MIDAS Modeling for Core Inflation Forecasting

Luis Libonatti

Central Bank of Argentina

June 4, 2018

Ninth BIS CCA Research Conference,
“Microdata and Economic Research at Central Banks”

The opinions expressed in this presentation are those of the author and do not reflect the views of the Central Bank of Argentina

Introduction

- In the past decade, inflation in Argentina has been high and volatile, particularly in recent years when the economy stagnated.
- Since the year 2012, real GDP remained flat while price inflation averaged 25 %.
- In December 2015, the new administration abandoned altogether the capital controls imposed in the previous administration and unified the multiple exchange rate regimes under a new unique flexible scheme. This produced a sudden depreciation of the currency and a subsequent acceleration in the inflation rate in the coming months.
- The government also gradually started to eliminate subsidies on public utility services in order to stabilize the fiscal deficit.

Introduction

- In September 2016, the Central Bank formally introduced annual **inflation targets** to coordinate agents' expectations about the path of disinflation. This became operational starting in early 2017.
- Bounds and center points were set for end of period headline inflation rates for the years 2017 (17 %), 2018 (10 %) and 2019 (5 %) in terms of the "most" National Core Price Index available[1].
- In December 2017, the path of disinflation was revised upward. The target for the year 2018 was changed to 15 % and the target for 2019 was set to 10 %. From 2020 onward, the Central Bank would aim for a 5 % yearly inflation rate.

[1] A National CPI was released in July 2017.

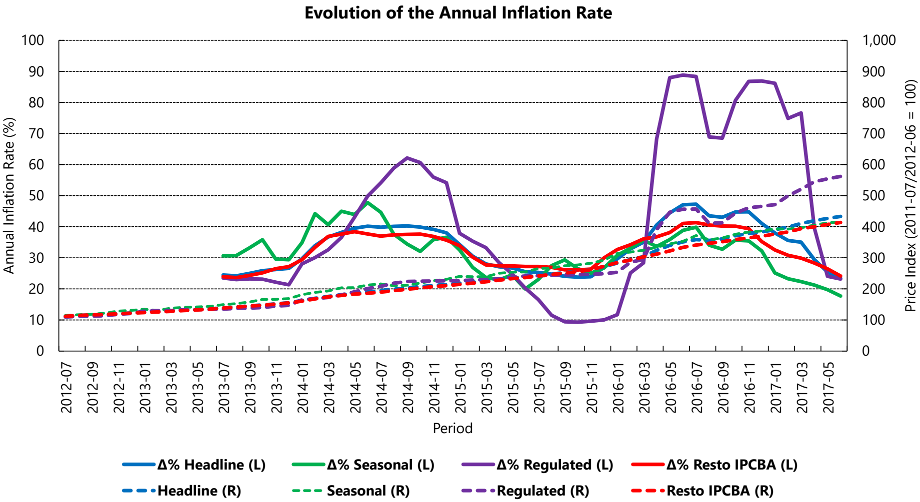
Introduction

- From 2007 to 2015, practically all statistical series emanating from the National Institute of Statistics and Censuses (INDEC) were considered unreliable by the public.
- In early 2016, INDEC resumed the publication of "quality" figures, including producer price indexes and a consumer price index for the Buenos Aires Metropolitan Area. However, as of today, these time series at most consist of less than two years of monthly data points.
- From an operational point of view, the monetary authorities closely monitor some (consumer) provincial price indexes and online price indexes by private economic consultants in addition to the above.

IPCBA

- A consumer price index that is particularly relevant for the monetary authorities is the consumer price index compiled by the Government of the Autonomous City of Buenos Aires (**IPCBA**), which falls under the category of provincial price index and covers only the City of Buenos Aires.
- A more restricted version of the index is also published, called "resto IPCBA" (**rIPCBA**), which serves as a measure of core inflation. It excludes products with strong seasonal patterns and regulated prices (e.g., public utility services) and represents 78 % percent of the headline index.
- Focusing on **rIPCBA** is relevant because the gradual phasing out of subsidies on public utility services severely distorts the headline index when there are price hikes in these regulated markets.
- **rIPCBA** is available from July 2012 onward and is released monthly, with a two-week publication lag.

