.

Once we are provided with the key ingredients (3.23)-(3.24) to perform a bond portfolio analysis, we apply the mean-variance portfolio problem formulated by Markowitz (1952) to derive the optimal portfolio choice. To this end, it is assumed that investors allocate their wealth to risky assets based on the trade-off between expected bond log-returns and risk associated with a decrease in their market prices. In particular, the p-vector of optimal portfolio weights w_t at time t is computed by minimizing the portfolio variance for a particular vector of expected bond log-returns:

$$\min_{w_t} \quad w_t' \, \bar{\sigma}^2_{R,t+1} \, w_t - \frac{1}{\gamma} \, w_t' \, \bar{R}_{t+1} \quad \text{subject to} \quad w_t' \, \imath = 1 \text{ and } w_t \geq 0, \tag{3.25}$$

where $\imath$ is an appropriately sized vector of ones and γ is the investor's risk aversion coefficient. Thus, the optimization problem is subject to non-negative individual portfolio weights and the budget constraint[19].

Evaluating portfolio performance

Implementation of the portfolio problem (3.25) here takes into account three investment scenarios based on different risk aversion coefficients: aggressive ($\gamma = 0.1$), moderate ($\gamma = 20$), and conservative ($\gamma = 100$). Portfolio analysis will then consist of computing optimal weights w_t at each recursive out-of-sample window such that a daily portfolio rebalancing is performed and w_t is constantly updated for each investment scenario.

To evaluate the performance of optimal portfolios we use some usual criteria in finance. First, the average excess return relative to the risk-free rate (1-month Treasury rate) indicates how profitable the mean-variance portfolio strategy can be compared to constantly holding the risk-free asset[20]. Second, the Sharpe ratio provides a risk-adjusted measure of the portfolio's excess return, whereas the ex-post portfolio's volatility is given by the standard deviation of excess returns. Ultimately, the portfolio's average duration indicates how much exposure to yield changes the portfolio carries.

Table 3.2 reports the performance of mean-variance portfolios using density forecasts of competing models while Figure 3.3 provides an easier-to-grasp comparison. Wishart modeling strategies deliver higher excess returns and Sharpe ratios in the aggressive investment scenario with portfolios mostly invested in long-term maturities. To be precise,

is generally more volatile than yields on long-term bonds.

[19]The former constraint represents a restriction on short positions while the latter ensures that the investor's wealth is fully invested in risky bonds with outstanding maturities $\tau_1, \tau_2, \ldots, \tau_p$ by the time of decision-making.

[20]It is worth mentioning that computation of excess returns disregards any transaction costs, which is a reasonable assumption these days. Hence the empirical analysis can simply overlook the impact of the portfolio's turnover on excess returns.

the EWMA-TVP strategy exhibits 0.142% daily excess returns in the 2008 crisis period - and about 0.12% adjusted per unit of exposed volatility - while WSV-TVP carries an average of 0.10% excess returns throughout the pandemic - and 0.085% adjusted to risk. In the moderate investment scenario, RW forecasts achieve the best overall performance in terms of Sharpe ratio for both periods, nevertheless, similar average returns are achieved by EWMA-TVP in the Great Recession while DNS specifications consistently outperform RW's returns by a great margin in 2020. Lastly, stochastic volatility models also dominate in the conservative scenario with portfolios mostly allocated in short- and medium-term maturities, although the differences in excess returns among all models are quite small. In particular, the high-risk aversion coefficient ($\gamma = 100$) yields the best performance on a risk-adjusted basis across all investment scenarios even though excess returns are much smaller, meaning that sharp declines in the portfolio's total risk offset diminishing returns.

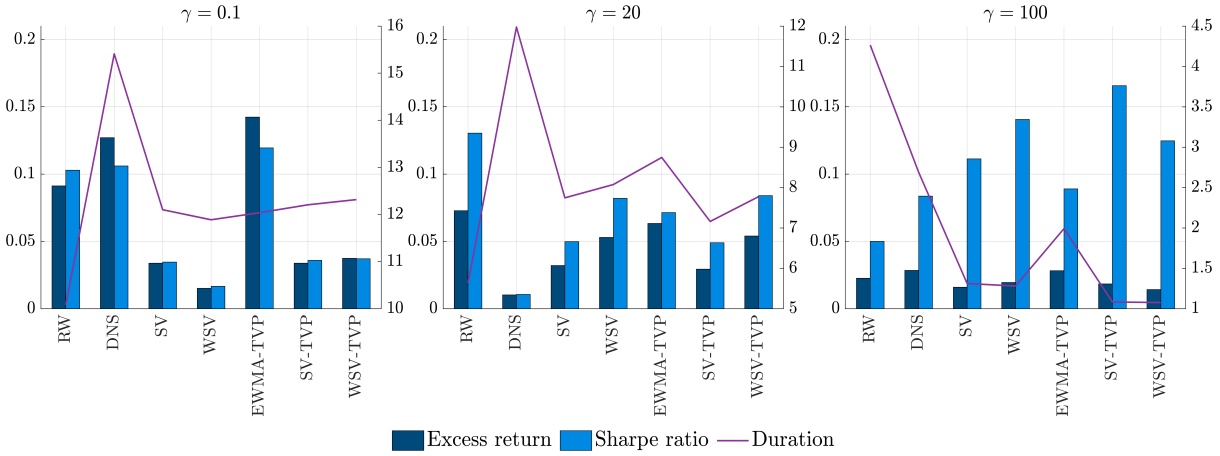
Table 3.2: Performance of optimal bond portfolios with daily rebalancing

	Panel A: Great Recession predictive sample							**Panel B: COVID-19 predictive sample**						
					Time-Varying Parameters							Time-Varying Parameters		
	RW	DNS	SV	WSV	EWMA	SV	WSV	RW	DNS	SV	WSV	EWMA	SV	WSV
Aggressive investment scenario ($\gamma = 0.1$)														
Excess Return	0.091	0.127	0.033	0.015	**0.142**	0.033	0.037	0.078	0.069	0.083	0.095	0.093	0.100	**0.101**
Sharpe Ratio	0.103	0.106	0.034	0.016	**0.119**	0.035	0.037	0.061	0.057	0.069	0.080	0.081	0.084	**0.085**
Duration	10.07	15.41	12.10	11.88	12.03	12.20	12.31	20.00	17.99	16.62	15.90	14.43	15.10	14.61
Moderate investment scenario ($\gamma = 20$)														
Excess Return	**0.072**	0.010	0.032	0.052	0. 063	0.029	0.054	0.027	0.086	0.108	0.108	0.105	**0.122**	0.097
Sharpe Ratio	**0.130**	0.010	0.049	0.082	0.071	0.049	0.084	**0.160**	0.076	0.107	0.095	0.140	0.125	0.086
Duration	5.62	11.97	7.74	8.07	8.74	7.15	7.77	4.15	13.37	12.98	13.60	10.69	11.95	12.50
Conservative investment scenario ($\gamma = 100$)														
Excess Return	0.022	**0.028**	0.016	0.019	**0.028**	0.018	0.014	**0.031**	0.020	0.021	0.030	0.020	0.024	0.020
Sharpe Ratio	0.050	0.083	0.111	0.140	0.089	**0.165**	0.124	0.158	0.123	0.149	**0.199**	0.130	0.197	0.163
Duration	4.26	2.68	1.31	1.27	1.99	1.08	1.07	4.58	3.86	3.05	1.34	3.67	2.36	2.19

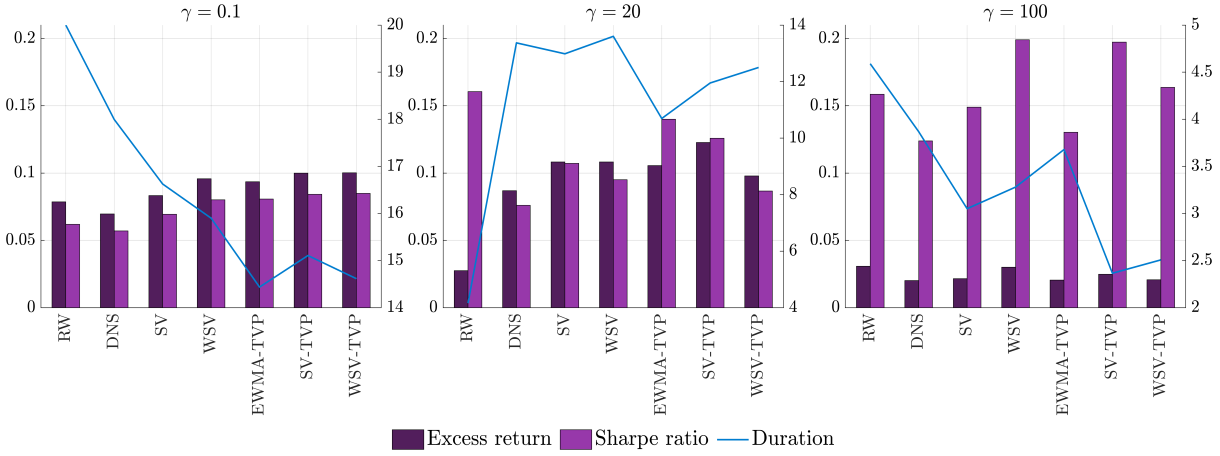
Notes: The table reports performance statistics for optimal portfolios using one-step ahead forecasts of each competing model. Daily average excess returns (percent) are computed using the 1-month Treasury bill as the risk-free rate while duration is measured in years. Higher (preferred) portfolio's excess returns and Sharpe ratios are given in **bold**.

Figure 3.10 in Appendix 3.E.3 illustrates period-wise cumulative compounded excess returns for DNS competing models. In general, the picture reinforces the overall superiority of EWMA-TVP in most investment scenarios. Although the equally weighted portfolio outperforms all mean-variance strategies at some point prior to the 3rd quarter of 2008, it is clear that under the aggressive scenario, the baseline DNS and EWMA-TVP significantly outperform in the last quarter of 2008, delivering 3 times more gains than the equally weighted portfolio. On the other hand, the DNS factor structures substantially dominate the equally weighted portfolio throughout the COVID-19 period - in particular, the aggressive and moderate investors allocated in stochastic volatility strategies can respectively achieve annual excess returns from 25% up to 30% on average compared to 6.5% when allocated in the equally weighted portfolio. Thereby, a more aggressive investment profile can highly benefit from using factor Wishart stochastic volatility to

Figure 3.3: Performance of optimal bond portfolios under alternative investment scenarios



(a) Great Recession evaluation period



(b) COVID-19 evaluation period

Notes: The figure reports the performance of optimal portfolios in terms of daily average excess returns (percent), Sharpe ratio, and duration (in years) for the Great Recession (upper panel) and COVID-19 (lower panel) predictive samples.

exploit the sharp declines in bond yields during turbulent times. Altogether, it pays off to account for time-varying volatility in bond markets - especially via Wishart strategies - in order to construct smart investment decisions in terms of cumulative excess returns as well as risk-adjusted performance.

3.4.5 Value-at-risk forecasting and backtesting

One major application of volatility models in the context of bond markets is the computation of VaR forecasts to measure the risk associated with unexpected rises in bond yields. To this end, the VaR measure quantifies the potential loss in value of a risky portfolio over a defined time interval at a given confidence level. Hence, the VaR measure is a simple quantile on the conditional distribution of expected bond portfolio returns. Its practical

relevance, ease of implementation, and convenient backtesting established the VaR as the most popular method in financial risk forecasting to aid in decision-making and managing market risk[21].

