MIDAS Modeling for Core Inflation Forecasting

Luis Libonatti

Abstract

This paper presents a forecasting exercise that assesses the predictive potential of a daily price index based on online prices. Prices are compiled using web scraping services provided by the private company PriceStats in cooperation with a finance research corporation, State Street Global Markets. This online price index is tested as a predictor of the monthly core inflation rate in Argentina, known as "resto IPCBA" and published by the Statistics Office of the City of Buenos Aires. Mixed frequency regression models offer a convenient arrangement to accommodate variables sampled at different frequencies and hence many specifications are evaluated. Different classes of these models are found to produce a slight boost in out-of-sample predictive performance at immediate horizons when compared to benchmark naïve models and estimators. Additionally, an analysis of intra-period forecasts, reveals a slight trend towards increased forecast accuracy as the daily variable approaches one full month for certain horizons.

Keywords: MIDAS, distributed lags, core inflation, forecasting.

JEL Classification: C22, C53, E37.

The author is affiliated with the Central Bank of Argentina. E-mail: luis.libonatti@bcra.gob.ar. This research was undertaken within the framework of CEMLA's Joint Research Program 2017 coordinated by the Central Bank of Colombia. The author gratefully acknowledges counseling and technical advisory provided by the Financial Stability and Development (FSD) Group of the Inter-American Development Bank in the process of writing this document. The opinions expressed in this publication are those of the author and do not reflect the views of CEMLA, the FSD group, the InterAmerican Development Bank or the Central Bank of Argentina.

1. INTRODUCTION

Forecasting inflation has become increasingly important in Argentina as it is essential for economic agents to adjust wages and prices—particularly in recent years—in a context of high and volatile inflation. Having timely updates about the future trajectory of the inflation rate is essential for conducting monetary policy, specially, since the Central Bank is transitioning towards an inflation targeting regime. Recent developments in the use of "big data" have greatly facilitated tracking macroeconomic variables in real-time. A remarkable example is the construction of online price indexes that are sampled daily, rather than monthly, as it is standard for traditional price indexes from statistical offices. The question naturally arises of whether this information can help predict the future trajectory of traditional consumer price indexes. Ghysels *et al.* (2004) introduced a regression framework that allows for the exploitation of time series sampled at different frequencies, known in the literature as Mixed Data Sampling (MIDAS) regression models. The methodology reduces to fitting a regression model to a low-frequency variable using high-frequency data as regressors. As it will be shown later, this technique closely resembles distributed lag models. This paper employs this methodology to assess whether the combination of price series sampled at different frequencies is an effective tool for improving forecast accuracy compared to naïve models, using the online price index constructed by PriceStats in cooperation with State Street Global Markets.

The rest of the paper is organized as follows. In the next section, a brief introduction to MIDAS models is presented. In the third section, existing theoretical research on MIDAS regressions as well as some applications in forecasting inflation are briefly reviewed. In the fourth section, the forecasting exercise is described, and the results are discussed. And finally, the fifth section concludes.

2. MIDAS REGRESSION MODELS

MIDAS regression models propose a data-driven method to aggregate high frequency variables into lower-frequency predictors. They provide an alternative to the well-known "bridge" approach (Schumacher, 2016) in which high frequency variables are aggregated with equal weights (flat aggregation).[1] Ghysels *et al.* (2004) suggested combining y_t, a low frequency process, and x_τ, 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, in a plain regression equation,