Evolution of IPCBA



PriceStats

- According to the official website, the Billions Prices Project (BPP) at MIT is an academic initiative that uses prices collected from hundreds of online retailers around the world on a daily basis to conduct research in macro and international economics, and was founded in 2008 by Alberto Cavallo and Roberto Rigobón.
- PriceStats, a spin-off company that emerged from the BPP, is among the first organization, company or institution to produce real-time (daily) macro series. The company processes online prices for over 20 countries to track general price inflation and other related metrics (e.g., PPP series).
- Nowadays, this online price index is developed in cooperation with a finance research corporation, State Street Global Markets. Thus, I refer to it as the **SSPS** price index.

PriceStats

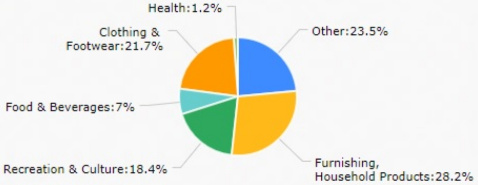
- The company aggregates online price information from a multiplicity of retailers' websites by "web scraping" and replicates the methodology of a traditional consumer price index to compile its daily online price index.
- For most countries, these indexes are only accessible with a subscription. However, the website **InflacionVerdadera.com** provides real-time information for Venezuela.
- Data for Argentina is available from November 1, 2007, with a three-day publication lag. Micro data is not available for the general public, but most aspects of the methodology can be found in Cavallo's original paper.
- Cavallo also provides comparisons between online and offline price indexes for Argentina, Brazil, Chile, Colombia, and Venezuela and shows that online price indexes can track the dynamic behavior of inflation rates over time fairly well, with the exception of Argentina.

Composition of the SSPS Price Index

ECONOMIC SECTORS

PriceStats gathers price information for key economic sectors: food & beverages; furnishing & household products; recreation & culture; clothing & footwear; housing, electricity & fuel; and health. We continue to update this as more data becomes available.

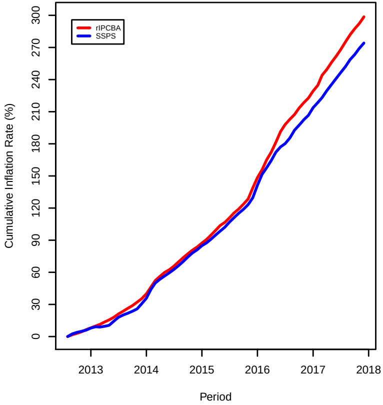
This chart includes information from all countries where PriceStats collects data.

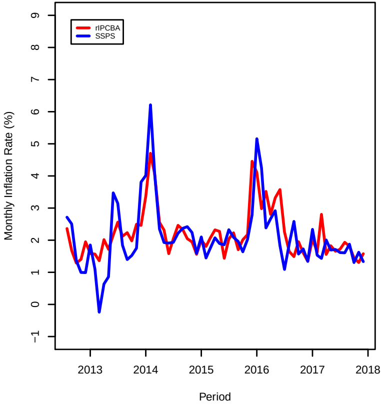


Updated: February 2018

Comparison of Price Indexes

- The following plots present a comparison between **rIPCBA** inflation and **SPSS** inflation aggregated at monthly frequency.





MIDAS Regression Models: Motivation

- Providing the monetary authorities with short term forecasts would aid policymakers to keep track of the disinflation process and act in accordance.
- **rIPCBA** and **SSPS** are both series that are closely monitored by the Central Bank. A combination of both in a single model could produce better short term forecasts of the monthly inflation rate.
- Mixed-frequency regression models offer a convenient arrangement to accommodate variables sampled at different frequencies.
- Ghysels, Santa-Clara and Valkanov (2004) popularized these models, known as Mixed Data Sampling Regression Models (MIDAS). Originally intended for financial time series, this framework quickly turned to macroeconomics.

MIDAS Regression Models: Structure

- Suppose that y_t is a **low-frequency** process and that x_τ is a **high-frequency** process that is observed a discrete and fixed number of times m each time a new value of y_t is observed. Then, both series can be combined in a simple regression equation

$$y_t = \sum_{j=0}^{m-1} \theta_j x_{t-j/m} + u_t.$$

  Here I have used the fact $\tau = t - 1 + j/m$ for $j = 1, \ldots, m$ since m is fixed, where $x_{t-0/m}$ would be the most recent observation.

