Harnessing machine learning for real-time inflation nowcasting*

By Richard Schnorrenberger (Kiel University), Aishameriane Schmidt
(Erasmus Universiteit, Rotterdam, Tinbergen Institute, and De Nederlandsche Bank),
and Guilherme Valle Moura (Federal University of Santa Catarina)

In a recent paper (Schnorrenberger et al., 2024, available here), we study the effectiveness of machine learning in generating weekly inflation nowcasts for the Brazilian economy, using high-frequency macrofinancial data and a daily survey from professional forecasts. We show that a well-designed mixedfrequency ML approach delivers major nowcasting gains during the onset of the COVID-19 crisis, where professional forecasters underestimated the rapidly evolving inflationary environment. In general, we find that crafting an effective ML model for inflation nowcasting depends on three key ingredients: (i) accurate timely signals from price indicators, (ii) informed judgment entailed in SPF expectations, and (iii) variable selection performed via shrinkage, especially the LASSO. These insights provide valuable implications for monetary policy and economic forecasting, particularly in volatile economic environments.

Inflation nowcasting and the potential of machine learning

The demand for accurate and timely inflation nowcasts has become increasingly urgent in the light of recent economic disruptions, including the COVID-19 pandemic. In rapidly changing inflationary environments, tracking inflation dynamics in real-time falls short given the publication lag of official price statistics. In the meantime, high-frequency (e.g., weekly or daily) and quickly released data have become particularly useful for forecasting the current state ("nowcasting") of the inflation process. Timelier signals extracted from these data allow us to better anticipate swift inflationary shocks, as well as inflationary trends that may escalate or dwindle.

Machine learning (ML) methods have already been proven as valuable tools for inflation forecasting in a data-rich environment (see, among others, Medeiros et al., 2016; Hauzenberger et al., 2023), exhibiting substantial improvements over traditional econometric models (e.g., Stock and Watson, 2007). In our study, we assess the potential of ML methods for inflation nowcasting (see, for example, Modugno, 2013; Monteforte and Moretti, 2013; Breitung and Roling, 2015; Knotek and Zaman, 2017; Knotek II and Zaman, 2023; Beck et al., 2023). Specifically, we provide guidance on key modeling choices when using ML to construct inflation nowcasts in a real-time setup. By leveraging on high-frequency macrofinancial data and a daily survey of professional forecasters (SPF), we show that ML helps us refine inflation nowcasts in normal times and significantly improve the model's responsiveness and accuracy in changing economic conditions, such as those experienced in the COVID-19 crisis. This capability is vital for informed monetary policy decisions and economic planning, proving particularly advantageous in turbulent times.

The unique features of the Brazilian inflation data

Brazil's unique economic environment, marked by frequent inflationary episodes and significant economic fluctuations, presents a valuable case study for inflation nowcasting. The availability of alternative high-frequency and timely price indicators and daily surveys of professional forecasters by the Brazilian Central Bank (BCB) provides a robust dataset for real-time analysis. This section provides context to this data and motivates its use for inflation nowcast.

We have constructed a novel real-time database of macro-financial series specifically tailored for the purpose of inflation nowcasting within the Brazilian economy. Our dataset predominantly comprises timely price indicators, financial variables, and expert forecasts that are indicative of the current month's inflation rate. In addition to the target Consumer Price Index (CPI) variable, we have systematically organized publicly available data on price indicators issued by both public and private institutions, financial indicators, and daily surveys of professional forecasters with aggregate predictions for the target variable. Our comprehensive real-time dataset spans from June 2004 to December 2022, encompassing 222 monthly observations, with information on release dates 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. Figure 1a shows the IPCA evolution since mid-2001, shortly after the BCB adopted the inflation targeting regime. For most periods, IPCA has remained between 5% and 10%, with spikes around economic and political crises. The plot also displays other price indexes, that are available either on a monthly or weekly frequency and display a strong correlation with the IPCA. An example of the release dates is in Figure 1b. While the IPCA is available around the 10th day of the subsequent month, several price indexes are announced during the current month. In total, we utilize five monthly price indices and six weekly indicators of consumer and energy prices.