2.1
$$y_t = \sum_{(j=0)}^{(m-1)} \theta_j x_{t-j/m} + u_t,$$

or more compactly,

2.2
$$y_t = \left(\theta' x_t'\right)' + u_t$$

where $\boldsymbol{x}_t \equiv \left[x_t \cdots x_{t-(m-1)/m} \right]$ is a $1\times m$ row vector that collects all the x_τ corresponding to period t and $\boldsymbol{\theta} \equiv \left[\theta_0 \ldots \theta_{m-1}\right]'$ is the $m\times 1$ vector of weight coefficients.[2] Each j high frequency observation $x_{t-j/m}$ within the low frequency period t enters the model linearly as a variable accompanied by its specific weight, θ_j, totaling m explanatory variables and m weights, plus an error term. The high frequency subindex τ needs to be represented in terms of the low frequency index t by noting that $\tau = t-1+j/m$ for $j=1,\ldots,m$ since m is fixed, where $x_{t-0/m}$ would be the most recent observation. This structure actually conceals a high frequency lag polynomial $\theta\left(L^{1/m}\right) \equiv \sum_{j=0}^{m-1} \theta_j L^{j/m} x_t$ so that $L^{j/m} x_t = x_{t-j/m}$ is similar in fashion to a distributed lags model.

To provide a clearer perspective, it is perhaps easier to introduce matrix notation. Defining $X \equiv [x_1' \ldots x_T']$ as the $T\times m$ matrix that groups all the $\boldsymbol{x}_t$ vectors together; $\boldsymbol{y} \equiv \left[y_1 \cdots y_T \right]'$, the collection of the low frequency observations of size $T\times 1$; and $u \equiv [u_1 \ldots u_T]$ the residuals of the same length as $\boldsymbol{y}$, it is possible to unveil a simple multiple regression equation,

1 In fact, this can be considered a special case of a MIDAS regression.

2 This equation may also include constants, trends, seasonal terms or other low frequency explanatory variables.

2.3

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

Indeed, this problem can be solved by ordinary least squares (OLS) and this method will produce consistent coefficient estimates. Equation (2.1) is usually referred to as the unrestricted MIDAS regression model (U-MIDAS).[3] However, an inconvenience arises when m, the length of the vector θ, is large relative to the sample size T, as is usually the case in MIDAS regressions. When this occurs, the models suffer from parameter proliferation and OLS induces poor estimates and consequently, poor forecasts. A straightforward way to overcome this deficiency is to impose restrictions 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 $w_j(\bullet)$. Equation (2.1) is transformed to,

2.4

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

where λ is an impact parameter and the weights are normalized so that they sum up to unity. Ghysels *et al.* (2004) initially recommended what is known as the exponential Almon polynomial as a candidate for weight function as it allows for many different shapes and depends only on a few parameters. This is an exponentiated version of an Almon lag polynomial, which is well known in the distributed lags literature,[4]

---

[3] Foroni et al. (2015) present a detailed assessment of this strategy.

[4] See for example the book by Judge *et al.* (1985).

2.5 $$w_j\left(\gamma_1,\ldots,\gamma_q;j\right)=e^{\sum_{s=1}^{q}\gamma_s j^s}.$$

Another conventional candidate is the beta probability density,

2.6 $$w_j\left(\gamma_1,\gamma_2;j\right)=z_j^{\gamma_1-1}\left(1-z_j\right)^{\gamma_2-1},$$

with $z_j\equiv j/\left(m-1\right),\gamma_1>0$ and $\gamma_2>0$.

Parameterization as in equation (2.5) has proved to be quite popular and has become the standard among researchers, particularly when *q=2*.

The introduction of constrained coefficients has many far-reaching implications. The model turns nonlinear and lacks a closed form solution. It is necessary to resort to nonlinear least squares and approximate the solution by numerical optimization routines. Additionally, the constraints are highly likely to introduce a bias in each θ_j. However, based on Monte Carlo simulations, when the sample size is small relative to the number of parameters, Ghysels *et al.* (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.