- This structure conceals a **high-frequency** lag polynomial $\theta(L^{1/m}) \equiv \sum_{j=0}^{m-1} \theta_j L^{j/m} x_t$ so that $L^{j/m} x_t = x_{t-j/m}$ is similar to a distributed lags model.

MIDAS Regression Models: Structure

- It is perhaps easier to visualize the problem if we write the equation in matrix form:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{T-1} \\ y_T \end{bmatrix}}_{\boldsymbol{y}} = \underbrace{\begin{bmatrix} x_1 & \cdots & x_{1-(m-1)/m} \\ x_2 & \cdots & x_{2-(m-1)/m} \\ \vdots & \ddots & \vdots \\ x_{T-1} & \cdots & x_{T-1-(m-1)/m} \\ x_T & \cdots & x_{T-(m-1)/m} \end{bmatrix}}_{\boldsymbol{X}} \underbrace{\begin{bmatrix} \theta_0 \\ \vdots \\ \theta_{m-1} \end{bmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{T-1} \\ u_T \end{bmatrix}}_{\boldsymbol{u}}.$$

- In principle, this model can be estimated by OLS and this would the U-MIDAS model, but if m is really large relative to the sample size T, then the model will suffer from parameter proliferation.

MIDAS Regression Models: Structure

- Parameter proliferation can be overcome if restrictions are imposed on the coefficients of the **high-frequency** lag polynomial and restate each θ_j as a function of some q hyperparameters and its subindex j (its position within the **low-frequency** lag polynomial) in such a way that $q \gg m$.
- Each θ_j is redefined as $\theta_j \equiv w_j(\gamma; j)$ where the vector γ is the collection of q hyperparameters that characterize the weight function $w_j(\cdot)$. The previous equation is transformed to,

$$y_t = \lambda \sum_{j=0}^{m-1} \left(\frac{w_j(\gamma; j)}{\sum_{j=0}^{m-1} w_j(\gamma; j)} \right) x_{t-j/m} + u_t,$$

where λ is an impact parameter and the weights are usually normalized so that they sum up to unity.

MIDAS Regression Models: Structure

- Some weight functions that allow for many different shapes found in the literature include

$$w_j(\gamma_1, \ldots, \gamma_q; j) = \sum_{s=1}^{q} \gamma_s j^s, \qquad \text{(Almon)}$$

$$w_j(\gamma_1, \ldots, \gamma_q; j) = \exp\left(\sum_{s=1}^{q} \gamma_s j^s\right), \qquad \text{(Exp. Almon)}$$

$$w_j(\gamma_1, \gamma_2; j) = z_j^{\gamma_1 - 1}(1 - z_j)^{\gamma_2 - 1}, \qquad \text{(Beta)}$$

  with $z_j \equiv j/(m-1)$, $\gamma_1 > 0$ and $\gamma_2 > 0$. A frequent choice for q is $q = 2$ or $q = 3$ at most.

- Even "nonparametric" restrictions are also possible (Breitung and Rolling, 2015). This closely resembles ridge regression.

MIDAS Regression Models: Forecasting

- MIDAS models are intended as a direct forecasting tool. Estimation depends on the time displacement of the variables, $d \in \mathbb{Q}$, and the forecast horizon $h \in \mathbb{N}$. The direct strategy requires estimation of as many models as per pair (d, h) is required.
- If T_Y is the time index of latest y_t available for **estimation**, and T_X is the time index of the latest x_τ available for both **estimation and forecasting**, then d can be defined as $d \equiv T_Y - T_X$. If $W(L^{1/m};\gamma) \equiv \sum_{j=0}^{m-1} w_j(\gamma;j)L^{1/m}$, a forecast can be computed with,

$$\hat{y}_{T+h} = \hat{\lambda}_{d,h} W(L^{1/m};\hat{\gamma}_{d,h}) x_{T-d}.$$

- For example, the "nowcast" can be retrieved when $d = -1$ and $h = 1$.
- The fact that d is a rational number implies that it is possible to generate intra-period forecasts.
- To arrive at the previous equation, it would be necessary to estimate,

$$y_t = \lambda W(L^{1/m};\gamma) x_{t-h-d} + u_t.$$

MIDAS Regression Models: Extensions

- The basic MIDAS model can be extended by including p_X lags of the vector $\boldsymbol{x}_t \equiv [x_t \;\; \dots \;\; x_{t-(m-1)/m}]$ that collects all the x_τ corresponding to period t, totaling $m \times L_X$ high-frequency variates where $L_X = p_X + 1$

$$y_t = \sum_{r=0}^{p_X} \sum_{j=0}^{m-1} \theta_{r,j} x_{t-r-j/m} + u_t. \qquad \text{(MIDAS-DL)}$$

  where in general a single weight function for all L_X variates is employed.