To map the randomness in volatile interest rates to the risk in a bond, let us first denote by $\bar{Q}_{t+1} = w_t' \bar{R}_{t+1}$ the expected bond portfolio log-return at time $t+1$. The bond portfolio VaR at time t for a one-step-ahead holding period and confidence level ϑ, hereafter denoted as VaR_t^{ϑ}, is then given by the ϑ-quantile of the conditional distribution of the expected bond portfolio log-return:

$$\text{VaR}_t^{\vartheta} = F_{t+1}^{-1}(\vartheta), \tag{3.26}$$

where $F_{t+1}^{-1}(\vartheta)$ is the inverse of the cumulative distribution function of $\bar{Q}_{t+1}$. Using the assumption that forecasted bond log-returns $\hat{R}_{t+1}$ are conditionally normally distributed[22], the portfolio VaR can also be computed as a function of the conditional moments (3.23)-(3.24). More precisely,

$$\text{VaR}_t^{\vartheta} = \bar{Q}_{t+1} + q \,.\, \bar{\sigma}_{Q,t+1}, \tag{3.27}$$

where $\bar{\sigma}_{Q,t+1}^2 = w_t' \, \bar{\sigma}_{R,t+1}^2 \, w_t$ denotes the portfolio conditional covariance and q is the $(1-\vartheta)$-quantile of the standard normal distribution since unexpected bond returns are here assumed to be iid with zero mean and unit variance. The application focuses on daily one-step-ahead VaR forecasts at confidence levels $\vartheta = 90\%, 95\%, 99\%$, and thus there is a $(1-\vartheta)$ probability that bond portfolio returns will fall below VaR_t^{ϑ} over a one-business-day holding period.

To evaluate the performance of VaR forecasts, backtesting compares one-step-ahead VaR_t^{ϑ} forecasts with realized bond portfolio returns at $t+1$, hereby $Q_{t+1} = w_t' R_{t+1}$, where R_{t+1} is the p-vector of realized bond returns. Backtesting is then based on the hit sequence that corresponds to the VaR violation indicator variable $I_{t+1} = \mathbb{1}\left[Q_{t+1} \leq \text{VaR}_t^{\vartheta}\right]$. Hence, if the VaR_t^{ϑ} forecast exceeds the actual portfolio return Q_{t+1}, then a VaR violation is said to have occurred.

The behavior of the hit sequence of I_{t+1}'s is of major interest in risk management and, equally importantly, whether these VaR violations are clustered in time and have the correct coverage such that the expectation of I_{t+1} equals the $(1-\vartheta)$ probability. The problem of assessing the accuracy of predicted VaR can be reduced to a few statistical tests for VaR backtesting. First, we implement the proportion of failures (POF) test, which is a likelihood-ratio (LR) test proposed by Kupiec (1995) to assess whether the proportion of VaR violations is synchronized with the implied probability $(1-\vartheta)$ under the null. Second, the conditional coverage independence (CCI) LR test proposed by

[21] See Nieto and Ruiz (2016) for a survey on recent developments in VaR forecasting and backtesting.

[22] Although not reported here, the Gaussian distribution indeed provides a reasonable approximation to the empirical distribution of forecasted yields.

Christoffersen (1998) checks whether the elements of the hit sequence are independent of each other over time. Hence it tests for serial independence of VaR failures between consecutive days against the alternative of first-order Markov dependence.

Table 3.3 reports the hit rate for optimal portfolios under the aggressive, moderate, and conservative investment scenarios, whereas preferred values are displayed in bold. Turning to VaR statistical accuracy, grey-shaded cells provide information on the rejection of the CCI test while rejections of the POF test are indicated by the † symbol. Overall, stochastic volatility models reveal the best performance for both predictive samples. This result is even more evident for the aggressive and moderate investment scenarios, signaling that stochastic volatility plays an important role for risk-seeking investors. In particular, SV and WSV models consistently produce the most accurate VaR forecasts for the 2008 crisis period while Wishart stochastic volatility models outperform during the pandemic. Furthermore, the larger the defined VaR level ϑ, the more accurate the VaR predictions prove to be. Hence, Wishart strategies overcome the typical concern of VaR models failing during structural breaks. Lastly, it is worth emphasizing that the VaR predicted by RW and the constant volatility DNS are generally imprecise and significantly underpredict the risk over 2008, leading to much higher VaR violations than expected from a valid VaR model.

Table 3.3: Value-at-Risk forecasting results for optimal portfolios.

	Panel A: Great Recession predictive sample							**Panel B: COVID-19 predictive sample**						
					Time-Varying Parameters							Time-Varying Parameters		
	RW	DNS	SV	WSV	EWMA	SV	WSV	RW	DNS	SV	WSV	EWMA	SV	WSV
Aggressive investment scenario ($\gamma = 0.1$)														
1%	2.78†	3.18†	**0.39**	**0.39**	1.99	0†	**0.39**	1.59	1.59	0.79	**1.19**	**1.19**	0.79	**1.19**
5%	7.96†	11.95†	7.96†	7.17	10.35†	**6.77**	7.96†	3.18	3.98	5.57	5.97	3.18	4.38	**5.17**
10%	**13.94**†	25.49†	14.74†	14.74†	17.52†	14.34†	19.12†	4.78†	12.74	13.54	15.13†	**8.76**	11.95	13.14
Moderate investment scenario ($\gamma = 20$)														
1%	7.96†	3.58†	**0.79**	0.39	1.99	0.39	**0.79**	**1.19**	1.59	0.79	**1.19**	.0.39	0.79	**1.19**
5%	15.13†	11.95†	7.56	**6.77**	10.75†	**6.77**	9.16†	2.39†	3.98	5.57	5.97	3.18	4.38	**5.17**
10%	18.72†	25.89†	**14.34**†	15.53†	18.72†	**14.34**†	19.92†	3.98†	11.95	13.14	14.74†	**9.16**	11.55	13.14
Conservative investment scenario ($\gamma = 100$)														
1%	7.56†	4.78†	0.39	0.39	2.39	0.39	**0.79**	**1.19**	**1.19**	0†	0†	0.39	0†	0†
5%	14.34†	9.96†	1.59†	2.78†	8.36†	1.59†	**3.58**	**2.78**	1.99†	0.39†	0†	1.99†	0†	0†
10%	16.33†	18.72†	4.38†	**6.37**†	14.74†	2.39†	4.78†	4.38†	3.18†	0.79†	1.19†	**5.17**†	0.39†	0.39†

Notes: The table reports the hit rate (percent) for the 1%, 5% and 10% VaR forecasts. Preferred values (closest observed failures to expected VaR level) are given in **bold**. Results for the conditional coverage independence (CCI) test are indicated by grey-shaded cells (significance at the 5% level). Results for the proportion of failures (POF) test are indicated by the symbol † (5% level).

As for serial independence of VaR failures, preferred $\text{VaR}_t^{99\%}$ pass the CCI test for the Great Recession period although forecasts for the 5% and 10% VaR using stochastic volatility usually fail when aggressive and moderate scenarios are considered. This suggests that multiple VaR violations can occur in a short span (violation clustering), signaling that such VaR models might react slowly to changing market conditions. In fact, VaR violations might not be spreading out over time due to the high dependence between

the volatility episodes that characterize both recession periods. In contrast, independence of VaR failures is not an issue for investors with a conservative profile mainly exposed to the short- and medium-maturity spectrum of the term structure, however, VaR forecasts are less accurate in such investment scenarios. Turning to constant volatility models, RW and DNS can generate independent VaR failures in most cases, but they regularly fail the POF test due to highly inaccurate VaR forecasts. A similar pattern, as described above, can be observed for the COVID-19 period, however predicted VaR of EWMA-TVP passed the CCI test across all scenarios considered. Altogether, there is strong evidence that stochastic volatility VaR models better estimate the risk in volatile bond markets - especially for risk-taking portfolios via SV and WSV models in 2008, including their TVP versions, and via EWMA-TVP in 2020. Nevertheless, VaR backtesting occasionally identifies some problems with the independence assumption on hit events, which suggests that one VaR violation is more likely to follow other violations in subsequent days.

3.5 Concluding remarks

The increasing volatility observed in bond markets, especially during the financial events of 2008 and the COVID-19 pandemic, has opened up good opportunities in portfolio allocation but also made risk management a more challenging task. This paper further extends a large literature on yield curve modeling by introducing Wishart stochastic volatility into the DNS factor structure to better capture structural instability and improve density forecasts in turbulent times. Specifically, the study exploits the role of Wishart stochastic volatility as a key ingredient for improving density forecasts of the yield curve and how this translates into outperformance in terms of portfolio allocation and risk management. As for estimation, a one-step Bayesian MCMC framework is developed for evaluating the integrated likelihoods of factor models with stochastic volatility and TVP-VAR features.

Using data on daily US Treasury yields, the results produced in the empirical analysis have interesting implications and provide useful guidelines for practitioners, being twofold. First, the DNS factor structure along with Wishart stochastic volatility improves the accuracy of density forecasts from a no-change random walk - especially during highly volatile periods - although random walk point forecasts represent a hard-to-beat benchmark. Second, we find overwhelming support for Wishart strategies in terms of optimal mean-variance portfolio performance and VaR forecasting. These findings become even more evident for scenarios that consider risk-seeking investors, mainly exposed to medium- and long-term maturity bonds, whereas higher excess returns and Sharpe ratios can be achieved in addition to better estimates for interest rate risk. It is worth noting that most of these economic gains come from allowing for a Wishart structure in the factor

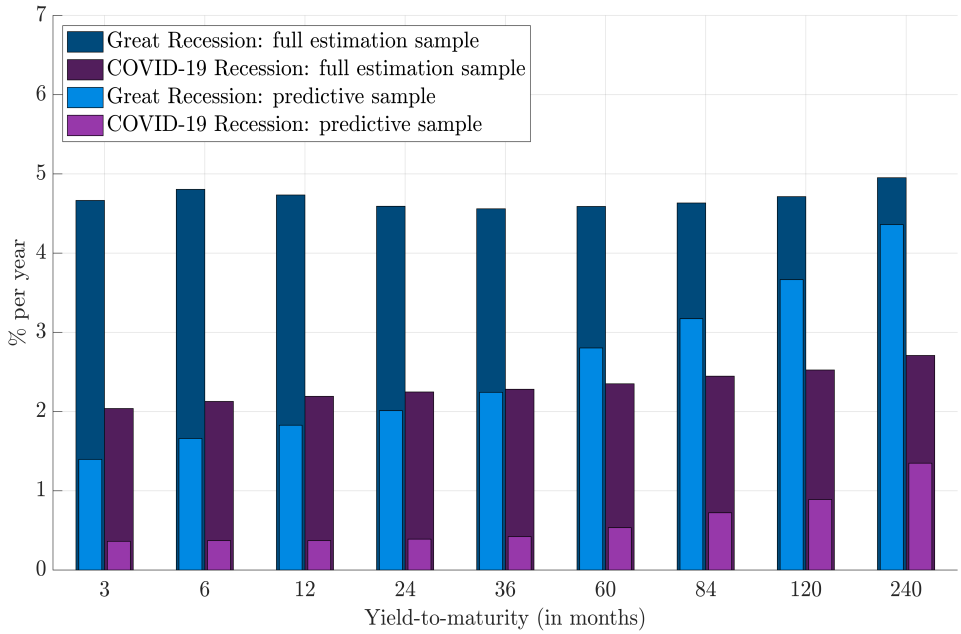
volatilities rather than time variation in the VAR coefficients, which in turn only delivers marginal benefits. Hence, bond fund managers can indeed benefit from incorporating Wishart stochastic volatility into the factor dynamics that govern the term structure of interest rates.

Nevertheless, there is still some room for improvement. Validation tests have shown that the proposed models tend to overpredict future yields level and volatility under quadratic loss and fail to fully capture the underlying volatility at short-term maturities. As possibilities for future research, one could evaluate the impact of asymmetric loss functions on the empirical results provided hereby. Moreover, high-frequency “nowcasts" of macroeconomic conditions could be incorporated into the factor structure of the proposed yield curve model. Daily nowcasts of inflation and business cycle dynamics comprise the natural channel that could link macroeconomic risks to yield curve movements, whereas Wishart processes provide a flexible and robust structure to capture dynamic interrelationships and correlations among yield-macro factors.

