Inflation nowcasting in persistently high inflation environments

Aishameriane Schmidt
(Erasmus University/Tinbergen Institute/De Nederlandsche Bank)

Jointly with: Richard Schnorrenberger (Kiel University)
& Guilherme Valle Moura (Federal University of Santa Catarina)

EEA 2023 Conference, Barcelona

August 2023

Can we improve upon market-based inflation nowcasts in rising inflation periods with the help of ML?

- Persistent macroeconomic uncertainty and spiraling inflation expectations posed a challenge for forecasters, especially in light of recent events (e.g. Covid-19 pandemic and Russia's invasion of Ukraine).
- This uncertainty increases the demand for reliable inflation nowcasts . . .
  - policymakers can make better-informed decisions.
  - market participants can better adjust their investment strategies.
- How to deal with a high-dimensional mixed-frequency setting?

Can we improve upon market-based inflation nowcasts in rising inflation periods with the help of ML?

- Persistent macroeconomic uncertainty and spiraling inflation expectations posed a challenge for forecasters, especially in light of recent events (e.g. Covid-19 pandemic and Russia's invasion of Ukraine).

- This uncertainty increases the demand for reliable inflation nowcasts . . .
  - policymakers can make better-informed decisions.
  - market participants can better adjust their investment strategies.

- How to deal with a high-dimensional mixed-frequency setting?
  - We can ~~borrow~~ adapt methodologies from **GDP nowcasts**

Can we improve upon market-based inflation nowcasts in rising inflation periods with the help of ML?

- Persistent macroeconomic uncertainty and spiraling inflation expectations posed a challenge for forecasters, especially in light of recent events (e.g. Covid-19 pandemic and Russia's invasion of Ukraine).

- This uncertainty increases the demand for reliable inflation nowcasts . . .
  - policymakers can make better-informed decisions.
  - market participants can better adjust their investment strategies.

- How to deal with a high-dimensional mixed-frequency setting?
  - We can ~~borrow~~ adapt methodologies from **GDP nowcasts**
  - and combine with **machine learning methods**

This work

1. We construct a **real-time database** of key macro-financial variables for the Brazilian economy.
    - (most) Information is publicly available but scattered
    - We matched with release dates
    - And combine with predictions from professional forecasters

This work

1. We construct a **real-time database** of key macro-financial variables for the Brazilian economy.
   - (most) Information is publicly available but scattered
   - We matched with release dates
   - And combine with predictions from professional forecasters

2. We focus on two types of ML methods: **shrinkage vs. tree-based algorithms**.
   - Shrinkage methods, such as **LASSO** and the **sparse-group LASSO** (Babii et al. 2021), consistently outperform tree-based algorithms, like the random forest (Breiman 2001) and BART (Chipman et al. 2012), for nowcasting Brazilian inflation.

This work

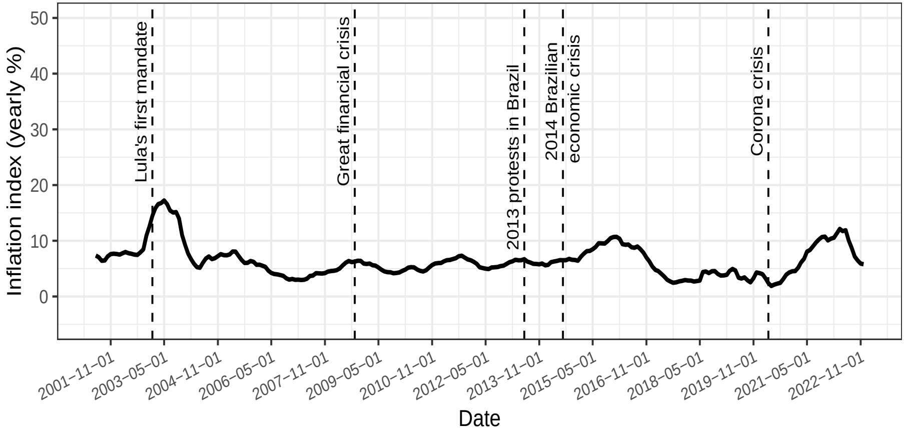
1. We construct a **real-time database** of key macro-financial variables for the Brazilian economy.
   - (most) Information is publicly available but scattered
   - We matched with release dates
   - And combine with predictions from professional forecasters