- Additionally, another important modification is the inclusion of **low-frequency** autoregressive augmentations. This can be done in two ways. Clements and Galvão (2008) favor the latter

$$y_t = \phi y_{t-1} + \sum_{r=0}^{p_X} \sum_{j=0}^{m-1} \theta_{r,j} x_{t-r-j/m} + u_t, \qquad \text{(MIDAS-ADL)}$$

$$(1-\phi L) y_t = (1-\phi L) \sum_{r=0}^{p_X} \sum_{j=0}^{m-1} \theta_{r,j} x_{t-r-j/m} + u_t. \qquad \text{(MIDAS-ADL-CF)}$$

MIDAS Regression Models: Caveats

- The constraints on the parameters turn the model nonlinear and in most cases, there is not a closed form solution. Numerical optimization methods and nonlinear least squares are required.
- The functional constraints are highly likely to introduce a bias/variance trade off due to model misspecification.
- However, Ghysels, Kvedaras and Zemlys (2016) argue that both, parameter estimation precision and out-of-sample forecast accuracy, gained by the increase in degrees of freedom, far offset the effects of the bias generated by misspecified constraints.
- If x_τ is not sampled at regular intervals (e.g., a monthly/daily combination), m needs to be treated as a fixed number.

Forecasting Exercise

- MIDAS regressions turn out to be intuitive for combining **rIPCBA** and **SSPS** since the monthly inflation rate can be approximately decomposed as the aggregation of daily inflation rates of the corresponding month, when evaluated in logarithmic differences,

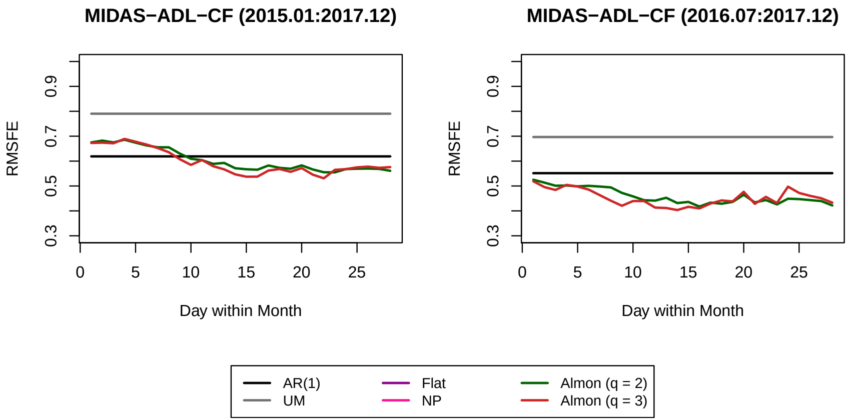
$$\pi_t^m \approx \sum_{\tau \in t} (\log p_\tau^d - \log p_{\tau-1}^d).$$

- In a first stage, the out-of-sample performance of the MIDAS-DL, MIDAS-ADL, and MIDAS-ADL-CF models for various weight functions with $q = 2$ and $q = 3$ are compared with $h \in \{1, 2, 3\}$ and for $L_X \in \{1, 2, 3\}$, with a balanced dataset in terms of RMSFE.
- Flat aggregation and the nonparametric model mentioned earlier are also considered.
- A balanced dataset means that there is exact frequency matching: m daily observations from the same month or L_X groups of m daily observations from the same months correspond to a specific **low-frequency** monthly observation of the dependent variable.