MIDAS models are generally intended as a direct forecasting tool since this could prove to be more robust against misspecification (Marcellino *et al.*, 2006). This implies that estimation additionally depends on the time displacement of the variables, $d\in\mathbb{Q}$, and the forecast horizon,[5] $h\in\mathbb{N}$. 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 $d\equiv T_Y-T_X$. Setting $W\left(L^{1/m};\boldsymbol{\gamma}\right)\equiv\sum_{j=0}^{m-1}w_j\left(\boldsymbol{\gamma};j\right)L^{1/m}$, a forecast can be computed with,

2.7 $$\hat{y}_{T+h}=\hat{\lambda}_{d,h}W\left(L^{1/m};\hat{\boldsymbol{\gamma}}_{d,h}\right)x_{T-d}.$$

[5] How many periods into the future it is necessary to forecast.

The "nowcast" can be retrieved when d=–1 and h=1. Note also that, the fact that d is a rational number implies that it is possible to generate intra-period forecasts.

To arrive at equation (2.7), it is first necessary to estimate,

2.8
$$y_t = \lambda W\left(L^{1/m};\hat{\gamma}\right)x_{t-h-d} + u_t,$$

and then compute $\hat{y}_{T+h}$ with the estimated parameters, $\hat{\lambda}_{d,h}$ and $\hat{\gamma}_{d,h}$, and the vector x_{T-d}.

It is possible to extend the MIDAS model by allowing for more than m high frequency regressors. For example, by including p_X lags of the vector x_t totaling $m \times L_X$ high frequency variates where $L_X = p_x + 1$, the MIDAS-DL model is formed,

2.9
$$y_t = \sum_{r=0}^{p_X}\left(\boldsymbol{\theta}_r' \boldsymbol{x}_{t-r}'\right)' + u_t,$$

or equivalently,

2.10
$$y_t = \sum_{r=0}^{p_X}\sum_{j=0}^{m-1}\boldsymbol{\theta}_{r,j}x_{t-r-j/m} + u_t.$$

In matrix notation, this can be represented by,

2.11

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{T-1} \\ y_T \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_{1-(m-1)/m} & \cdots & x_{1-p_X} & \cdots & x_{1-p_X-(m-1)/m} \\ x_2 & \cdots & x_{2-(m-1)/m} & \cdots & x_{2-p_X} & \cdots & x_{2-p_X-(m-1)/m} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{T-1} & \cdots & x_{T-1-(m-1)/m} & \cdots & x_{T-p_X-1} & \cdots & x_{T-p_X-1-(m-1)/m} \\ x_T & \cdots & x_{T-(m-1)/m} & \cdots & x_{T-p_X} & \cdots & x_{T-p_X-(m-1)/m} \end{bmatrix} \begin{bmatrix} \theta_{0,0} \\ \vdots \\ \theta_{0,m-1} \\ \vdots \\ \theta_{p_X,0} \\ \vdots \\ \theta_{p_X,m-1} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_1 \\ \vdots \\ u_{T-1} \\ u_T \end{bmatrix}.$$

If different weight functions for each θ_r in equation (2.9), then the multiplicative or aggregates-based MIDAS model is obtained (Ghysels *et al.*, 2016). On the contrary, employing a single weight function for all $m \times L_X$ coefficients vectors θ_r is also possible. The first method allows for greater flexibility but at the cost of more parameters to estimate, so this possibility will not be considered, as this may not be convenient for a very short sample size.

Other possible extensions include constructing high frequency factors (Marcellino and Schumacher, 2010), incorporating cointegration relations (Miller, 2013), integrating Markov switching (Guérin and Marcellino, 2013), estimating multivariate models (Ghysels *et al.*, 2007), using infinite polynomials (Ghysels *et al.*, 2007) or adding low frequency autoregressive augmentations (Ghysels *et al.*, 2007; Clements and Galvão, 2008; Duarte, 2014), for example. Foroni and Marcellino (2013) provide a comprehensive survey of possible extensions in a recent survey about mixed frequency models.