2. We focus on two types of ML methods: **shrinkage vs. tree-based algorithms**.
   - Shrinkage methods, such as **LASSO** and the **sparse-group LASSO** (Babii et al. 2021), consistently outperform tree-based algorithms, like the random forest (Breiman 2001) and BART (Chipman et al. 2012), for nowcasting Brazilian inflation.

3. We learn two key lessons for reliably updating inflation nowcasts:
   - Shrinkage-based models combined with **timely releases of non-official consumer price indices and market expectations** better anticipate the inflation surge following the COVID-19 pandemic.
   - In a mixed-frequency setting with ragged-edge data, the **real-time flow of data releases** must guide model specifications to produce good-quality nowcasts.

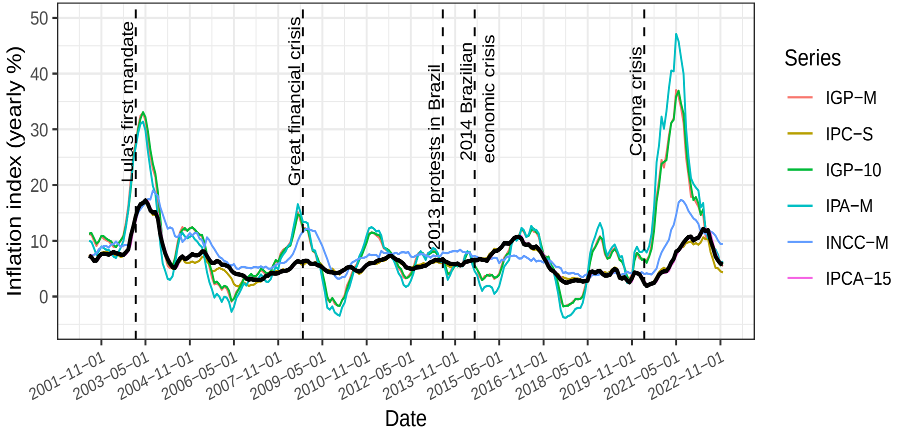
Brazil’s long history with high inflation rates might offer valuable insights . . .

Figure: Official Brazilian consumer price index (black line, IPCA).



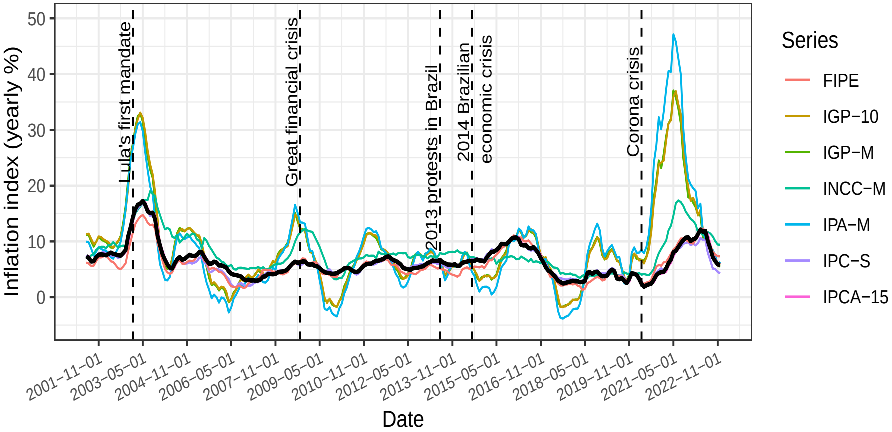
Brazil's long history with high inflation rates might offer valuable insights . . .

Figure: Official Brazilian consumer price index (black line) and non-official monthly price indices (colored lines).