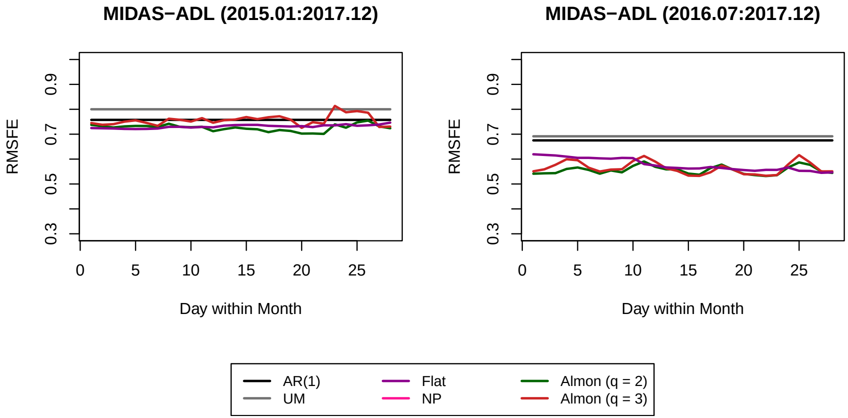
Forecasting Exercise

- In total, two sets of RMSFE were computed, one corresponding to the large sample (36 obs.) and the other to a reduced subsample (18 obs.) using recursive (expanding) windows.
- The frequency of **SPSS** was assumed fixed at $m = 28$, while $d = -1$; so days 29, 30 and 31 of each month are discarded.
- In a second stage, intra-period forecasts are computed for the best selected models for each L_X based on the results of the large sample.
- When intra-period forecasts are computed, d is a fraction in the interval $[-1, 0)$. More specifically, $d = -1 + i/m$ for $i \in 1, \ldots, m$, where m is again the frequency.
- The performance of the forecasts from an AR(1) and the unconditional mean are used as benchmarks. These remain the same throughout the month.

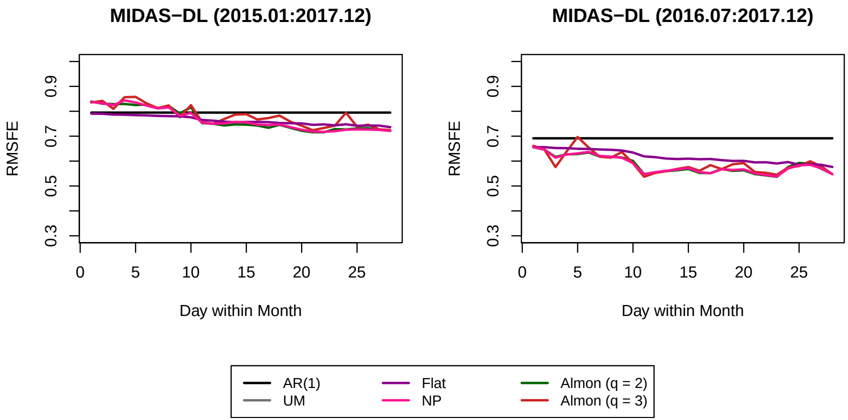
Evolution of the RMSFE for Horizon $h = 1$ within a Month for Selected Models with $L_X = 3$



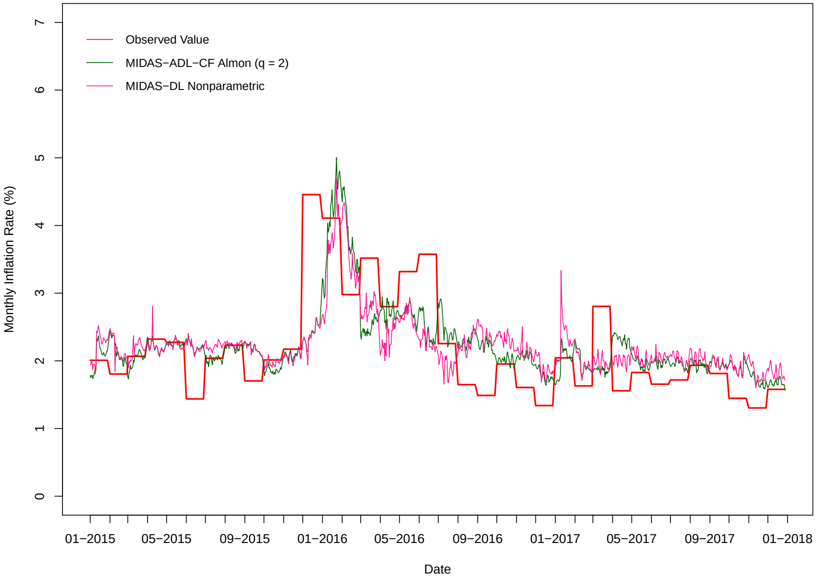
Evolution of the RMSFE for Horizon $h = 2$ within a Month for Selected Models with $L_X = 3$