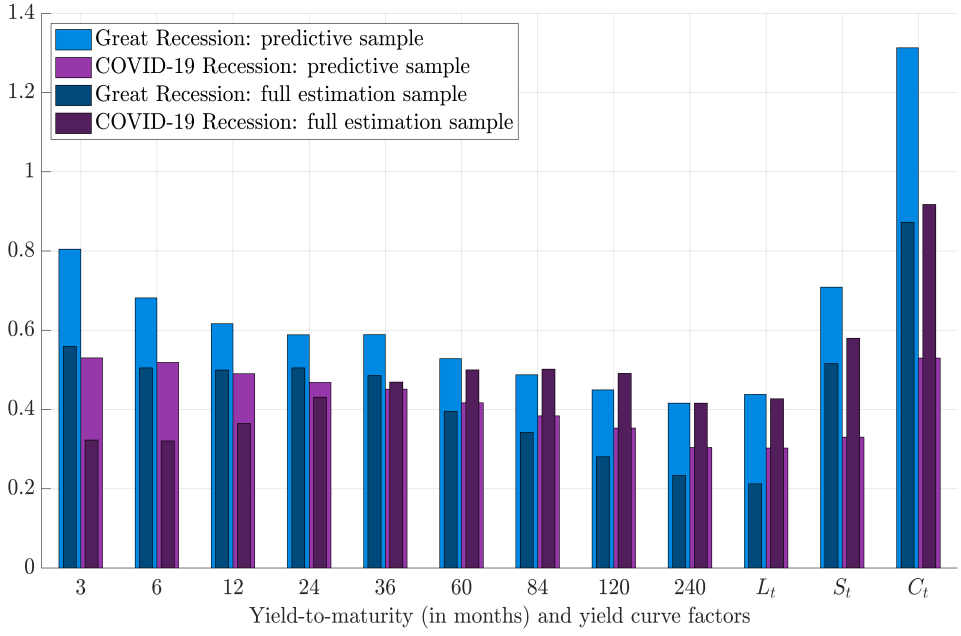
Appendix 3

3.A Descriptive statistics of US Treasury yields

Figure 3.4: Sample mean and standard deviation of US Treasury yields



(a) Sample mean



(b) Standard deviation

Notes: The figure reports the sample mean (upper panel) and standard deviation (lower panel) of daily Treasury yields at constant maturities and yield curve Level (L_t), Slope (S_t) and Curvature (C_t) factors for the full estimation (training + in-sample) and predictive samples of the Great Recession and COVID-19 Recession selected periods. The yield curve factors are extracted from the baseline DNS model using the estimation procedure suggested by Diebold and Li (2006).

Table 3.4: Descriptive statistics of US Treasury yields

Panel A: Great Recession									**Panel B: COVID-19 Recession**								
τ	Mean	Std. Dev.	Min.	Max.	Skewness	Kurtosis	$\hat{\rho}(1)$	$\hat{\rho}(10)$	τ	Mean	Std. Dev.	Min.	Max.	Skewness	Kurtosis	$\hat{\rho}(1)$	$\hat{\rho}(10)$
Full estimation period from Jan 3, 2006 until Dec 31, 2007									Full estimation period from Jan 2, 2018 until Dec 31, 2019								
3	4.66	0.55	2.87	5.19	-1.63	4.82	0.98	0.84	3	2.03	0.32	1.39	2.49	-0.25	1.81	0.99	0.92
6	4.80	0.50	3.17	5.33	-1.68	5.00	0.98	0.87	6	2.12	0.32	1.55	2.58	-0.28	1.82	0.99	0.92
12	4.73	0.49	3.09	5.30	-1.75	5.30	0.98	0.87	12	2.19	0.36	1.51	2.74	-0.34	1.81	0.99	0.93
24	4.59	0.50	2.90	5.29	-1.69	5.42	0.98	0.86	24	2.24	0.43	1.39	2.98	-0.35	1.83	0.99	0.94
36	4.55	0.48	2.91	5.26	-1.65	5.51	0.98	0.85	36	2.28	0.46	1.34	3.05	-0.42	1.85	0.99	0.94
60	4.58	0.39	3.23	5.23	-1.30	4.84	0.98	0.84	60	2.35	0.49	1.32	3.09	-0.53	1.88	0.99	0.95
84	4.63	0.34	3.48	5.23	-0.97	4.22	0.98	0.83	84	2.44	0.50	1.40	3.18	-0.57	1.93	0.99	0.95
120	4.71	0.28	3.83	5.26	-0.39	3.22	0.97	0.81	120	2.52	0.49	1.47	3.24	-0.64	2.05	0.99	0.95
240	4.95	0.23	4.29	5.44	0.06	2.47	0.97	0.80	240	2.70	0.41	1.76	3.37	-0.70	2.29	0.99	0.94
L_t	4.91	0.21	4.46	5.39	0.24	2.26	0.97	0.78	L_t	2.72	0.42	1.72	3.34	-0.75	2.36	0.99	0.93
S_t	-0.04	0.51	-1.61	0.73	-0.75	2.96	0.98	0.83	S_t	-0.63	0.57	-1.72	0.51	-0.04	1.73	0.99	0.95
C_t	-1.10	0.87	-4.10	0.32	-1.13	4.32	0.96	0.80	C_t	-0.60	0.91	-2.58	0.77	-0.17	1.77	0.99	0.94
Predictive period from Jan 2, 2008 until Dec 31, 2008									Predictive period from Jan 2, 2020 until Dec 31, 2020								
3	1.39	0.80	0	3.27	-0.09	2.70	0.97	0.71	3	0.36	0.53	0	1.59	1.74	4.16	0.98	0.80
6	1.66	0.68	0.14	3.32	-0.35	3.32	0.96	0.69	6	0.37	0.51	0.02	1.58	1.77	4.26	0.98	0.80
12	1.82	0.61	0.34	3.17	-0.58	3.12	0.96	0.69	12	0.37	0.49	0.09	1.56	1.79	4.37	0.98	0.80
24	2.01	0.58	0.65	3.05	-0.53	2.44	0.97	0.77	24	0.38	0.46	0.11	1.58	1.77	4.38	0.98	0.80
36	2.24	0.58	0.88	3.38	-0.42	2.43	0.97	0.78	36	0.42	0.45	0.10	1.61	1.75	4.39	0.98	0.80
60	2.80	0.52	1.26	3.73	-1.04	3.88	0.96	0.72	60	0.53	0.41	0.19	1.67	1.76	4.51	0.97	0.79
84	3.17	0.48	1.59	3.95	-1.47	5.03	0.96	0.67	84	0.72	0.38	0.36	1.79	1.70	4.46	0.97	0.77
120	3.66	0.44	2.08	4.27	-1.97	6.75	0.96	0.59	120	0.88	0.35	0.52	1.88	1.60	4.34	0.97	0.75
240	4.35	0.41	2.86	4.86	-2.16	7.38	0.96	0.58	240	1.34	0.30	0.87	2.21	1.25	3.83	0.96	0.71
L_t	4.63	0.43	2.98	5.22	-2.29	8.04	0.96	0.52	L_t	1.38	0.30	0.84	2.20	1.01	3.31	0.95	0.70
S_t	-2.99	0.70	-4.64	-0.87	0.68	4.72	0.96	0.63	S_t	-0.84	0.33	-1.66	0.09	0.87	3.61	0.97	0.70
C_t	-3.83	1.31	-6.66	-1.33	-0.21	2.09	0.96	0.73	C_t	-2.16	0.53	-3.13	-0.91	0.21	2.04	0.96	0.74

Notes: The table presents descriptive statistics - Mean, Standard Deviation (Std. Dev.), Minimum (Min), Maximum (Max), Skewness, Kurtosis and Sample Autocorrelations $\hat{\rho}(k)$ at lag order k - for daily Treasury yields at constant maturities ranging from 3 to 240 months, and for the yield curve Level (L_t), Slope (S_t) and Curvature (C_t) factors extracted from the baseline DNS model using the estimation procedure suggested by Diebold and Li (2006).

3.B DNS-WSV likelihood and joint posterior

The likelihood function of the DNS-WSV framework (3.6)-(3.9) discussed in Section 3.2.2 is given by:

$$
\begin{aligned}
p(y_{1:T}|\theta) &= \prod_{t=1}^{T} f_\theta(y_t, \alpha_t, \beta_t, \Sigma_t | y_{1:t-1}, \alpha_{1:t-1}, \beta_{1:t-1}, \Sigma_{1:t-1}) \\
&= \prod_{t=1}^{T} f_\theta(y_t|\alpha_t)\, f_\theta(\alpha_t|\alpha_{1:t-1}, \beta_t, \Sigma_t)\, f_\theta(\beta_t|\beta_{1:t-1})\, f_\theta(\Sigma_t|\Sigma_{1:t-1}) \\
&\propto \prod_{t=1}^{T} \Bigg[|\Omega^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(y_t - \Lambda\alpha_t)'\Omega^{-1}(y_t - \Lambda\alpha_t)\right) \\
&\quad |\Sigma_t^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\alpha_t - B_t'\alpha_{t-1})'\Sigma_t^{-1}(\alpha_t - B_t'\alpha_{t-1})\right) \\
&\quad |Q^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\beta_t - \beta_{t-1})'Q^{-1}(\beta_t - \beta_{t-1})\right) \\
&\quad \frac{|S_{t-1}^{-1}|^{\frac{\nu}{2}} |\Sigma_t^{-1}|^{\frac{\nu-k-1}{2}}}{2^{\frac{\nu k}{2}}\Gamma_k(\nu)} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{t-1}^{-1}\Sigma_t^{-1}\right]\right)\Bigg],
\end{aligned} \tag{3.28}
$$

where $\Gamma_k(\nu) = \prod_{i=1}^{k}\left(\frac{\nu+1-i}{2}\right)$ with $\Gamma(\cdot)$ being the gamma function, and $y_{1:t-1}$ is suppressed for convenience in conditional notation. The vector of yield curve factors at $t-1$ is here assumed to be a $(k+1)$-vector $\alpha_{t-1} = (1, \alpha_{t-1}')'$ to account for VAR intercepts in the factor equation.

By Bayes theorem, the joint posterior distribution of the parameter set θ is proportional to the product of the likelihood (3.28) and priors defined in Section 3.3.2,

$$
\begin{aligned}
p(\theta|y_{1:T}) &\propto p(y_{1:T}|\theta)\, p(\theta) \\
&\propto |\Omega^{-1}|^{\frac{\underline{\nu}_\epsilon+p+1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\underline{S}_\epsilon\Omega^{-1}\right]\right) |Q^{-1}|^{\frac{\underline{\nu}_e+m+1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\underline{S}_e^{-1}\right]\right) \\
&\quad |A^{-1}|^{\frac{\underline{\nu}_A-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\underline{S}_A^{-1}A^{-1}\right]\right) \nu^{(\underline{\alpha}_\nu-1)} e^{-\underline{\beta}_\nu \nu} \\
&\quad \prod_{t=1}^{T} \Bigg[|\Omega^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(y_t - \Lambda\alpha_t)'\Omega^{-1}(y_t - \Lambda\alpha_t)\right) \\
&\quad |\Sigma_t^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\alpha_t - B'\alpha_{t-1})'\Sigma_t^{-1}(\alpha_t - B'\alpha_{t-1})\right) \\
&\quad |Q^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\beta_t - \beta_{t-1})'Q^{-1}(\beta_t - \beta_{t-1})\right) \\
&\quad \frac{|S_{t-1}^{-1}|^{\frac{\nu}{2}} |\Sigma_t^{-1}|^{\frac{\nu-k-1}{2}}}{2^{\frac{\nu k}{2}}\Gamma_k(\nu)} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{t-1}^{-1}\Sigma_t^{-1}\right]\right)\Bigg].
\end{aligned} \tag{3.29}
$$

3.C DNS-WSV conditional posterior densities

The corresponding conditional posterior distributions and sampling procedures are shown in this section.