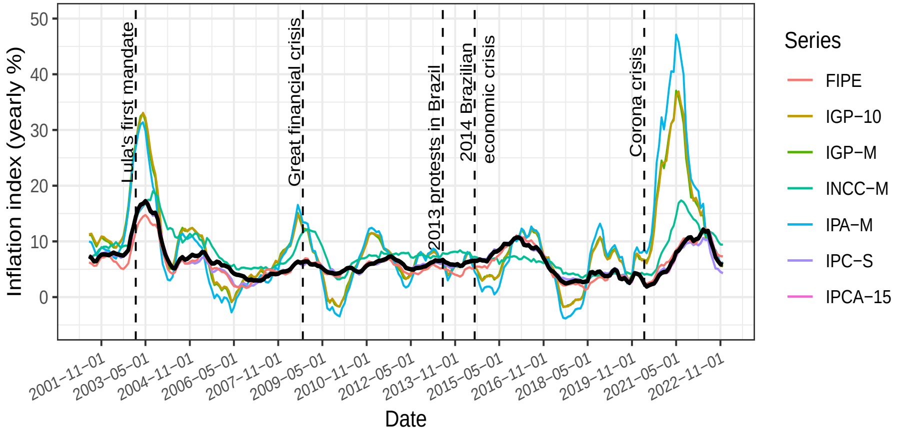
Brazil's long history with high inflation rates might offer valuable insights . . .

Figure: Official Brazilian consumer price index (black line) and non-official monthly and weekly price indices (colored lines).



Brazil's long history with high inflation rates might offer valuable insights . . .

Figure: Official Brazilian consumer price index (black line) and non-official monthly and weekly price indices (colored lines).



- Moreover, the Brazilian central bank conducts a **daily survey of professional forecasters**.

The Brazilian CB survey of professional forecasters (FOCUS)

- The Brazilian Central Bank (BCB) collets market expectations regarding different macro variables: GDP, exchange rate, consumer price indexes, interest rate, BoP variables, etc
  - The survey started in 1999
  - Currently about 160 institutions participate (banks, investment funds, academia)
  - The data collection includes monthly, quarterly, and annual forecasts, depending on the variable

The Brazilian CB survey of professional forecasters (FOCUS)

- The Brazilian Central Bank (BCB) collets market expectations regarding different macro variables: GDP, exchange rate, consumer price indexes, interest rate, BoP variables, etc
  - The survey started in 1999
  - Currently about 160 institutions participate (banks, investment funds, academia)
  - The data collection includes monthly, quarterly, and annual forecasts, depending on the variable
- Participants can inform their expectations on any single day
  - To be considered in the calculation for the next business day, information needs to be typed until 5pm
  - The system eliminates projections from institutions that have not had an input in the last 30 days

The Brazilian CB survey of professional forecasters (FOCUS)

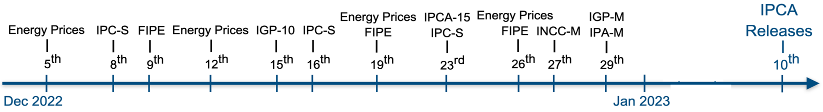
- The Brazilian Central Bank (BCB) collets market expectations regarding different macro variables: GDP, exchange rate, consumer price indexes, interest rate, BoP variables, etc
  - The survey started in 1999
  - Currently about 160 institutions participate (banks, investment funds, academia)
  - The data collection includes monthly, quarterly, and annual forecasts, depending on the variable
- Participants can inform their expectations on any single day
  - To be considered in the calculation for the next business day, information needs to be typed until 5pm
  - The system eliminates projections from institutions that have not had an input in the last 30 days
- There are two main series that are followed closely
  - The **median forecasts**, which are published every Monday, regularly at 8:30 AM, with data collected up to 5 pm of the previous Friday [our benchmark]
  - The average forecast from the **top-5** forecaster institutions, which are released monthly

The mixed-frequency price dataset (I) and timeline of data releases

Series	Mnemonic	Reference time span	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 CPI	IGP-M	21^{st}_{t-1} to 20^{th}_{t}	last week, reporting month	7	2003M1	FGV
General CPI - 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