3. LITERATURE REVIEW

Clements and Galvão (2008) were among the first to study applications of MIDAS regressions to macroeconomic variables. In their paper, they forecast U.S. real quarterly output growth in combination with three different monthly variables: *i)* industrial production, *ii)* employment growth, and *iii)* capacity utilization. They find a slight increase in out-of-sample forecast accuracy with both vintage and revised data compared to two benchmarks models, an autoregression and an ADL model in particular, for short-term horizons. They also derive and assess a model with autoregressive dynamics introduced as a common factor shared by the low and the high-frequency lag polynomials. Based on comments by Ghysels *et al.* (2007), they argue that including an autoregressive term in a standard MIDAS model, as in the next equation,

3.1 $$y_t = \phi y_{t-1} + \lambda W\left(L^{1/m};\gamma\right) x_t + u_t,$$

induces a seasonal response from y_t to x_t irrespective of whether x_t exhibits a seasonal pattern. They suggest further restricting the model by adding a common lag polynomial shared between y_t and x_t,

$$(1-\phi L)y_t = \lambda(1-\phi L)W(L^{(1/m)};\gamma)x_t + u_t, \tag{3.2}$$

so that when writing the model in distributed lag representation, the polynomial in L cancels out, eliminating the spurious seasonal response. A multi-step generalization of (3.2) for h-step-ahead forecasts would be,

$$\left(1-\phi L^h\right)y_t = \lambda\left(1-\phi L^h\right)W\left(L^{1/m};\gamma\right)x_t + u_t. \tag{3.3}$$

Armesto *et al.* (2010) analyze the performance of MIDAS models for the US economy for four different variable combinations: *i)* quarterly GDP growth and monthly employment growth; *ii)* monthly CPI inflation and daily Fed funds rate; *iii)* monthly industrial production growth and a measure of term spread, and *iv)* employment growth and again a measure of term spread. They contrast the results of flat aggregation, the exponential Almon polynomial and a step weight function, but are unable to find a dominant model specification. They provide detailed results for one-step-ahead intra-period forecasting performance of the models, computed by accumulating leads[6] as the high frequency variable approaches a full low frequency period. They find an erratic pattern for the root mean square forecast error (RMSFE) of the models as a function of the leads included in the regression. Thus, in a real-time setting, which intra-period forecasts could be the most accurate would not be trivial.

Monteforte and Moretti (2013) develop MIDAS models to forecast the euro area harmonized price index inflation. They put forward a two-step approach involving low and high frequency variables. In the first place, they estimate a generalized dynamic factor model (Forni *et al.*, 2000) for the inflation rate based on a set of variables,

[6] In this instance "lead" refers to an observation of the high-frequency predictor that corresponds to the same temporal period of the low frequency variable.

and then they extract a common component and separate that into a long-run and a cyclical, or short-run, component. The second step consists in fitting the model of Clements and Galvão (2008) to capture short-term dynamics and use financial time series as high frequency regressors, in addition to the long-run component previously estimated as well as other low frequency variables. They design three MIDAS models, M1, M2 and M3, each with different high frequency regressors: *i)* M1 includes the short-term interest rate, changes in interest rate spread and oil future prices; *ii)* M2 uses changes in the wheat price, oil future quotes and the exchange rate; and finally, *iii)* M3 consists of long-term rates, changes in the interest rate spreads, and changes in the short-term rate. They contrast the out-of-sample performance in terms of RSMFE of these models against the equations for the inflation rate of two different low frequency vector autoregressions, and univariate random walks, autoregressions and autoregressive-moving average models. They compute all the intra-period forecasts for the MIDAS models and the monthly average of these daily forecasts, and compare this average to all the low frequency models. All the analysis is conducted for one-month-ahead and two-month-ahead forecasts. They find on average a 20% reduction in forecast error dispersion. The authors also provide a final empirical exercise by using forecast combinations with the MIDAS models and the inflation rate implied by financial derivatives, but this approach does not produce any significant gains.