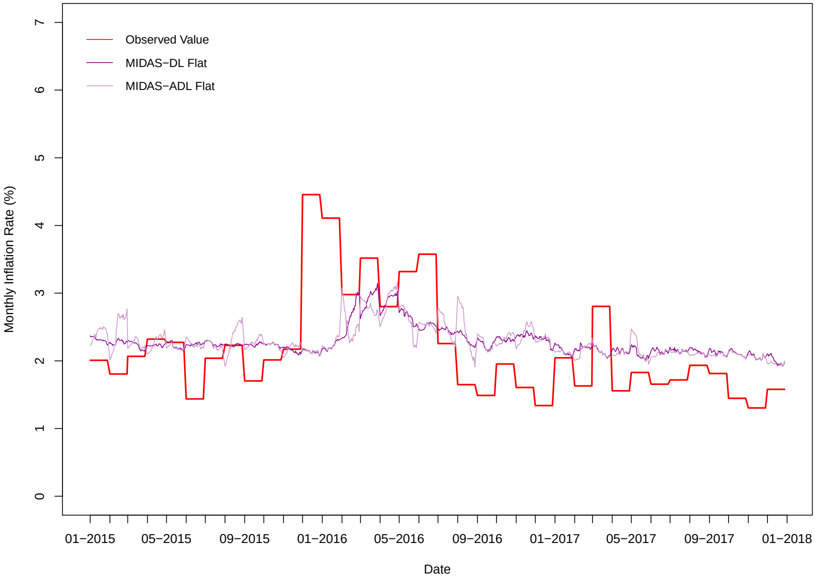
Evolution of the RMSFE for Horizon $h = 3$ within a Month for Selected Models with $L_X = 1$



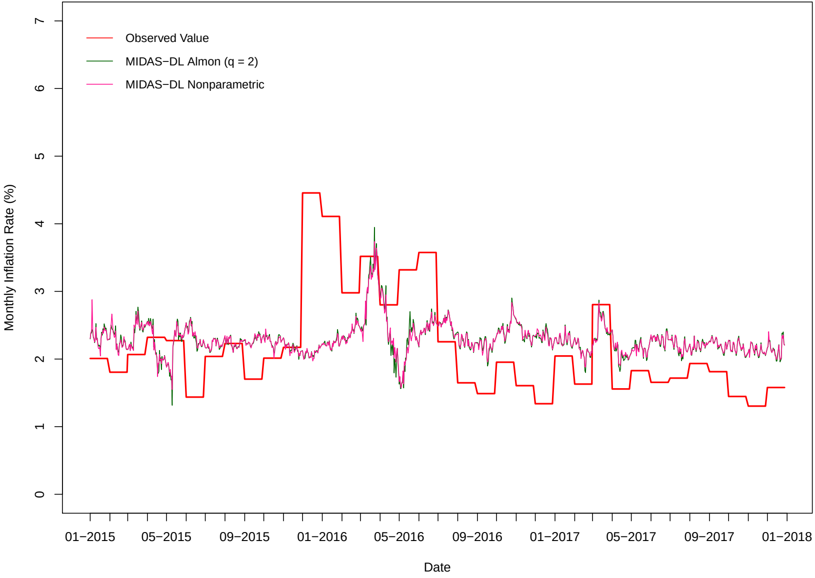
One-Step-Ahead Intra-Period Forecasts



Two-Step-Ahead Intra-Period Forecasts



Three-Step-Ahead Intra-Period Forecasts



Results and Final Thoughts

- For $h = 1$, in day 1 to day 28 point-to-point comparison, the RMSFE is reduced by approximately 20 % and particularly, in the second half of the month, the models start to surpass the accuracy of the autoregression by up to 15 % for some days.
- For $h = 2$ this is less evident. In the case of $h = 3$, a similar pattern of increasing forecast accuracy emerges.
- When considering the subsample, the results suggest that it is possible to obtain even better forecasts as the inflation rate stabilizes.
- Despite intra-period forecasts evidencing some volatility within the month, this does not seem to be a major concern as inflation stabilizes at the end of the sample.
- Nevertheless, the short length of the time series of Argentina difficult arriving a robust conclusion about the predictive potential of the daily series and the MIDAS models.