The mixed-frequency price dataset (I) and timeline of data releases

Series	Mnemonic	Reference time span	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 CPI	IGP-M	21^{st}_{t-1} to 20^{th}_{t}	last week, reporting month	7	2003M1	FGV
General CPI - 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



The mixed-frequency dataset (II)

Series	Mnemonic	Reference time span	Publication timing	Avg. delay	Starting date	Source
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
Brazil credit default swaps	CDS	end of day	real-time	0	2007M12D19	B3
Bloomberg commodity index	BCOM	end of day	real-time	0	2003M1	Bloomberg
Daily expectations from the FOCUS survey of professional forecasters						
SPF (median)	FOCUS	full day	subsequent day	1	2003M1	BCB

Note: The reference time span relates to the data collection period. The publication timing provides the regular release calendar with respect to 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 in order to guarantee stationarity of the time series.
*DI-rates are yields of Brazilian interbank deposit future contracts negotiated at the B3 stock exchange.

Literature developments

1. Increasing availability of high-frequency macro-financial data to monitor the (now-)state of the economy . . .
   - broad evidence in real-time analysis of GDP growth, for instance, Giannone et al. (2008), Bańbura et al. (2012), Andreou et al. (2013), Bok et al. (2018), Cimadomo et al. (2021).
   - nevertheless, the recent literature on inflation nowcasting is making up for the disparity: Modugno (2013), Breitung & Roling (2015), Knotek & Zaman (2017), Aliaj et al. (2023).
2. Use of machine learning tools to forecast inflation dynamics . . .
   - effective solution to handle a large set of covariates: Garcia et al. (2017), Medeiros et al. (2021), **?**), Paranhos (2021), Clark et al. (2022), Araujo & Gaglianone (2022), Hauzenberger et al. (2023).
3. Complementing model-based forecasts with market-based expectations . . .
   - survey data of expert's forecasts as covariates in the forecasting model (Garcia et al. 2017, Banbura et al. 2021).

The nowcasting setup

We use an unrestricted MIDAS (U-MIDAS) approach to organize the mixed-frequency dataset:

$$\underbrace{\phi(L)\, y_{t+h}}_{\text{low-frequency target}} = c + \underbrace{\sum_{k=1}^{K} B(L^{1/m})\, x_{k,t}^{(m)}}_{\text{high-frequency predictors}} + \underbrace{\sum_{j=1}^{J} \alpha_j\, x_{j,t}}_{\text{low-frequency predictors}} + \underbrace{\sum_{i=1}^{11} \gamma_i\, d_{i,t+h}}_{\text{seasonal dummies}} + \varepsilon_{t+h} \tag{1}$$

where

matrix notation

- y_t represents the **official month-on-month CPI rate** with monthly time indices $t = 1.\dots, T$.
- $x_{k,t}^{(m)}$ is a weekly or daily predictor of inflation **sampled m times more frequently** than y_t.
- $B(L^{1/m}) = \sum_{i=0}^{m-1} \beta_{k,i}\, L^{i/m}$ are unrestricted distributed lag coefficients $\Rightarrow$ **linear model structure**.
- we assume a convenient **month/week frequency ratio of $m = 4$** with hypothetical information sets:

$$\underbrace{\textbf{Day 8:}}_{h=3/4}\; y_t \sim \left(y_{t-1}, x^{(m)}_{t-\frac{3}{4}}, x^{(m)}_{t-\frac{3}{4}}, x^{(m)}_{t-\frac{3}{4}}, x^{(m)}_{t-\frac{3}{4}}, x_t \right)$$

$$\underbrace{\textbf{Day 22:}}_{h=1/4}\; y_t \sim \left(y_{t-1}, x^{(m)}_{t-\frac{1}{4}}, x^{(m)}_{t-\frac{1}{4}}, x^{(m)}_{t-\frac{2}{4}}, x^{(m)}_{t-\frac{3}{4}}, x_t \right)$$