In addition to the price indicators, we also utilize as predictors for the IPCA the Survey of Professional Forecasters from the Brazilian Central Bank, namely the FOCUS survey data. The FOCUS survey was initiated by the Brazilian Central Bank (BCB) in the late 90s alongside the implementation of the inflation-targeting regime. Primarily involving pre-screened financial and economic entities, the BCB gathers daily expert forecasts on key macroeconomic indicators like GDP, inflation, and exchange rates. The BCB releases aggregate daily statistics from the survey with a one-day delay and also issues a ranking of the top five forecasting institutions. Historically, the survey has over 100 active participants, and their median forecasts, especially for the IPCA, are critical both as a predictor in our models and as a benchmark. These forecasts are closely monitored through weekly BCB reports and are compared against those from the top five performers, which are determined based on their previous forecasting accuracy.

We also include in our dataset a group of predictors that contains daily information from financial markets, including movements in the yield curve or interest rate spreads, commodity and stock price indices, and exchange rates.

## Setting up a mixed-frequency framework where machine learning can thrive

In Schnorrenberger et al. (2024), we describe an unrestricted mixed-frequency ML framework that 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.

Figure 1: Brazilian price indicators

(a) Time series of the official Brazilian CPI (IPCA, in black) and alternative price indicators (colored lines)
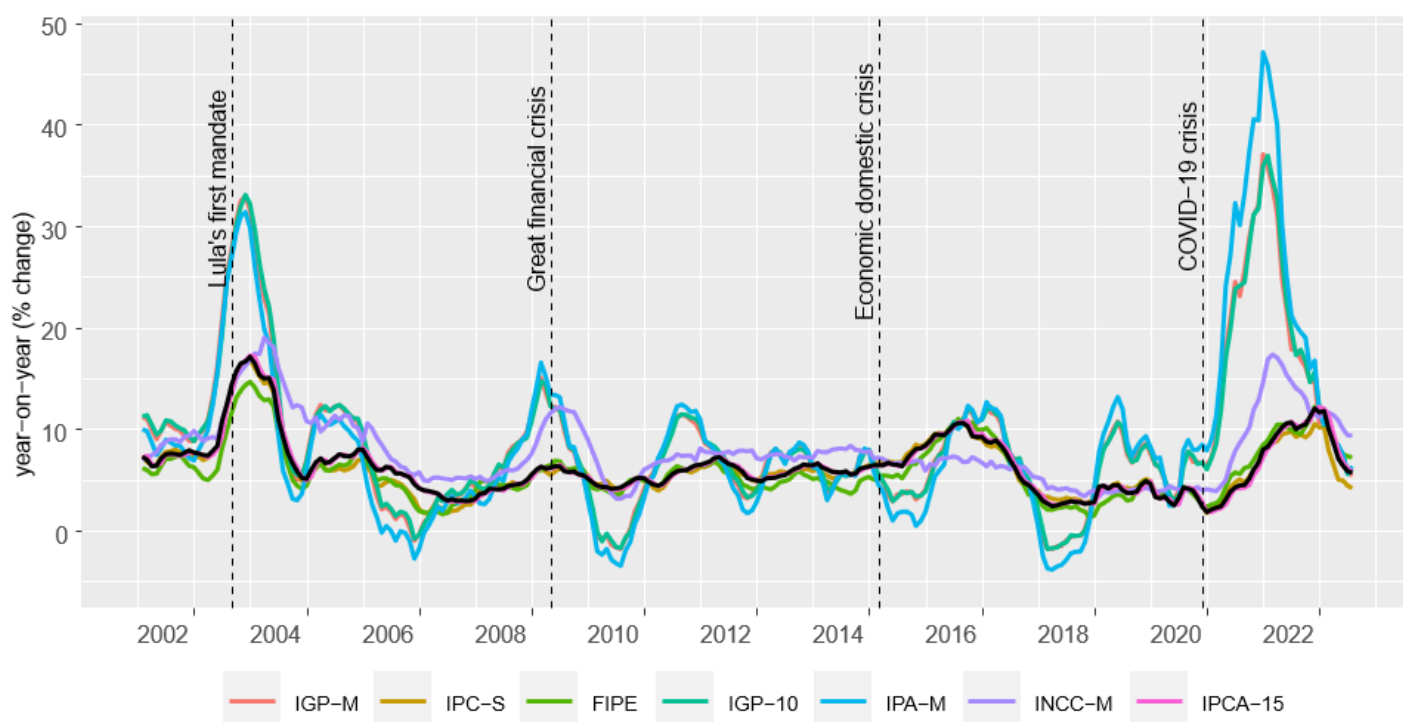
(b) Release calendar of Brazilian price indicators in December 2022
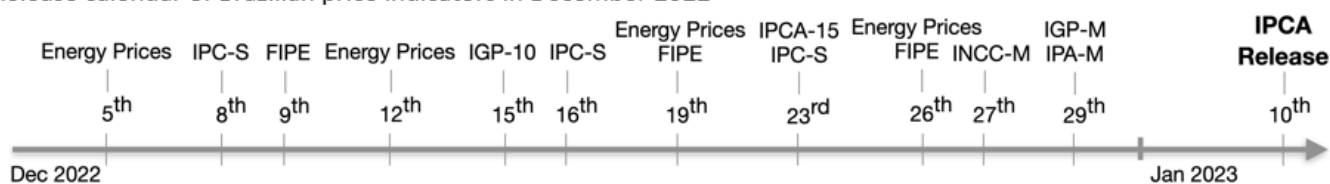
We update our nowcasts at four different points within the reporting month: days 8, 15, 22, and end-of-month. Given the mixed-frequency environment, we take a stance on how to incorporate high-frequency information on these four nowcast days. We assume that the daily information from financial predictors and SPF can be weekly summarized by the latest available value on the nowcast day. For the weekly variables, they only enter in a given nowcast if their contemporaneous value has been released at the nowcast day. Therefore, in the first week we include only high-frequency variables, while at the end of the month, our nowcast matrix of predictors will be much larger, comprising all the variables in the database. Hence, we constantly assess the real-time data availability of monthly predictors by the time of the nowcast.