Duarte (2014) discusses in detail the implications of autoregressive augmentations in MIDAS regression models and diverse ways to incorporate them. She explores the out-of-sample performance of MIDAS models with autoregressive augmentations with no restrictions, with an autoregressive augmentation with a common factor restriction, and models with autoregressive augmentations with no restrictions and a multiplicative scheme to aggregation. She then compares these models to the same models but without the autoregressive component, and to two low frequency benchmark models, a low frequency autoregression and multiple regression model. She computes forecasts for quarterly euro area GDP growth based on three different series: *i)* industrial production, *ii)* an economic sentiment indicator and *iii)* the Dow Jones Euro Stoxx index. She disregards the seasonal spikes impulse responses as the relevant impulse responses, as she argues that it is not possible to single out a particularly relevant impulse response for a mixed-frequency process since responses vary depending on when the shocks occur within the low-frequency process. Although there is no superior model among all tested, Duarte finds once again that there are sizable gains compared to the benchmarks at all horizons.

Breitung and Roling (2015) propose a "nonparametric" MIDAS model to forecast monthly inflation rates using a daily predictor. Instead of imposing any particular polynomial parameterization, the nonparametric approach consists on enforcing some degree of smoothness to the lag distribution by minimizing a penalized least squares cost function,

$$S(\theta) = (y - \boldsymbol{X}\theta)' (y - \boldsymbol{X}\theta) + \eta\theta'\boldsymbol{D}'D\theta \tag{3.4}$$

where D is a $(m-1)\times(m+1)$ matrix such that

$$D = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & -2 & 1 \end{bmatrix}, \tag{3.5}$$

and η is a pre-specified smoothing parameter. They refer to this estimator the Smoothed Least Squares estimator, and its structure closely resembles the well-known Hodrick-Prescott filter. If η is not known, they suggest solving for the η that minimizes the Akaike Information Criterion. Their target variable is the harmonized index of consumer prices for the euro area and they use a commodity price index as a high frequency regressor. They compare their model against the unconditional mean and the parametric MIDAS model (exponential Almon weights) for two different forecast horizons. They conclude that the commodity index paired with the nonparametric MIDAS results in a reasonably good one-month-ahead forecasts. Additionally, the authors conduct a Monte Carlo experiment and compare their model to four parametric MIDAS alternatives: *i)* the exponential Almon polynomial, *ii)* a hump shaped function, *iii)* a declining linear function, and *iv)* a sinusoidal function. They find that the nonparametric method performs on par with the parametric competitors.

4. DATA, EXERCISE, AND RESULTS

The out-of-sample predictive performance of an online price index will be analyzed to forecast the core inflation rate in real-time. To be more specific, this will be assessed using many different MIDAS specifications discussed in the previous sections and these estimations will be compared with benchmark single frequency naïve models and estimators. MIDAS turn out to be intuitive for this purpose 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, $\pi_t^m \approx \sum_{\tau \in t} \left(\log p_\tau^d - \log p_{\tau-1}^d \right)$

Atkeson and Ohanian (2001), Stock and Watson (2007) and Faust and Wright (2009) have shown that simple benchmarks are not easily beaten by more sophisticated models (at least in the case of the US economy), and so these could serve as a good starting point to gauge the predictive power of the daily series.

4.1 Data

The online price index is compiled by the company PriceStats in cooperation with State Street Global Markets, a leading financial research corporation. PriceStates is a spin-off company that emerged from the Billion Prices Project at MIT, founded by professors Alberto Cavallo and Roberto Rigobón. It is the first company, institution, or organization to apply a big data approach to produce real-time (daily) price indexes to track general price inflation and other related metrics. Essentially, they collect daily data of prices from online retailers by "web scraping" (i.e. recording price information contained inside specific HyperText Markup Language tags in the retailers' websites) and aggregate the data by replicating the methodology of a traditional consumer price index, as is done by National Statistics Offices with offline prices. Cavallo (2013) goes through the methodology and provides comparisons between online and offline price indexes for Argentina, Brazil, Chile, Colombia, and Venezuela. He concludes that online price indexes can track the dynamic behavior of inflation rates over time fairly well with the exception of Argentina. In fact, the construction of online price indexes was initially motivated by the desire to provide the public with an alternate measure of the inflation rate in Argentina because from the years