$$\underbrace{\textbf{Day 15:}}_{h=2/4}\; y_t \sim \left(y_{t-1}, x^{(m)}_{t-\frac{2}{4}}, x^{(m)}_{t-\frac{2}{4}}, x^{(m)}_{t-\frac{2}{4}}, x^{(m)}_{t-\frac{3}{4}}, x_t \right)$$

$$\underbrace{\textbf{EoM:}}_{h=0}\; y_t \sim \left(y_{t-1}, x^{(m)}_{t}, x^{(m)}_{t-\frac{1}{4}}, x^{(m)}_{t-\frac{2}{4}}, x^{(m)}_{t-\frac{3}{4}}, x_t \right)$$

Machine learning methods: shrinkage vs. tree-based models

Model	Short name	R function (package)	Tuning parameters and Cross-validation
LASSO	LASSO	glmnet (glmnet)	λ using timeslice cross validation
Elastic Net	EN	glmnet (glmnet)	α, λ using timeslice cross validation
Ridge	Ridge	glmnet (glmnet)	λ using timeslice cross validation
Sparse-Group LASSO	sg-LASSO	cv.sgl.fit (midasml)	α, λ using timeslice cross validation and optimal L oos
Random Forest	RF	randomForest (randomForest)	mtry using timeslice cross validation
Generalized Random Forest	GRF	regression_forest (grf)	sample fraction, mtry, min node size, honesty, α
Local Linear Forest	LLF	ll_regression_forest (grf)	sample fraction, mtry, minimum node size, honesty, α
Bayesian Additive Regression Trees	BART	rbart (rbart)	200 trees, 1000 posterior drws, d=0.95, prob of death = 0.7

Note: Time slice cross-validation (when used) starts with a 36-month window and subsequent 12-month fold slides.

Shrinkage methods Tree-based methods

For the evaluation . . .

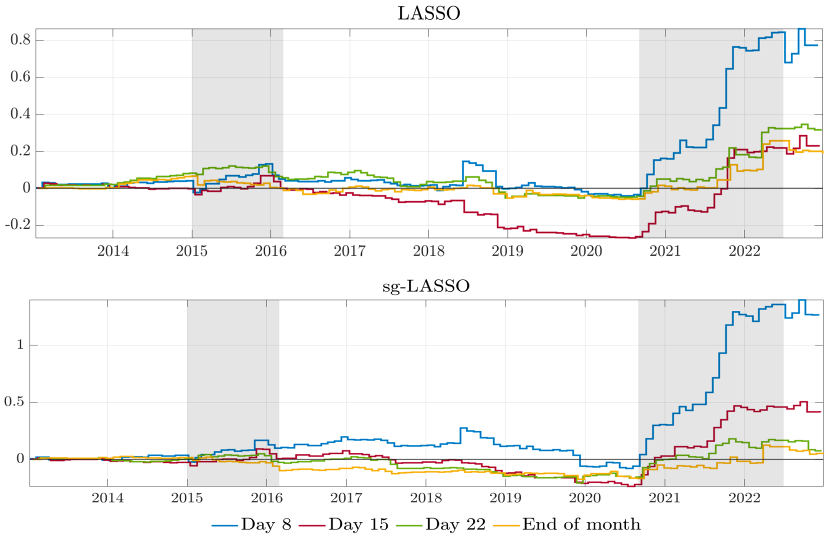
- we run the model at each nowcasting day (8, 15, 22 and end-of-month), whereas the predictor set is only composed of variables with available contemporaneous data by the time of the nowcast.
- we use a rolling window from Jan 2003 to Dec 2022 with an evaluation period starting in Jan 2013.
- we compare nowcasts by means of RMSE, MAE, CUMSFE and Fluctuation Tests.