1. Sample Ω from an inverse Wishart conditional posterior, $\Omega \sim i\mathcal{W}(\bar{\nu}_\epsilon, \bar{S}_\epsilon)$, where

$$\bar{\nu}_\epsilon = T + \underline{\nu}_\epsilon \quad \text{and} \quad \bar{S}_\epsilon^{-1} = \underline{S}_\epsilon + \sum_{t=1}^{T}(y_t - \Lambda\alpha_t)(y_t - \Lambda\alpha_t)'. \tag{3.30}$$

2. Sample Q from an inverse Wishart conditional posterior, $Q \sim i\mathcal{W}(\bar{\nu}_e, \bar{S}_e)$, where

$$\bar{\nu}_e = T + \underline{\nu}_e \quad \text{and} \quad \bar{S}_e^{-1} = \underline{S}_e + \sum_{t=1}^{T}(\beta_t - \beta_{t-1})(\beta_t - \beta_{t-1})'. \tag{3.31}$$

3. The conditional posterior distribution of A^{-1} is derived as follows

$$\begin{aligned}
p(A^{-1}|\Xi) &\propto p(A^{-1}) \prod_{t=1}^{T} p(\Sigma_t^{-1}|\Sigma_{1:t-1}^{-1};\theta) \\
&\propto |A^{-1}|^{\frac{\underline{\nu}_A-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\underline{S}_A^{-1}A^{-1}\right]\right) \prod_{t=1}^{T}|S_{t-1}|^{-\frac{\nu}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{t-1}^{-1}\Sigma_t^{-1}\right]\right) \\
&\propto |A^{-1}|^{\frac{\underline{\nu}_A+T\nu-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\underline{S}_A^{-1}A^{-1}\right]\right) \prod_{t=1}^{T} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{t-1}^{-1}\Sigma_t^{-1}\right]\right) \\
&\propto |A^{-1}|^{\frac{\underline{\nu}_A+T\nu-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\underline{S}_A^{-1}A^{-1} + \sum_{t=1}^{T}\nu(A^{\frac{1}{2}'})^{-1}(\Sigma_{t-1})^d(A^{\frac{1}{2}})^{-1}\Sigma_t^{-1}\right]\right)
\end{aligned}$$

The prior d.o.f. $\underline{\nu}_A$ is set to $k+1$ while the prior scale matrix $\underline{S}_A^{-1}$ equals the identity matrix $\mathbf{I}_k$. Since $p(A^{-1}|\Xi)$ does not follow a known distribution, it is appropriate to employ a MH algorithm to sample from the conditional posterior of A^{-1}. A random walk chain MH algorithm with the below proposal density is implemented:

$$(A^{-1})^* = (A^{-1})^{(s-1)} + \epsilon_A, \quad \epsilon_A \sim \mathcal{N}(0, \sigma_A^2\mathbf{I}),$$

where ϵ_A follows a symmetric matrix variate normal distribution. Note that ϵ_A has $k(k+1)/2$ different random elements, whereas each element of ϵ_A can be randomly generated from a normal distribution. The acceptance ratio is obtained as follows

$$AR = \left(\frac{|(A^{-1})^*|}{|(A^{-1})^{(s-1)}|}\right)^{\frac{(\underline{\nu}_A+T\nu-k-1)}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\underline{S}_A^{-1}((A^{-1})^* - (A^{-1})^{(s-1)}) + (\mathbf{Q}_\mathbf{A}^* - \mathbf{Q}_\mathbf{A}^{(\mathbf{s-1})})\right]\right),$$

where

$$\begin{aligned}\mathbf{Q}_{\mathbf{A}}^{*} &= \sum_{t=1}^{T} \nu((A^{*})^{\frac{1}{2}'})^{-1}(\Sigma_{t-1})^{d}((A^{*})^{\frac{1}{2}})^{-1}\Sigma_{t}^{-1} \\ \mathbf{Q}_{\mathbf{A}}^{(\mathbf{s-1})} &= \sum_{t=1}^{T} \nu((A^{(s-1)})^{\frac{1}{2}'})^{-1}(\Sigma_{t-1})^{d}((A^{(s-1)})^{\frac{1}{2}})^{-1}\Sigma_{t}^{-1}\end{aligned}$$

4. The conditional posterior distribution of ν is derived as follows

$$\begin{aligned}p(\nu|\Xi) &\propto p(\nu)\prod_{t=1}^{T} p(\Sigma_{t}^{-1}|\Sigma_{1:t-1}^{-1};\theta) \\ &\propto \nu^{(\underline{\alpha}_{\nu}-1)}e^{-\underline{\beta}_{\nu}\nu}\prod_{t=1}^{T}\frac{|S_{t-1}^{-1}|^{\frac{\nu}{2}}|\Sigma_{t}^{-1}|^{\frac{\nu-k-1}{2}}}{2^{\frac{\nu k}{2}}\Gamma_{k}(\nu)}\exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{t-1}^{-1}\Sigma_{t}^{-1}\right]\right) \\ &\propto \nu^{(\underline{\alpha}_{\nu}-1)}e^{-\underline{\beta}_{\nu}\nu}\left(\frac{|\nu A^{-1}|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}}\Gamma_{k}(\nu)}\right)^{T}\prod_{t=1}^{T}|\Sigma_{t-1}^{-1}|^{-\frac{\nu d}{2}}|\Sigma_{t}^{-1}|^{\frac{\nu-k-1}{2}}\exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{t-1}^{-1}\Sigma_{t}^{-1}\right]\right)\end{aligned}$$

The prior hyperparameters $\underline{\alpha}_{\nu} = 2$ and $\underline{\beta}_{\nu} = 0.1$ are defined. The conditional posterior $p(\nu|\Xi)$ does not follow a known distribution, and hence we need to apply MH to sample from ν. A random walk chain MH algorithm with the following proposal density is implemented:

$$\nu^{*} = \nu^{(s-1)} + \epsilon_{\nu}, \quad \epsilon_{\nu} \sim \mathcal{N}(0, \sigma_{\nu}^{2}),$$

where ϵ_{ν} follows a normal distribution. The acceptance ratio is obtained as follows

$$\begin{aligned}AR = &\left(\frac{\nu^{*}}{\nu^{(s-1)}}\right)^{(\underline{\alpha}_{\nu}-1)}e^{-\underline{\beta}_{\nu}(\nu^{*}-\nu^{(s-1)})}\left(\frac{|\nu^{*}A^{-1}|^{\frac{\nu^{*}}{2}}}{2^{\frac{\nu^{*}k}{2}}\Gamma_{k}(\nu^{*})}\right)^{T}\left(\frac{|\nu^{(s-1)}A^{-1}|^{\frac{\nu^{(s-1)}}{2}}}{2^{\frac{\nu^{(s-1)}k}{2}}\Gamma_{k}(\nu^{(s-1)})}\right)^{-T} \\ &\frac{\prod_{t=1}^{T}|\Sigma_{t-1}^{-1}|^{-\frac{\nu^{*}d}{2}}|\Sigma_{t}^{-1}|^{\frac{\nu^{*}-k-1}{2}}}{\prod_{t=1}^{T}|\Sigma_{t-1}^{-1}|^{-\frac{\nu^{(s-1)}d}{2}}|\Sigma_{t}^{-1}|^{\frac{\nu^{(s-1)}-k-1}{2}}}\exp\left(-\frac{1}{2}\operatorname{tr}\left[\mathbf{Q}_{\nu}^{*}-\mathbf{Q}_{\nu}^{(\mathbf{s-1})}\right]\right)\end{aligned}$$

where

$$\begin{aligned}\mathbf{Q}_{\nu}^{*} &= \sum_{t=1}^{T}\nu^{*}(A^{\frac{1}{2}'})^{-1}(\Sigma_{t-1})^{d}(A^{\frac{1}{2}})^{-1}\Sigma_{t}^{-1} \\ \mathbf{Q}_{\nu}^{(\mathbf{s-1})} &= \sum_{t=1}^{T}\nu^{(s-1)}(A^{\frac{1}{2}'})^{-1}(\Sigma_{t-1})^{d}(A^{\frac{1}{2}})^{-1}\Sigma_{t}^{-1}\end{aligned}$$

3.C DNS-WSV simulation smoother for Σ_t^{-1}

1. **Inverse factor covariances Σ_t^{-1} for $t = 1, ..., T-1$.** The conditional posterior distribution of Σ_t^{-1} follows

$$\begin{aligned}
p(\Sigma_t^{-1}|\Xi) &\propto p(\alpha_t|\alpha_{1:t-1}, \Sigma_t^{-1}; \theta)\, p(\Sigma_t^{-1}|\Sigma_{1:t-1}^{-1}; \theta)\, h_{t+1}(\Sigma_{t+1}^{-1}|\underline{\Sigma}_t^{-1}; \theta) \\
&\propto |\Sigma_t^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\alpha_t - B_t'\alpha_{t-1})'\Sigma_t^{-1}(\alpha_t - B_t'\alpha_{t-1})\right) \\
&\quad |\Sigma_t^{-1}|^{\frac{\nu-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{t-1}^{-1}\Sigma_t^{-1}\right]\right) |S_t|^{-\frac{\nu}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_t^{-1}\Sigma_{t+1}^{-1}\right]\right) \\
&\propto |\Sigma_t^{-1}|^{\frac{\nu-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[(S_{t-1}^{-1} + (\alpha_t - B_t'\alpha_{t-1})(\alpha_t - B_t'\alpha_{t-1})')\Sigma_t^{-1}\right]\right) \\
&\quad |\Sigma_t^{-1}|^{\frac{1-\nu d}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_t^{-1}\Sigma_{t+1}^{-1}\right]\right) \\
&\propto \mathcal{W}_k(\nu, \tilde{S}_{t-1}) \cdot |\Sigma_t^{-1}|^{\frac{1-\nu d}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_t^{-1}\Sigma_{t+1}^{-1}\right]\right)
\end{aligned}$$

   Although the first term of $p(\Sigma_t^{-1}|\Xi)$ follows a Wishart distribution of the form $\mathcal{W}_k(\nu, \tilde{S}_{t-1})$ with $\tilde{S}_{t-1} = (S_{t-1}^{-1} + \Theta_t)^{-1}$, and where $\Theta_t = (\alpha_t - B_t'\alpha_{t-1})(\alpha_t - B_t'\alpha_{t-1})'$, the conditional posterior $p(\Sigma_t^{-1}|\cdot)$ does not have a known distributional form. Hence, a MH algorithm for multivariate state-space modeling proposed by Carlin et al. (1992) is employed to sample from Σ_t^{-1} for $t = 1, ..., T-1$. An independence chain MH step with a proposal density of the form $\mathcal{W}_k(\nu, \tilde{S}_{t-1})$ is employed. The acceptance ratio is obtained as follows

$$AR = \frac{|(\Sigma_t^{-1})^*|^{\frac{1-\nu d}{2}}}{|(\Sigma_t^{-1})^{(s-1)}|^{\frac{1-\nu d}{2}}} \frac{\exp\left(-\frac{1}{2}\operatorname{tr}\left[\nu(A^{\frac{1}{2}'})^{-1}(\Sigma_t^*)^d(A^{\frac{1}{2}})^{-1}\Sigma_{t+1}^{-1}\right]\right)}{\exp\left(-\frac{1}{2}\operatorname{tr}\left[\nu(A^{\frac{1}{2}'})^{-1}(\Sigma_t^{(s-1)})^d(A^{\frac{1}{2}})^{-1}\Sigma_{t+1}^{-1}\right]\right)},$$

   which can be further simplified to

$$AR = \left(\frac{|(\Sigma_t^{-1})^*|}{|(\Sigma_t^{-1})^{(s-1)}|}\right)^{\frac{1-\nu d}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\mathbf{Q}_{\boldsymbol{\Sigma}}((\Sigma_t^*)^d - (\Sigma_t^{(s-1)})^d)\right]\right),$$

   where $\mathbf{Q}_{\boldsymbol{\Sigma}} = \nu(A^{\frac{1}{2}})^{-1}\Sigma_{t+1}^{-1}(A^{\frac{1}{2}'})^{-1}$.