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. 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. In our comparison, we produced nowcasts using 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, which naturally embeds the serial dependence across different high-frequency lags.

On the other group of models, we have tree-based methods, which are nonparametric methods that recursively divide the predictor space according to a pre-determined splitting rule. They can uncover potentially underlying non-linear structures of the data, potentially enhancing predictions. 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).

Does machine learning help us improve inflation nowcasts?

Our empirical exercise underscores the effectiveness of shrinkage models over tree-based methods. This indicates that a targeted selection of strong determinants is beneficial in a high-dimensional setting for inflation nowcasting. Notably, variable selection done via the LASSO leads to exceptional prediction accuracy at longer nowcast horizons. For example, when nowcasting on day 8 or 15 of the reporting month, LASSO yields significant predictive gains of 17% and 13.5% in terms of RMSE, respectively, compared to median SPF expectations. The difference between model predictions narrows when nowcasts are computed at the end-of-month, where the competitive advantage of LASSO compared to the median SPF is 8.5%. This result aligns with the timing of price indicator releases, which predominantly occur towards the end of the month, prompting professional forecasters to adjust their expectations more frequently as the information set expands within the reporting month.
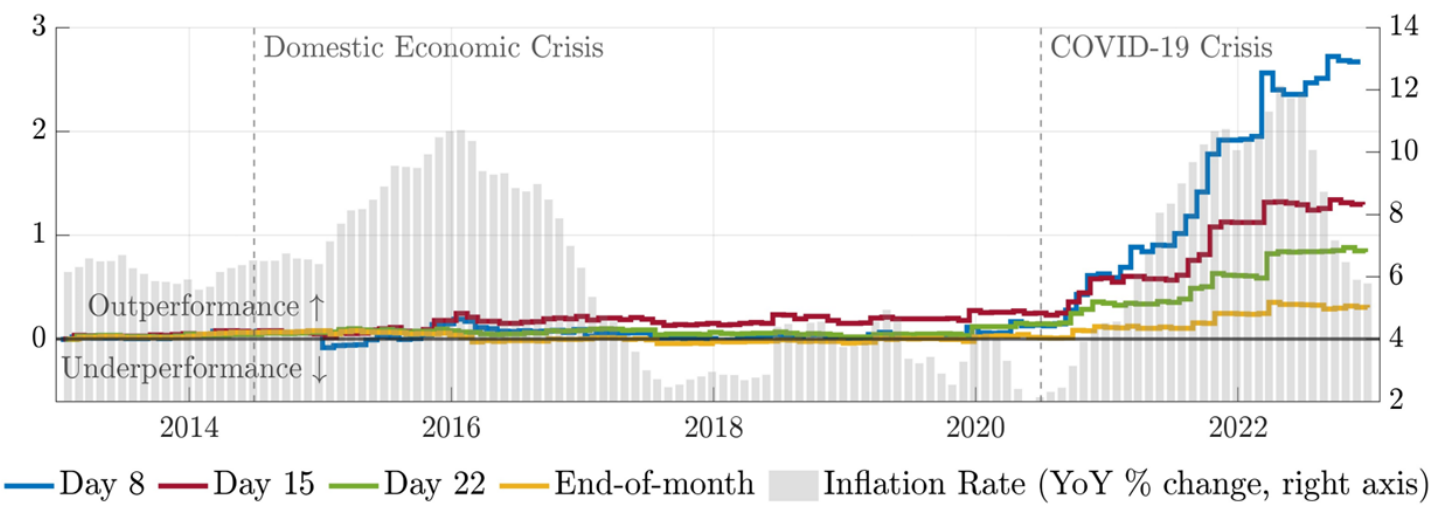
The evolution of loss differentials illustrated in Figure 2 reveals that LASSO predictive gains mostly build up during the COVID-19 inflation surge, when professional forecasters underestimated the swift change in economic conditions. Moreover, the importance of timely updates across different information sets is accentuated during the COVID-19 crisis. This can be seen by the superiority of LASSO at longer horizons (days 8, 15, and 22) compared to the moderate gains at shorter horizons (end-of-month). As for nowcasts purely based on tree-based models, we observe underperformance in relation to the tough SPF benchmark across most horizons, being highly detrimental during calm times.