RMSE and MAE relative to the survey of professional forecasters (SPF)

Metric	Horizon	SPF	LASSO	EN	Ridge	sgLASSO	RF	GRF	LLF	BART
RMSE	day 8	1	0.95	0.95	0.95	**0.92**	1.05	1.09	1.07	1.02
RMSE	day 15	1	0.98	0.98	1.03	**0.96**	1.21	1.26	1.25	1.17
RMSE	day 22	1	**0.94**	0.94	0.99	0.99	1.39	1.53	1.50	1.34
RMSE	end-of-month	1	**0.95**	0.95	1.03	0.99	1.55	1.77	1.79	1.48
MAE	day 8	1	0.96	0.96	0.98	**0.94**	1.05	1.09	1.09	1.04
MAE	day 15	1	0.99	0.99	1.06	**0.97**	1.21	1.24	1.26	1.16
MAE	day 22	1	**0.98**	0.98	1.02	0.98	1.36	1.47	1.45	1.32
MAE	end-of-month	**1**	1.01	1.02	1.11	1.06	1.53	1.71	1.73	1.46

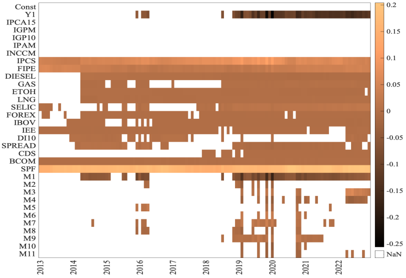
- Shrinkage methods generally perform better than SPF nowcasts. ▸ Fluctuation Test
- Relative performance increases with the nowcast horizon.
- Tree-based methods are not able to outperform SPF nowcasts.

Cumulative sum of loss differentials: shrinkage vs. market-based expectations

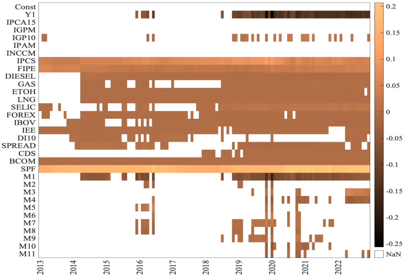


Note: The gray shaded areas correspond to rising inflation periods.

Variable selection in sg-LASSO (I)

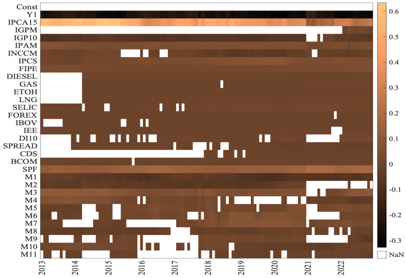


(a) Day 8



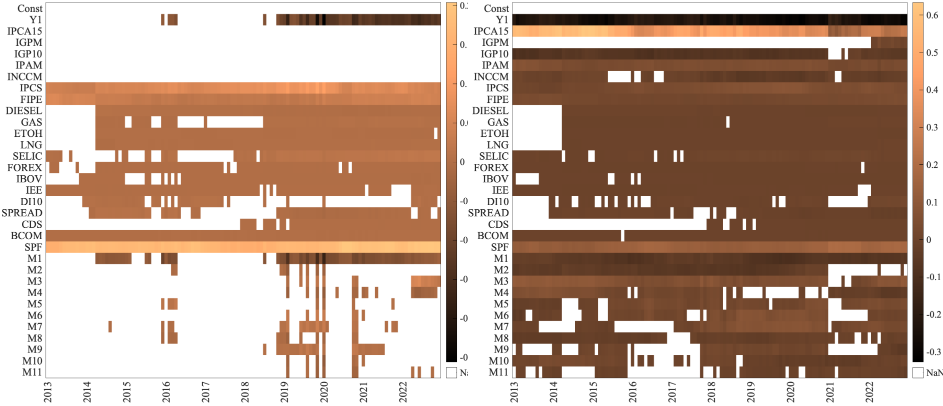
(b) Day 15

(c) Day 22



(d) End-of-month

Variable selection in sg-LASSO (II)



(a) Day 8 **(b)** End-of-month

Figure: Heatmap of coefficient estimates using one of the best performing methods: sg-LASSO ($L = 0$). Empty cells represent a coefficient estimate equal to zero, and thereby a predictor that has not been selected for a given period t in the evaluation period.