2. **Last period inverse factor covariance Σ_T^{-1}.** The conditional posterior of Σ_T^{-1}, however, does not depend on the next period's volatility process $h_{T+1}(\Sigma_{T+1}^{-1}|\underline{\Sigma}_T^{-1}; \theta)$ since such density at $T+1$ is not available, and directly follows a Wishart distribu-

tion:

$$
\begin{aligned}
p(\Sigma_T^{-1}|\Xi) &\propto p(\alpha_T|\alpha_{1:t-1}, \Sigma_T^{-1}; \theta)\, p(\Sigma_T^{-1}|\Sigma_{1:t-1}^{-1}; \theta) \\
&\propto |\Sigma_T^{-1}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\alpha_T - B_T'\alpha_{T-1})'\Sigma_T^{-1}(\alpha_T - B_T'\alpha_{T-1})\right) \\
&\quad |\Sigma_T^{-1}|^{\frac{\nu-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[S_{T-1}^{-1}\Sigma_T^{-1}\right]\right) \\
&\propto |\Sigma_T^{-1}|^{\frac{(\nu+1)-k-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left[(S_{T-1}^{-1} + (\alpha_T - B_t'\alpha_{T-1})(\alpha_T - B_t'\alpha_{T-1})')\Sigma_T^{-1}\right]\right) \\
&\propto \mathcal{W}_k(\nu+1, \tilde{S}_{T-1}),
\end{aligned}
$$

with $\tilde{S}_{T-1} = (S_{T-1}^{-1} + \Theta_T)^{-1}$.

3.E Supplementary results

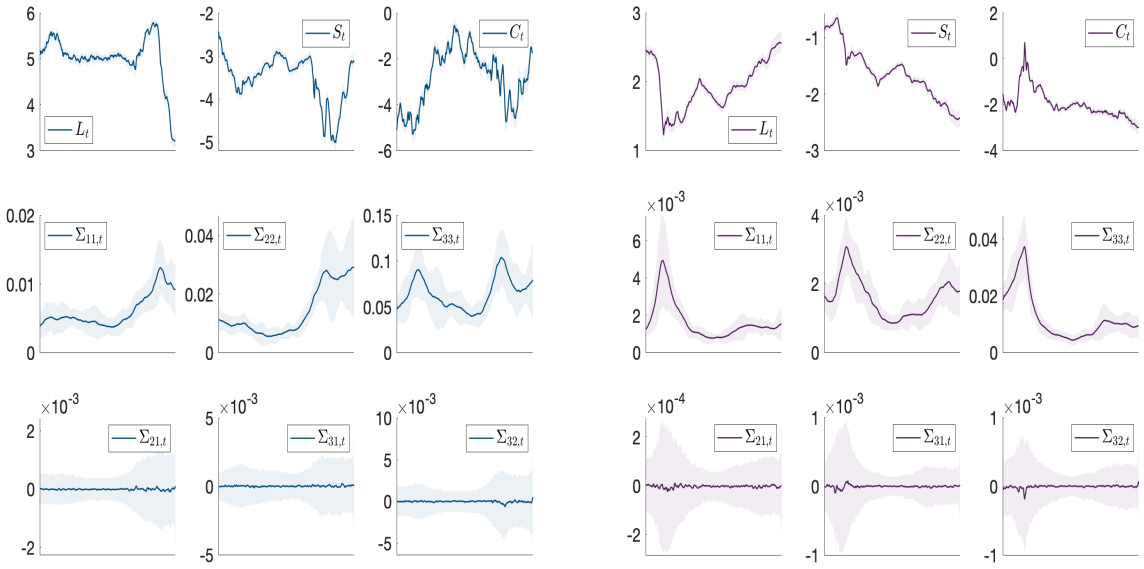
3.E.1 DNS-WSV posterior results

Table 3.5: MCMC posterior estimates of Wishart parameters in DNS-WSV

	Panel A: Dec 31, 2007 - Dec 29, 2008					**Panel B: Jan 2, 2020 - Dec 30, 2020**				
	Mean	Std Dev (×100)	NSE (×100)	ESS	CD	Mean	Std Dev (×100)	NSE (×100)	ESS	CD
$a_{11}^{\frac{1}{2}}$	1.0257	0.254	0.016	20.334	1.868	1.0350	0.258	0.016	22.119	-0.071
$a_{21}^{\frac{1}{2}}$	-0.0014	0.345	0.015	41.601	-0.848	-0.0062	0.331	0.016	34.961	0.795
$a_{31}^{\frac{1}{2}}$	0.0006	0.352	0.017	36.969	-0.180	-0.0018	0.338	0.018	33.790	-0.743
$a_{22}^{\frac{1}{2}}$	1.0227	0.274	0.023	17.217	1.161	1.0344	0.265	0.016	22.209	1.066
$a_{32}^{\frac{1}{2}}$	-0.0049	0.335	0.016	36.838	-0.128	-0.0065	0.340	0.017	36.324	0.125
$a_{33}^{\frac{1}{2}}$	1.0148	0.254	0.014	23.973	-1.655	1.0260	0.270	0.019	20.150	0.958
ν	343.2172	0.079	0.004	7.268	0.504	343.2172	0.077	0.004	8.678	-0.413

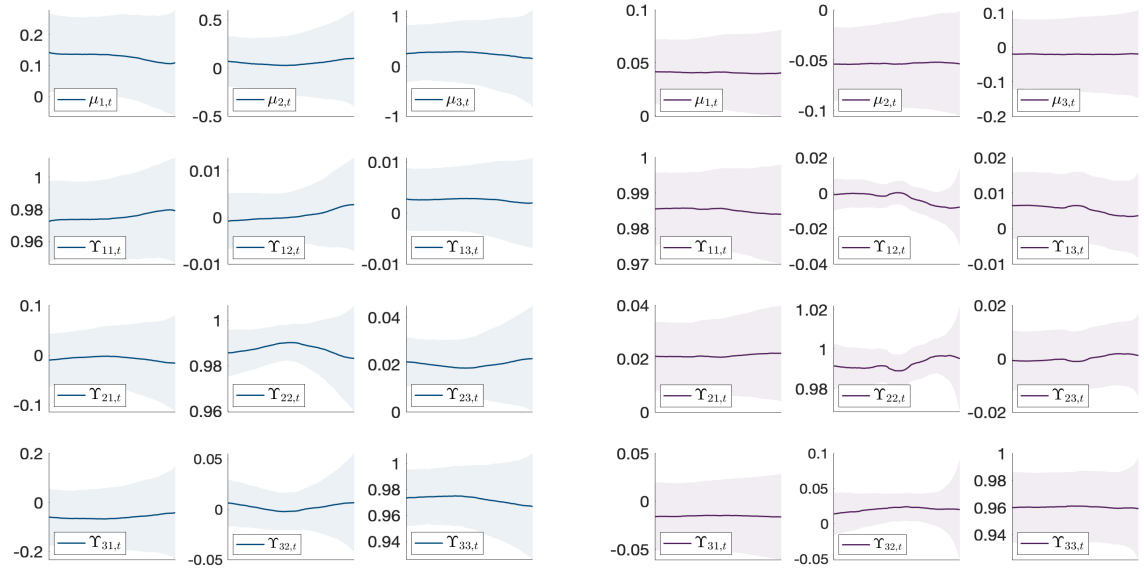
Notes: The table reports MCMC posterior results of WSV Wishart parameters for the final estimation windows – from Dec 31, 2007 to Dec 29, 2008 (Panel A) and from Jan 2, 2020 to Dec 30, 2020 (Panel B) – Mean, Standard Deviation (Std. Dev.), Numerical Standard Errors (NSE), Effective Sample Size (ESS) and CD convergence statistic with subsets divided into proportions of $0.2, 0.5$ and 0.3. The Newey-West estimator with 20 lag orders is applied to obtain a long-run variance estimate robust to serial correlation. Results are based on $20,000$ Gibbs iterations neglecting the first $15,000$ and saving every 10th draw after burn-in so that the posterior thinned sample is composed of 500 replications.

Figure 3.5: Posterior estimates of factor volatilities in DNS-WSV



Notes: The figure reports the posterior estimates of WSV factor variances and covariances Σ_t (mean and 95% HPDI) and yield curve Level (L_t), Slope (S_t) and Curvature (C_t) factors for the window periods from Dec 31, 2007 to Dec 29, 2008 (left panel) and from Jan 2, 2020 to Dec 30, 2020 (right panel).

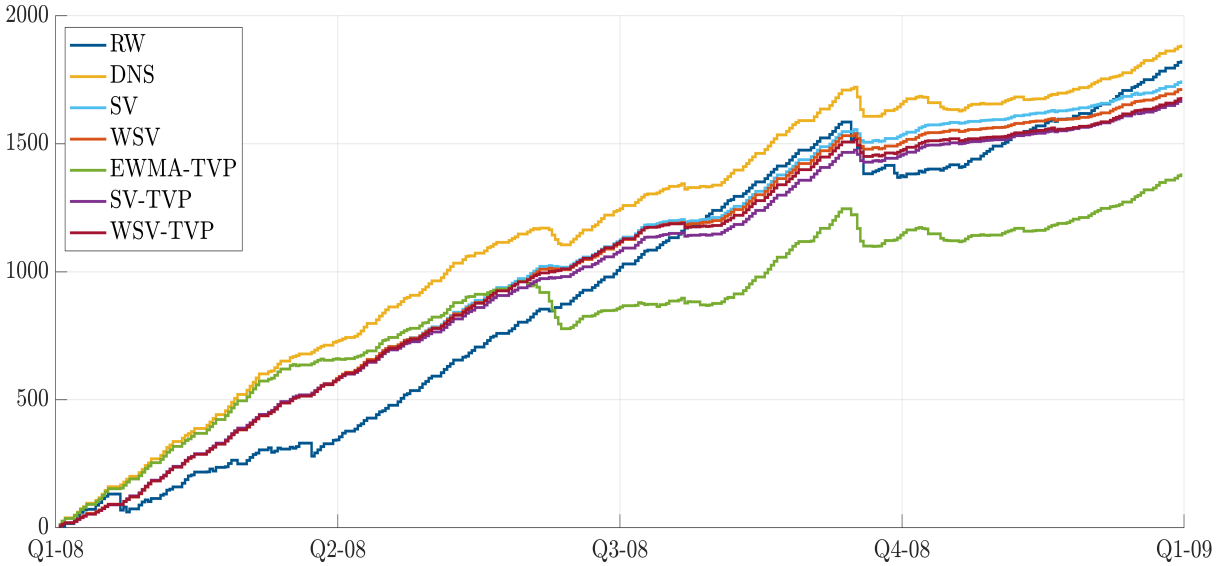
Figure 3.6: Posterior estimates of TVP parameters in DNS-WSV



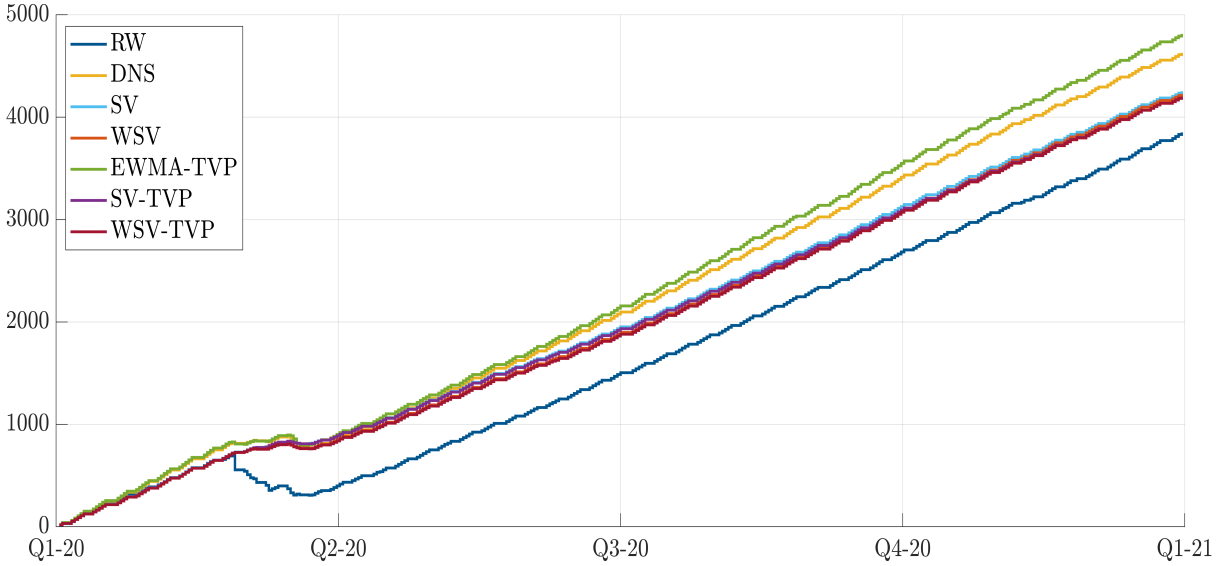
Notes: The figure reports the posterior estimates of WSV time-varying parameters μ_t and Υ_t (mean and 95% HPDI) for the window periods from Dec 31, 2007 to Dec 29, 2008 (left panel) and from Jan 2, 2020 to Dec 30, 2020 (right panel).