What drives the poor performance of tree-based methods? Although the flexibility of these methods is capable of uncovering complex nonlinear patterns within inflation data, they require extensive amounts of data to perform well. Tree-based methods might be ill-equipped to handle the limited samples of inflation series. We show that a LASSO pre-selection step helps mitigate the curse of dimensionality and significantly improves their predictive

performance. Additionally, given that nowcasts inherently cover a short horizon during which inflationary dynamics may exhibit minor nonlinearities, linear methods can achieve better approximations and benefit from greater parsimony.

Finally, variable selection performed via the LASSO facilitates model interpretation, as illustrated in Figure 3. We find that relatively sparse model structures prevail at longer nowcast horizons, whereas SPF expectations and weekly price indicators exert the most substantial impact on shaping our model-based nowcasts. Therefore, hard data information and informed judgment entailed in SPF expectations provide a good anchor for ML models, particularly when navigating the challenges posed by the pandemic period. At shorter horizons, a more dense structure emerges, driven by valuable data releases of monthly price indicators that contain accurate contemporaneous signals. On the other hand, financial variables play a minor role due to their limited informativeness about current inflation dynamics.

Figure 2: CUMSFE: LASSO versus the SPF benchmark

Notes: This Figure reports the evolution of loss differentials (cumulative sum of squared forecast error, CUMSFE) between LASSO and median SPF expectations when nowcasting occurs on days 8, 15, 22 and end-of-month.