What else we tried (without improvements)

- PCA and LASSO before the random forests
- BART and HBART
- Estimation step only monthly followed by data updating weekly
- Other variables: unemployment, industrial production,
- Removing or using simple cross-validation instead of time-slice cv

Summary

What has been done so far . . .

- we study the usefulness of high-frequency macro-financial indicators to nowcast price developments in an environment characterized by persistently high inflation, namely the Brazilian economy of the past decades.

Summary

What has been done so far ...

- we study the usefulness of high-frequency macro-financial indicators to nowcast price developments in an environment characterized by persistently high inflation, namely the Brazilian economy of the past decades.

Key findings ...

- shrinkage-based models combined with timely releases of non-official consumer price indices and market expectations better capture the inflation surge following the Covid-19 pandemic.
- using the real-time flow of data releases to guide model specifications leads to higher-quality nowcasts.

Outlook

Next steps . . .

- increase the benchmarks models' pool: unobserved components model, DFM, Bayesian MF-VAR, boosting.
- replace the median from the Focus by the top-5
- evaluate mixed-frequency strategies that avoid the ragged-edge problem when updating nowcasts.
- get deeper into the “persistently high inflation" character of the study.

Inflation nowcasting in persistently high inflation environments

Aishameriane Schmidt (EUR/Tinbergen Institute/DNB)
More about me: http://www.aishameriane.com

This is a joint work with: R. Schnorrenberger (Kiel University) & G. Valle Moura (UFSC)

EEA 2023 Conference, Barcelona

August 2023

The U-MIDAS structure in matrix notation

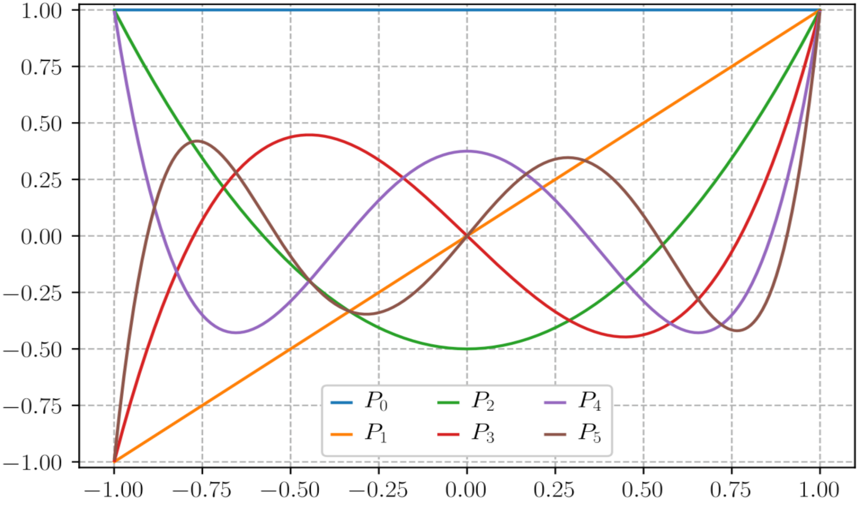
Conditional on data available up to month t of the nowcast, the unrestricted structure (1) has the following matrix representation (one-predictor case and neglecting seasonal dummies):

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & y_0 & x^{(m)}_{k,1} & x^{(m)}_{k,1-\frac{1}{4}} & x^{(m)}_{k,1-\frac{2}{4}} & x^{(m)}_{k,1-\frac{3}{4}} & x_{j,1} \\ 1 & y_1 & x^{(m)}_{k,2} & x^{(m)}_{k,2-\frac{1}{4}} & x^{(m)}_{k,2-\frac{2}{4}} & x^{(m)}_{k,2-\frac{3}{4}} & x_{j,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_{t-1} & \underbrace{x^{(m)}_{k,t}}_{\text{end-of-month}} & \underbrace{x^{(m)}_{k,t-\frac{1}{4}}}_{\text{day 22}} & \underbrace{x^{(m)}_{k,t-\frac{2}{4}}}_{\text{day 15}} & \underbrace{x^{(m)}_{k,t-\frac{3}{4}}}_{\text{day 8}} & x_{j,t-1} \end{bmatrix} \begin{bmatrix} c \\ \rho_1 \\ \beta_{k,1} \\ \beta_{k,2} \\ \beta_{k,3} \\ \beta_{k,4} \\ \alpha_j \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix} \tag{2}$$