3.E.2 Additional figures on forecasting performance

Figure 3.7: Accumulated log-predictive density scores

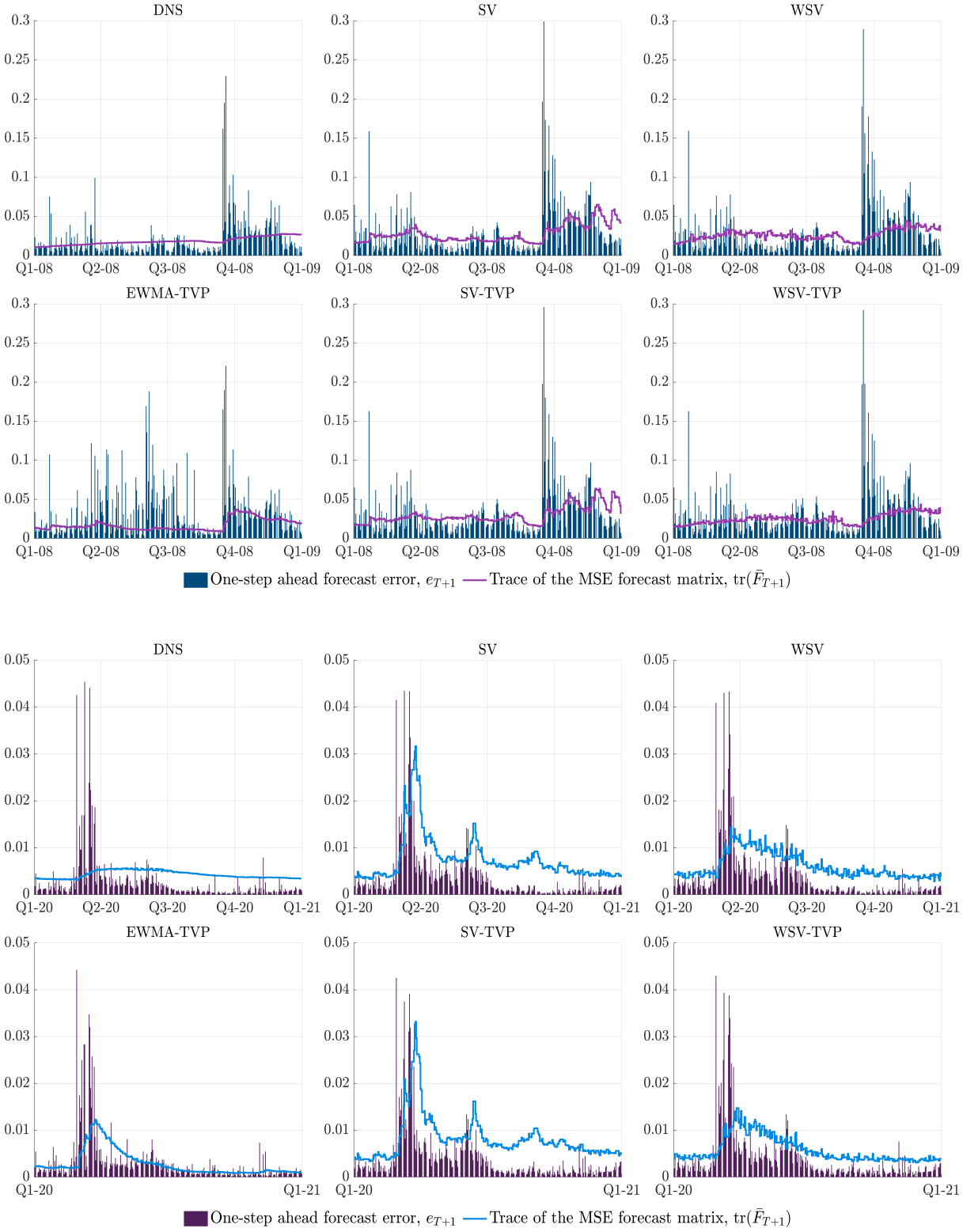


(a) Great Recession evaluation period



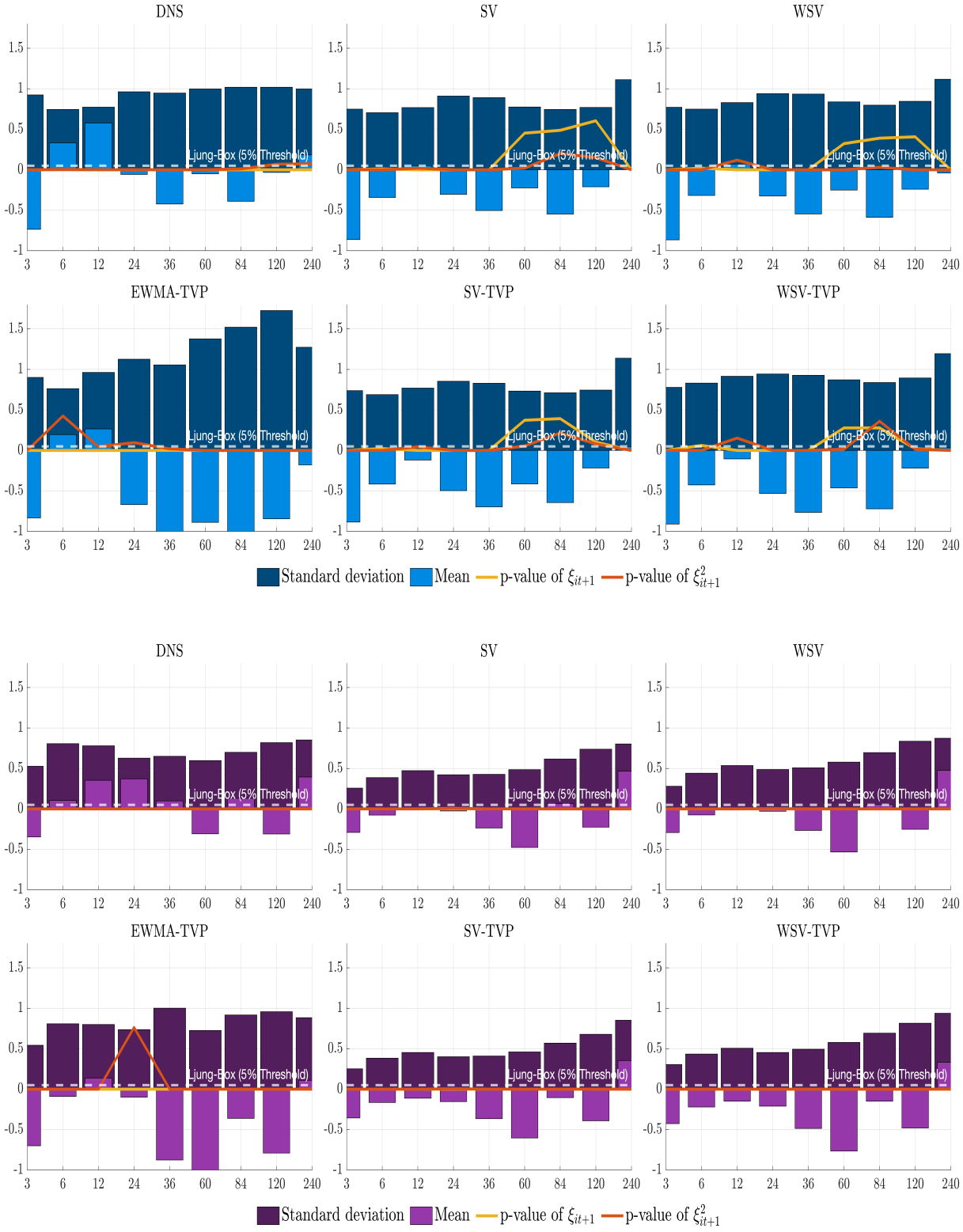
(b) COVID-19 evaluation period

Figure 3.8: Yields predictability in response to crisis shocks



Notes: The figure reports the one-step ahead forecast errors under MSE loss, e_{t+1}, and the corresponding trace of the MSE forecast matrix, tr($\bar{F}_{t+1} \times 10^{-1}$), for the Great Recession (upper panel) and COVID-19 (lower panel) predictive samples.

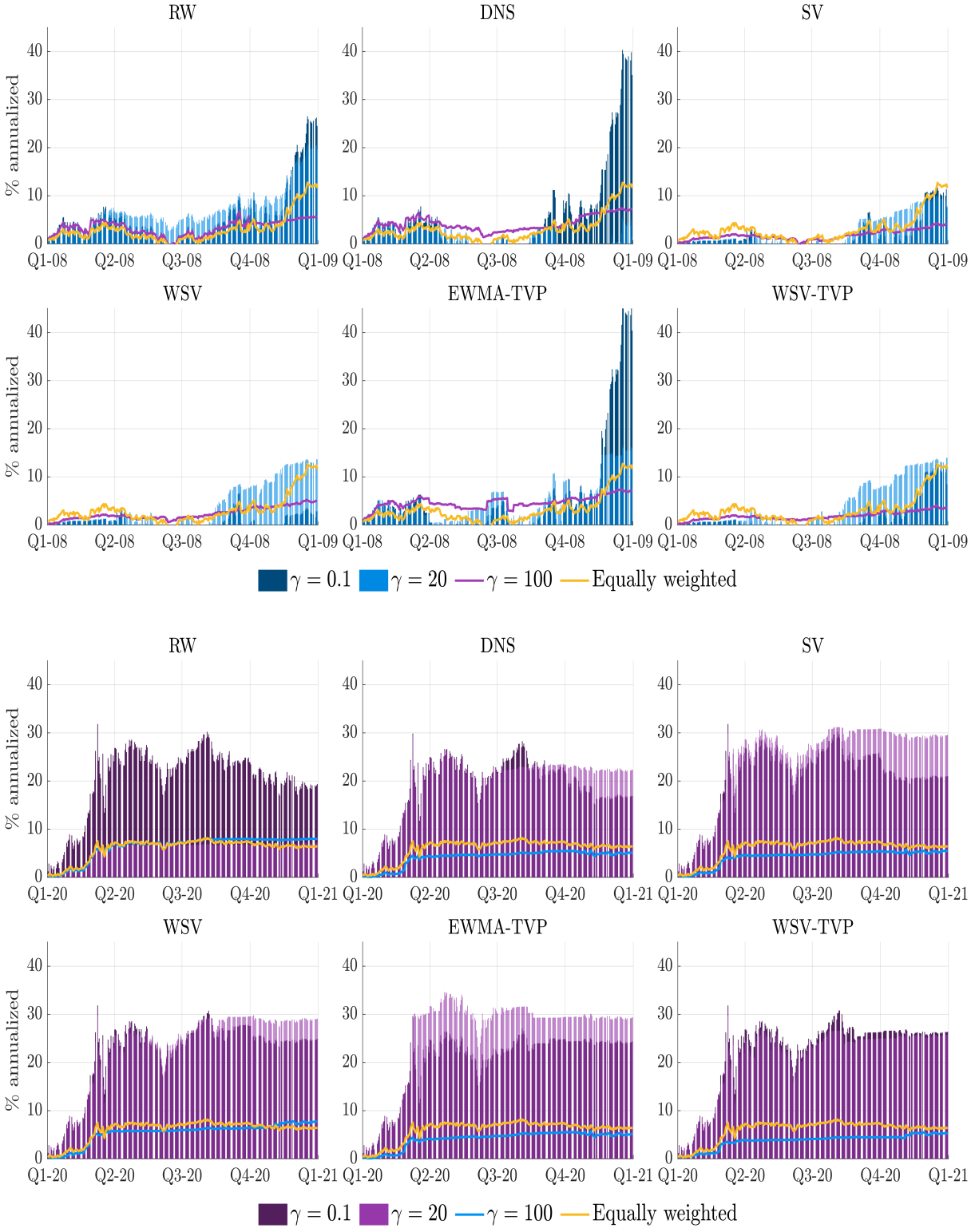
Figure 3.9: Sample statistics and autocorrelation test of predictive Pearson residuals



Notes: The figure reports the sample statistics of the predictive Pearson residuals ξ_{it+1} (bars) for each maturity τ_i ranging from 3 to 240 months, along with p-values of the Ljung-Box test statistic (lines) for the Great Recession (upper panel) and COVID-19 (lower panel) predictive samples. The Ljung-Box autocorrelation test for ξ_{it+1} and ξ^2_{it+1} is computed including 10 lags.