2007 to 2015 there were large discrepancies between the official price indexes compiled by the National Institute of Statistics and Census (INDEC) and price indexes compiled by provincial statistics offices or those compiled by private consultants. Throughout the rest of the paper, this price index will be referred to as the State Street PriceStats Index (SSPS). Data for Argentina is available since November 1, 2007 with a three-day publication lag.

A provincial price index that raised itself to prominence in recent years is the consumer price index compiled by the General Department of Statistics and Censuses of the Government of the Autonomous City of Buenos Aires, known as IPCBA. Although this index only takes into account the territory of the City of Buenos Aires (with a population close to 3 million), it should be reasonable to expect that price dynamics in the Buenos Aires Metropolitan Area (which encompasses a much larger population, close to 14 million or 1/3 of the total population of Argentina) share most of its features with the pricing structure of the City of Buenos Aires, resulting from arbitrage by reason of geographical proximity, as this should prevent large distortions, at least in nonregulated markets. A more restricted version of the index is also published, called "resto IPCBA" (rIPCBA) witch serves as a measure of core inflation. Compared to the headline version, it excludes products with strong seasonal patterns and regulated prices (e.g. public utility services) and represents 78.15% of the headline index. rIPCBA is available from July 2012 onward and is released monthly, with approximately a two-week publication lag.

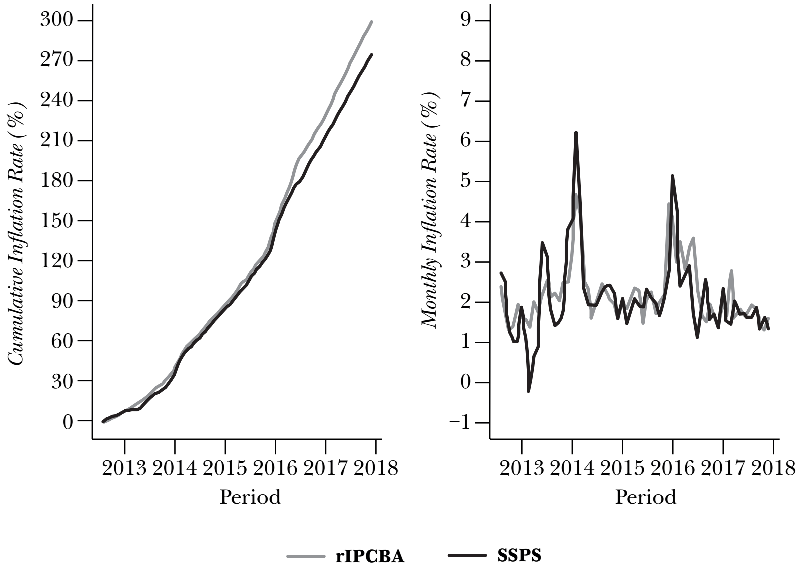
These two indexes, as well as other provincial private and public price indexes, are closely monitored by the monetary authorities as well as the general public. This is particularly true for INDEC's recently introduced National Consumer Price Index. As the name implies, this is the only index with full national coverage. However, this index so far consists of less than two years of data points and this limits the possibility of drawing any relevant inferences.

Inflation in Argentina in recent years has been high, unstable and volatile, particularly from 2012 to most of 2016 when Argentina experienced high monetization of fiscal deficits, strict capital controls and two major devaluations of the currency.[7] The average monthly

[7] The last one coinciding with the lifting of the majority of the capital controls in December 2015 and a subsequent transition to a flexible exchange rate regime and inflation targeting.