Note that the matrix representation (2) makes explicit the transformation of the high-frequency predictor space into $m = 4$ low-frequency vectors. The model is estimated every week using days 8, 15, 22, and end-of-the-month (or the next business day) as reference.

Back

Legendre polynomials



Back

Shrinkage methods

1. **Standard shrinkage**: unstructured LASSO, ridge and elastic net.

   The hybrid `elastic net` estimator solves the penalized least-squares problem:

$$\hat{\beta} = \min_{\hat{\beta}} ||\mathbf{y} - \mathbf{X}\beta||^2 + \lambda \left(\alpha \, |\beta|_1 + \frac{(1-\alpha)}{2} \, ||\beta||^2 \right), \tag{3}$$

   - $\alpha \in (0, 1]$ is a weight parameter: `LASSO` ($\alpha = 1$) and `ridge regression` (as $\alpha \to 0$).
   - λ controls for shrinkage in the parameter space β.
   - λ and α determined in a data-driven way using cross-validation ($k = 5$ folds) for optimal performance.

   Caveats . . .

   - cannot cope with strongly correlated predictors $\Rightarrow$ treats each high-frequency lag separately.

2. **Sparse-group LASSO-MIDAS** (Babii et al. 2021):

$$\hat{\beta} = \min_{\hat{\beta}} ||\mathbf{y} - \mathbf{X}\beta||^2 + 2\lambda \underbrace{\left(\alpha \, |\beta|_1 + (1-\alpha) \, ||\beta||_{2,1} \right)}_{\text{LASSO shrinkage + group structure}} \tag{4}$$

   - $||\beta||_{2,1} = \sum_{G \in \mathcal{G}} |\beta_G|_2$ is the group LASSO norm and $\mathcal{G}$ is a **group structure**.
   - a group is defined as all high-frequency lags of a single covariate.
   - high-frequency lags are aggregated via predetermined Legendre polynomials of degree L.

$\Rightarrow$ **"All-in-One" solution**: handles high-dimensional mixed-frequency prediction problems.

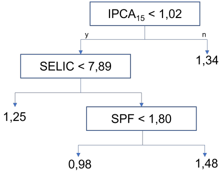
$\Rightarrow$ selects the relevant predictors and the appropriate high-frequency lag structure.

Back

Tree-based methods

1. **"Frequentist trees"**: Random Forest (Breiman 2001), Generalized Random Forest (Athey et al. 2019) and Local Linear Forest (Friedberg et al. 2020).

   A regression tree is a stepwise function based on binary partitions of selected covariates:



   - RF is an average of predictions from regression trees, decreasing the overfit of a single tree.
   - GRF uses predictions from an RF in a weighted regression problem.
   - LLF builts on GRF, but specifically uses a local linear regression for smoothness of the predictions.

2. **Bayesian Additive Regression Trees (BART)** (Chipman et al. 2012):

$$y_i = \sum_{j=1}^{m} \underbrace{g(x_i, \mathcal{T}_j, \mathcal{M}_j)}_{\text{maps } x_i \text{ to } \mu_{jk} \in \mathcal{M}_j} + \sigma\epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1) \tag{5}$$

   - $\mathcal{T}$ is the tree topology, $\mathcal{M}$ is a set of parameters μ_k, $k \in 1. \dots, K$.
   - BART induces sparsity via penalizing priors.
   - trees are *sequentially* combined to capture signals that remain in the residuals (similar to boosting).

Back

Fluctuation test (Giacomini and Rossi, 2010)