3.E.3 Additional figures on portfolio performance

Figure 3.10: Cumulative compounded excess returns under alternative investment scenarios



Notes: The figure reports the cumulative compounded excess returns (% annualized) for the Great Recession (upper panel) and COVID-19 (lower panel) predictive samples. The portfolio returns are provided for the aggressive ($\gamma = 0.1$), moderate ($\gamma = 20$), and conservative ($\gamma = 100$) investment scenarios, along with an equally weighted portfolio.

Bibliography

Aliaj, T., Ciganovic, M., and Tancioni, M. (2023). Nowcasting inflation with Lasso-regularized vector autoregressions and mixed frequency data. *Journal of Forecasting*, 42(3):464–480.

Alvarez, S. E. and Lein, S. M. (2020). Tracking Inflation on a Daily Basis. *Swiss Journal of Economics and Statistics*, 156(18).

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, 31(2):240–251.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342.

Anenberg, E. and Laufer, S. (2017). A More Timely House Price Index. *The Review of Economics and Statistics*, 99(4):722–734.

Aparicio, D. and Bertolotto, M. I. (2020). Forecasting Inflation with Online Prices. *International Journal of Forecasting*, 36(2):232–247.

Araujo, G. S. and Gaglianone, W. P. (2023). Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(2):100087.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4.

Asai, M. and McAleer, M. (2009). The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics*, 150(2):182–192.

Assenmacher, K., Glöckler, G., Holton, S., Trautmann, P., Ioannou, D., Mee, S., Bakk-Simon, K., Bergbauer, S., Catenaro, M., Charalampakis, E., et al. (2021). Clear, Consistent and Engaging: ECB Monetary Policy Communication in a Changing World. *ECB Occasional Paper*, 274:1–91.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25(1):2–11.

Babii, A., Ghysels, E., and Striaukas, J. (2021). Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106.

Banbura, M., Brenna, F., Paredes, J., and Ravazzolo, F. (2021). Combining Bayesian VARs with survey density forecasts: does it pay off? *ECB Working Paper*.

Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting*, volume 2, pages 195–237. Elsevier.

Bańbura, M., Leiva-Leon, D., and Menz, J.-O. (2023). Do Inflation Expectations Improve Model-Based Inflation Forecasts? *ECB Discussion Paper*, 2604:1–47.

Barbaglia, L., Frattarolo, L., Onorante, L., Pericoli, F. M., Ratto, M., and Pezzoli, L. T. (2023). Testing big data in a big crisis: Nowcasting under COVID-19. *International Journal of Forecasting*, 39(4):1548–1563.

Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., and Koenigstein, N. (2022). Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting*, 39:1145–1162.

Beck, G., Carstensen, K., Menz, J.-O., Schnorrenberger, R., and Wieland, E. (2022). Real-Time Food Price Inflation in Germany in Light of the Russian Invasion of Ukraine. *VOXEU Column, link*.

Beck, G., Carstensen, K., Menz, J.-O., Schnorrenberger, R., and Wieland, E. (2023). Nowcasting Consumer Price Inflation Using High-Frequency Scanner Data: Evidence from Germany. Deutsche Bundesbank Discussion Paper No. 34/2023.

Beck, G. and Jaravel, X. (2021). Prices and Global Inequality: New Evidence from Worldwide Scanner Data. *Discussion Paper*, pages 1–50.

Beck, G. W. and Lein, S. M. (2020). Price Elasticities and Demand-Side Real Rigidities in Micro Data and in Macro Models. *Journal of Monetary Economics*, 115:200–212.

Benalal, N., del Hoyo, J. L. D., Landau, B., Roma, M., and Skudelny, F. (2004). To Aggregate or not to Aggregate? Euro Area Inflation Forecasting. *ECB Discussion Paper*, 374:1–67.

Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.

Bermingham, C. and D'Agostino, A. (2013). Understanding and Forecasting Aggregate and Disaggregate Price Dynamics. *Empirical Economics*, 46:765–799.

Bianchi, F., Mumtaz, H., and Surico, P. (2009). The great moderation of the term structure of uk interest rates. *Journal of Monetary Economics*, 56(6):856–871.

Bobeica, E. and Hartwig, B. (2023). The COVID-19 shock and challenges for inflation modelling. *International journal of forecasting*, 39(1):519–539.

Borup, D., Rapach, D. E., and Schütte, E. C. M. (2023). Mixed-frequency machine learning: Nowcasting and backcasting weekly initial claims with daily internet search volume data. *International Journal of Forecasting*, 39(3):1122–1144.

Botha, B., Burger, R., Kotzé, K., Rankin, N., and Steenkamp, D. (2022). Big Data Forecasting of South African Inflation. *Empirical Economics*.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breitung, J. and Roling, C. (2015). Forecasting Inflation Rates Using Daily Data: A Nonparametric MIDAS Approach. *Journal of Forecasting*, 34(7):588–603.

Buda, G., Carvalho, V., Hansen, S., Ortiz, A., Rodrigo, T., and Rodríguez Mora, J. (2022). National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption. *Cambridge Working Paper in Economics*, 2244:1–72.

Buda, G., Carvalho, V. M., Corsetti, G., Duarte, J. B., Hansen, S., Ortiz, Á., Rodrigo, T., and Rodríguez Mora, J. V. (2023). Short and Variable Lags. *Cambridge Working Paper in Economics*, 2321:1–61.

Butters, A., Sacks, D., and Seo, B. (2022). How Do National Firms Respond to Local Cost Shocks? *American Economic Review*, 112(5):1737–1772.

Byrne, J. P., Cao, S., and Korobilis, D. (2017). Forecasting the term structure of government bond yields in unstable environments. *Journal of Empirical Finance*, 44:209–225.

Caldeira, J. F., Moura, G. V., and Santos, A. A. P. (2016). Bond portfolio optimization using dynamic factor models. *Journal of Empirical Finance*, 37:128–158.

Capistrán, C., Constandse, C., and Ramos-Francia, M. (2010). Multi-horizon inflation forecasts using disaggregated data. *Economic Modelling*, 27(3):666–677.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87(418):493–500.

Carriero, A., Clark, T. E., and Marcellino, M. (2015). Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 178(4):837–862.

Carriero, A., Clark, T. E., and Marcellino, M. (2019a). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212(1):137–154.

Carriero, A., Clark, T. E., and Marcellino, M. (2020). Nowcasting tail risks to economic activity with many indicators. *Federal Reserve Bank of Cleveland Working Paper*, (No.20-13).

Carriero, A., Clark, T. E., and Marcellino, M. (2021). No-arbitrage priors, drifting volatilities, and the term structure of interest rates. *Journal of Applied Econometrics*, 36(5):495–516.

Carriero, A., Galvao, A. B., and Kapetanios, G. (2019b). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, 35(4):1226–1239.

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Cascaldi-Garcia, D., Ferreira, T. R., Giannone, D., and Modugno, M. (2023). Back to the present: Learning about the euro area through a now-casting model. *International Journal of Forecasting*.

Cavallo, A., Cavallo, E., and Rigobon, R. (2014). Prices and Supply Disruptions during Natural Disasters. *Review of Income and Wealth*, 60:449–471.

Cavallo, A. and Kryvtsov, O. (2023). What can Stockouts Tell us About Inflation? Evidence from Online Micro Data. *Journal of International Economics*, page 103769.

Cavallo, A. and Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, 30(2):151–178.

Chan, J. C. and Eisenstat, E. (2018). Bayesian model comparison for time-varying parameter vars with stochastic volatility. *Journal of Applied Econometrics*, 33(4):509–532.

Chinn, M. D., Meunier, B., and Stumpner, S. (2023). Nowcasting world trade with machine learning: a three-step approach. Technical report, National Bureau of Economic Research.

Chipman, H. A., George, E. I., and McCulloch, R. E. (2012). BART: Bayesian additive regression trees. *Annals of Applied Statistics*, 6(1):266–298.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

Cimadomo, J., Giannone, D., Lenza, M., Monti, F., and Sokol, A. (2022). Nowcasting with Large Bayesian Vector Autoregressions. *Journal of Econometrics*, 231:500–519.

Clark, T. E., Leonard, S., Marcellino, M., and Wegmüller, P. (2022). Weekly Nowcasting US Inflation with Enhanced Random Forests. *Mimeo*.

Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575.

D'Acunto, F., Malmendier, U., Ospina, J., and Weber, M. (2021). Exposure to Grocery Prices and Inflation Expectations. *Journal of Political Economy*, 129(5):1615–1639.

Dahlhaus, T., Guénette, J.-D., and Vasishtha, G. (2017). Nowcasting BRIC+M in real time. *International Journal of Forecasting*, 33(4):915–935.

de Haan, J., Hendriks, R., and Scholz, M. (2021). Price Measurement Using Scanner Data: Time-Product Dummy Versus Time Dummy Hedonic Indexes. *Review of Income and Wealth*, 67(2):394–417.

Deutsche Bundesbank (2020). The Fiscal Stimulus Package Announced by the Coalition Parties. *Bundesbank Monthly Report*, Juni:28–29.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Diebold, F. X. and Rudebusch, G. D. (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*. Princeton University Press.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131(1):309–338.

Dietrich, A., Eiglsperger, M., Mehrhoff, J., and Wieland, E. (2021). Chain linking over December and methodological changes in the HICP: view from a central bank perspective. *ECB Statistics Paper Series*, 40.

Diewert, E. (2005). Weighted Country Product Dummy Variable Regressions and Index Number Formulae. *Review of Income and Wealth*, 51(4):561–570.

Doerr, S., Gambacorta, L., and Maria Serena, J. (2021). Big Data and Machine Learning in Central Banking. *BSI Working Paper*, 930:1–26.

Dubois, P., Griffith, R., and O'Connell, M. (2022). The Use of Scanner Data for Economics Research. *Annual Review of Economics*, 14:723–745.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*. Number 38. Oxford University Press.

Eraslan, S. and Götz, T. (2021). An unconventional weekly economic activity index for germany. *Economics Letters*, 204:109881.

Espasa, A. and Mayo-Burgos, I. (2013). Forecasting Aggregates and Disaggregates with Common Features. *International Journal of Forecasting*, 29:718–732.

Eurostat (2018). Harmonised index of Consumer Prices (HICP). Methodological Manual. *Publication Office of the European Union, Luxembourg*.

Eurostat (2022). Guide on Multilateral Methods in the Harmonised Index on Consumer Prices (HICP) - 2022 edition.