Figure 3: Variable relevance via coefficient estimates using LASSO
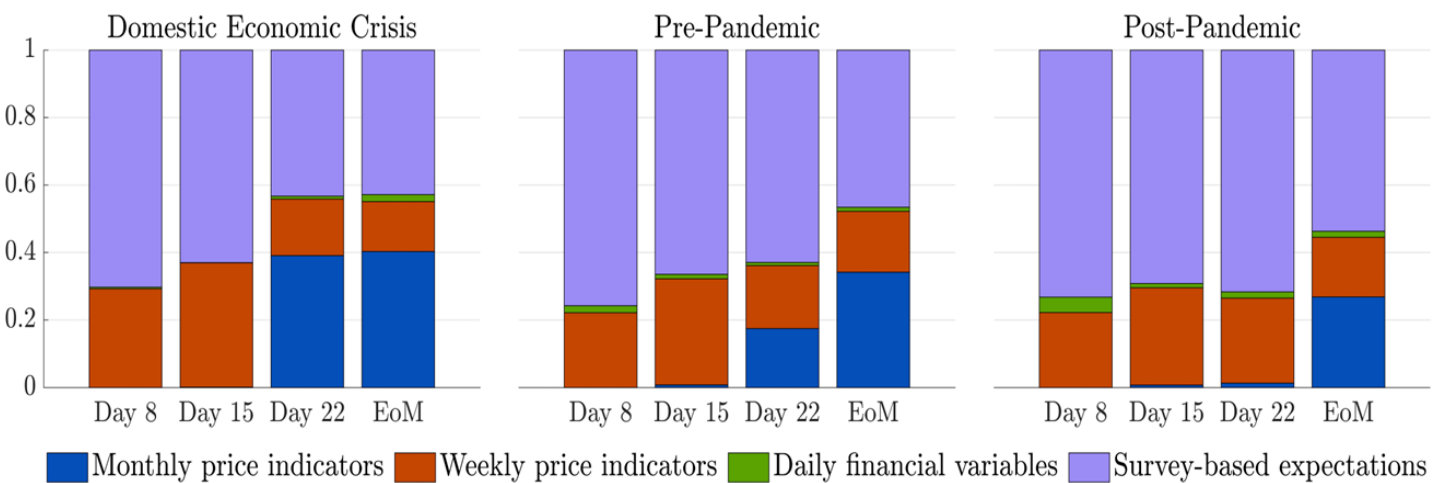
Notes: This Figure reports the weighted sum of absolute coefficient estimates fitted via the LASSO and grouped into different categories of predictors on days 8, 15, 22 and end-of-month (EoM). The "Domestic Economic Crisis" covers the period from 2013 to 2016 while March 2020 divides the "Pre-Pandemic" period and the start of the COVID-19 crisis ("Post-Pandemic").

Concluding Remarks

Machine learning methods are increasingly being adopted for macroeconomic nowcasting, especially in the context of managing high-frequency data from various economic sectors. This trend has been accelerated by the urgent need for accurate, real-time economic data following disruptive events such as the COVID-19 pandemic. Despite the popularity of ML in this field, there is still a significant gap in effectively using these techniques for real-time inflation nowcasting.

Our study specifically addresses this issue by comparing shrinkage methods with tree-based models within a scenario of persistent high inflation, highlighting the strengths of a well-crafted mixed-frequency ML framework. The study emphasizes the efficacy of variable selection through LASSO and the critical role of timely price indicators and expert judgment derived from SPF data in enhancing nowcast accuracy. The promising results suggest that these ML frameworks not only outperform SPF expectations but also surpass the accuracy of the top five SPF institutions, thereby presenting a strong case for further exploring ML methods against traditional econometric frameworks in economic forecasting. ▪

References

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*, 34.

Breitung, J. and Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach. *Journal of Forecasting*, 34(7):588–603.

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.

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.

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

Modugno, M. (2013). Now-casting inflation using high frequency data. International Journal of Forecasting, 29(4):664–675.

Monteforte, L. and Moretti, G. (2013). Real-Time Forecasts of Inflation: The Role of Financial Variables. *Journal of Forecasting*, 32(1):51–61.

Schnorrenberger, R., Schmidt, A., and Moura, G. V. (2024). Harnessing Machine Learning for RealTime Inflation Nowcasting. De Nederlandsche Bank Working Paper No. 806.

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.

About the authors

Richard Schnorrenberger is a researcher and teaching assistant at the Institute for Statistics and Econometrics of the Kiel University. His research focuses on time series forecasting models with applications in macroeconomics and finance. He has a PhD in econometrics from the Kiel University.

Aishameriane Schmidt is a PhD candidate at the Econometrics Institute from the Erasmus University Rotterdam & Tinbergen Institute in partnership with De Nederlandsche Bank. Her research interest lies in the intersection of machine learning models and macroeconomics.

Guilherme Valle Moura is professor at the Department of Economics and International Relations of the Federal University of Santa Catarina (UFSC) in Brazil. Before joining UFSC, he was assistant professor at Vrije Universiteit Amsterdam. He has a PhD from Kiel University.

SUERF Publications

Find more **SUERF Policy Notes and Briefs** at www.suerf.org/publications/suerf-policy-notes-and-briefs/