Faust, J. and Wright, J. H. (2013). Forecasting inflation. In *Handbook of economic forecasting*, volume 2, pages 2–56. Elsevier.

Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A*, 178(1):57–82.

Friedberg, R., Tibshirani, J., Athey, S., and Wager, S. (2020). Local linear forests. *Journal of Computational and Graphical Statistics*, 30(2):503–517.

Gagnon, E. and López-Salido, D. (2019). Small Price Responses to Large Demand Shocks. *Journal of the European Economic Association*, 18(2):792–828.

Garcia, M. G., Medeiros, M. C., and Vasconcelos, G. F. (2017). Real-Time Inflation Forecasting with High-Dimensional Models: The Case of Brazil. *International Journal of Forecasting*, 33(3):679–693.

Gautier, E., Conflitti, C., Faber, R. P., Fabo, B., Fadejeva, L., Jouvanceau, V., Menz, J.-O., Messner, T., Petroulas, P., Roldan-Blanco, P., Rumler, F., Santoro, S., Wieland, E., and Zimmer, H. (2023). New Facts on Consumer Price Rigidity in the Euro Area. *American Economic Journal Macroeconomics*.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics*, 4:641–649.

Geweke, J. and Amisano, G. (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2):216–230.

Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical science*, pages 473–483.

Ghysels, E. and Marcellino, M. (2018). *Applied Economic Forecasting Using Time Series Methods*. Oxford University Press.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. *Discussion paper UNC and UCLA*.

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Giannone, D., Lenza, M., Momferatou, D., and Onorante, L. (2014). Short-Term Inflation Projections: A Bayesian Vector Autoregressive Approach. *International Journal of Forecasting*, 30:635–644.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of monetary economics*, 55(4):665–676.

Gorgi, P., Hansen, P., Janus, P., and Koopman, S. (2018). Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model. *Journal of Financial Econometrics*.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5):920–964.

Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.

Gribisch, B. (2016). Multivariate Wishart stochastic volatility and changes in regime. *AStA Advances in Statistical Analysis*, 100(4):443–473.

Harchaoui, T. and Janssen, R. (2018). How Can Big Data Enhance the Timeliness of Official Statistics?: The Case of the U.S. Consumer Price Index. *International Journal of Forecasting*, 34(2):225–234.

Hautsch, N. and Yang, F. (2012). Bayesian inference in a stochastic volatility Nelson-Siegel model. *Computational Statistics & Data Analysis*, 56(11):3774–3792.

Hauzenberger, N., Huber, F., and Klieber, K. (2023). Real-Time Inflation Forecasting Using Non-Linear Dimension Reduction Techniques. *International Journal of Forecasting*, 39(2):901–921.

Hendry, D. and Hubrich, K. (2011). Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate. *Journal of Business & Economic Statistics*, 29(2):216–227.

Henn, K., Islam, C.-G., Schwind, P., and Wieland, E. (2019). Measuring Price Dynamics of Package Holidays with Transaction Data. *EURONA*, 2/2019:95–132.

Hindrayanto, I., Koopman, S. J., and de Winter, J. (2016). Forecasting and nowcasting economic growth in the euro area using factor models. *International Journal of Forecasting*, 32(4):1284–1305.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1):69–82.

Huber, F., Koop, G., Onorante, L., Pfarrhofer, M., and Schreiner, J. (2023). Nowcasting in a pandemic using non-parametric mixed frequency vars. *Journal of Econometrics*, 232(1):52–69.

Huwiler, M. and Kaufmann, D. (2013). Combining Disaggregate Forecasts for Inflation: The SNB's ARIMA Model. *Swiss National Bank Economic Studies*, 7:1–32.

Ibarra, R. (2012). Do Disaggregated CPI Data Improve the Accuracy of Inflation Forecasts? *Economic Modelling*, 29:1305–1313.

IMF, ILO, OECD, Eurostat, UNECE, and The World Bank (2020). Consumer Price Index Manual: Concepts and Methods. *IMF*.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

Jaravel, X. (2019). The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector. *Quarterly Journal of Economics*, 134(2):715–783.

Jaravel, X. and O'Connell, M. (2020). Real-Time Price Indices: Inflation Spike and Falling Product Variety During the Great Lockdown. *Journal of Public Economics*, 191:104–270.

Joseph, A., Kalamara, E., Kapetanios, G., Potjagailo, G., and Chakraborty, C. (2021). Forecasting UK inflation bottom up. *Bank of England staff working papers*, (915).

Joseph, A., Kalamara, E., Kapetanios, G., Potjagailo, G., and Chakraborty, C. (2022). Forecasting UK Inflation Bottom Up. *Bank of England Staff Working Paper*, 915:1–38.

Kaplan, G. and Schulhofer-Wohl, S. (2017). Inflation at the Household Level. *Journal of Monetary Economics*, 91:19–38.

Karadi, P., Amann, J., Bachiller, J. S., Seiler, P., and Wursten, J. (2023). Price setting on the two sides of the Atlantic-Evidence from supermarket scanner data. *Journal of Monetary Economics*.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65(3):361–393.

Kleppe, T. S., Liesenfeld, R., Moura, G. V., and Oglend, A. (2021). Analyzing Commodity Futures Using Factor State-Space Models with Wishart Stochastic Volatility. *Econometrics and Statistics*.

Knotek, E. S. and Zaman, S. (2017). Nowcasting US headline and core inflation. *Journal of Money, Credit and Banking*, 49(5):931–968.

Knotek II, E. S. and Zaman, S. (2023). Real-time density nowcasts of US inflation: A model combination approach. *International Journal of Forecasting*, 39(4):1736–1760.

Kohns, D. and Potjagailo, G. (2023). Flexible Bayesian MIDAS: time-variation, group-shrinkage and sparsity. *Bank of England staff working papers*, 1025.

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198.

Krüger, F., Clark, T. E., and Ravazzolo, F. (2017). Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business & Economic Statistics*, 35(3):470–485.

Ku, Y.-C., Bloomfield, P., and Ghosh, S. K. (2014). A flexible observed factor model with separate dynamics for the factor volatilities and their correlation matrix. *Statistical Modelling*, 14(1):1–20.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*, 3(2).

Laurini, M. P. and Caldeira, J. F. (2016). A macro-finance term structure model with multivariate stochastic volatility. *International Review of Economics & Finance*, 44:68–90.

Li, J., Liao, Z., and Quaedvlieg, R. (2022). Conditional Superior Predictive Ability. *The Review of Economic Studies*, 89(2):843–875.

Liesenfeld, R. and Richard, J.-F. (2008). Improving MCMC, using efficient importance sampling. *Computational Statistics & Data Analysis*, 53(2):272–288.

Macias, P., Stelmasiak, D., and Szafranek, K. (2023). Nowcasting Food Inflation with a Massive Amount of Online Prices. *International Journal of Forecasting*, 39(2):809–826.

Marcellino, M. and Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4):518–550.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Marques, A. B. C. (2012). Central Bank of Brazil's market expectations system: a tool for monetary policy. *IFC Bulletin*, 36:304–324.

Marsilli, C. (2014). Variable Selection in Predictive MIDAS Models. *Banque de France working paper*, 520.

McCracken, M. W., Owyang, M., and Sekhposyan, T. (2015). Real-time forecasting and scenario analysis using a large mixed-frequency Bayesian VAR. *FRB St. Louis Working Paper*, (2015-30).

Medeiros, M. C., Vasconcelos, G., and Freitas, E. (2016). Forecasting brazilian inflation with high-dimensional models. *Brazilian Review of Econometrics*, 36(2):223–254.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., and Zilberman, E. (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics*, 39(1):98–119.

Messner, T., Rumler, F., and Strasser, G. (2023). Cross-Country Price and Inflation Dispersion: Retail Network or National Border? *ECB Discussion Paper*, 2276:1–56.

Modugno, M. (2013). Now-casting inflation using high frequency data. *International Journal of Forecasting*, 29(4):664–675.

Mogliani, M. and Simoni, A. (2021). Bayesian MIDAS penalized regressions: estimation, selection, and prediction. *Journal of Econometrics*, 222(1):833–860.

Monteforte, L. and Moretti, G. (2013). Real-Time Forecasts of Inflation: The Role of Financial Variables. *Journal of Forecasting*, 32(1):51–61.

Nagengast, A., Bursian, D., and Menz, J.-O. (2021). Dynamic Pricing and Exchange Rate Pass-Through: Evidence from Transaction-Level Data. *European Economic Review*, 133:1–29.

Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787.

Nieto, M. R. and Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, 32(2):475–501.

Paranhos, L. (2021). Predicting Inflation with Neural Networks. *Warwick Economics Research Paper*, 1344.

Philipov, A. and Glickman, M. E. (2006a). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews*, 25(2-3):311–334.

Philipov, A. and Glickman, M. E. (2006b). Multivariate stochastic volatility via Wishart processes. *Journal of Business & Economic Statistics*, 24(3):313–328.

Powell, B., Nason, G., Elliott, D., Mayhew, M., Davies, J., and Winton, J. (2018). Tracking and Modelling Prices Using Web-Scraped Price Microdata: Towards Automated Daily Consumer Price Index Forecasting. *Journal of the Royal Statistical Society*, 181(3):737–756.

Richard, J.-F. and Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141(2):1385–1411.

Richardson, A., van Florenstein Mulder, T., and Vehbi, T. (2021). Nowcasting gdp using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, 37(2):941–948.

Riskmetrics, T. (1996). JP Morgan Technical Document.

Schnorrenberger, R., Schmidt, A., and Moura, G. V. (2024). Harnessing Machine Learning for Real-Time Inflation Nowcasting. DNB Working Paper No. 806.

Schorfheide, F. and Song, D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, 33(3):366–380.

Shin, M. and Zhong, M. (2017). Does realized volatility help bond yield density prediction? *International Journal of Forecasting*, 33(2):373–389.

Siliverstovs, B. (2017). Short-term forecasting with mixed-frequency data: a MIDASSO approach. *Applied Economics*, 49(13):1326–1343.

Stelmasiak, D., Szafranek, K., Macias, P., and Błażejowska, A. (2023). Online food prices and shocks to product availability since Covid-19. *Applied Economics Letters*, pages 1–7.

Stock, J. H. and Watson, M. W. (1999). Forecasting Inflation. *Journal of Monetary Economics*, 44(2):293–335.

Stock, J. H. and Watson, M. W. (2007). Why has us inflation become harder to forecast? *Journal of Money, Credit and banking*, 39:3–33.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tissot, B. and de Beer, B. (2020). Implications of COVID-19 for Official Statistics: A Central Banking Perspective. *BIS IFC Working Paper*, 20:1–23.

Uematsu, Y. and Tanaka, S. (2019). High-dimensional macroeconomic forecasting and variable selection via penalized regression. *The Econometrics Journal*, 22(1):34–56.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica*, pages 59–73.

Ulgazi, Y. and Vertier, P. (2022). Forecasting Inflation in France: An Update of MAPI. *Banque de France Discussion Paper*, 869:1–33.

Watanabe, T. (2020). The Responses of Consumption and Prices in Japan to the COVID-19 Crisis and the Tohoku Earthquake. *CJEB Working Papers - Columbia Business School*, 373:1–16.

Weber, M., Gorodnichenko, Y., and Coibion, O. (2022). The Expected, Perceived, and Realized Inflation of U.S. Households Before and During the COVID-19 Pandemic. *NBER Working Paper*, 29640.

Windle, J., Carvalho, C. M., et al. (2014). A tractable state-space model for symmetric positive-definite matrices. *Bayesian Analysis*, 9(4):759–792.

Wu, S.-J., Ghosh, S. K., Ku, Y.-C., and Bloomfield, P. (2018). Dynamic correlation multivariate stochastic volatility with latent factors. *Statistica Neerlandica*, 72(1):48–69.

Zhao, P. and Yu, B. (2006). On model selection consistency of Lasso. *The Journal of Machine Learning Research*, 7:2541–2563.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: series B (statistical methodology)*, 67(2):301–320.