Figure 1

COMPARISON BETWEEN rIPCBA INFLATION AND SPSS INFLATION AGGREGATED TO MONTHLY FREQUENCY



inflation rate has been fluctuating around 2.2% for rIPCBA and 2.1% for the monthly aggregated SSPS series, with coefficients of variation at 35% and 49% respectively. This should pose a significant challenge for economists' ability to formulate accurate forecasts. Figure 1 illustrates the comparison between these two indexes and provides a quick glimpse at the potential predictive power of the high-frequency index. Overall and for the scope of this work, rIPCBA is available from July 2012 to December 2017 (66 data points) while SSPS ranges from November 1, 2007 to December 31, 2017 (3,714 data points).

4.2 Forecasting Exercise

The MIDAS specifications tested were the MIDAS-DL, the unrestricted autoregressive MIDAS-DL (MIDAS-ADL), and the autoregressive MIDAS-DL with the common factor restriction (MIDAS-ADL-CF). All MIDAS specifications were evaluated with several high frequency regressors equal to $m \times L_X$,[8] with $L_X \in \{1,2,3\}$, and forecasts were computed for horizons $h \in \{1,2,3\}$ over a 36-observation evaluation sample, spanning from January 2015 to December 2017, and an 18 observation subsample from July 2016 to December 2017 (a period with a more stable inflation rate), using recursive (expanding) windows. MIDAS-ADL-CF models included quadratic and cubic variations of the standard Almon polynomial and the exponential version, as well as the Beta probability density function. MIDAS-ADL models further added flat aggregation (equal weights); and finally, MIDAS-DL models added the nonparametric (NP) model described in Section 3. Forecast combinations of the various MIDAS models with equal weights were also considered. In addition, all these models were compared to two benchmarks: *i)* the low-frequency unconditional mean and *ii)* a low-frequency first order autoregression.[9]

In a first stage, the models were estimated with a balanced dataset. In other words, 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. In total, two sets of RMSFE were computed, one corresponding to the large sample and the other to a reduced subsample. For all forecast horizons, d was set to $d=-1$.

A second stage involved estimating intra-period forecasts for the best selected L_X for each forecast horizon based on the results from the large sample of the first stage and briefly analyzing the stability of the forecasts as more recent information is incorporated in the models. When intra-period forecasts were 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 the frequency. Forecasts from the autoregression and the uncondditional mean remained the same throughout the month.

To account for the fact that SSPS is an irregularly spaced series, the frequency was assumed fixed at $m=28$, and so days 29, 30 and 31 of each month are discarded. Daily inflation rates were first computed with the full dataset and then the observations beyond day 28 of each month were discarded.

8 First order MIDAS-ADL-CF models include $m \times \left[L_X + \min\left(L_X, h\right)\right]$. high-frequency regressors since the common factor restriction increases the number of variates depending on the forecast horizon and the number of high frequency lags.

9 A detailed list of the models can be found in Appendix A.

Estimation was conducted in R with the midasr package developed by Ghysels *et al.* (2016) while optimization was performed with three routines included in optimx[10] for nonlinear models or with the lm function from the stats package for linear ones. Models that require optimx were solved simultaneously with three optimization routines (ucminf, nlminb and Nelder-Mead) for each model, forecast horizon h, number of high frequency regressors L_X, and out-of-sample period. Only the best solution was kept. The algorithm was initialized taking the hypothesis of equal weights and a null impact parameter as starting conditions. This strategy delivered reasonable results empirically and serves as a check on whether the high-frequency regressors are actually relevant.

4.3 Empirical Results

Tables 1 and 2 summarize the main results of the first stage. In general, for $h=1$ (nowcasts), larger values of L_X produce better results while this tends to reverse when forecasting further into the future, i.e. $h=3$. For $h=2$, the results are ambiguous and indicate that $L_X=2$ or $L_X=3$ perform best. All three classes of MIDAS models exhibit similar performance irrespective of the inclusion of the autoregressive term or how it is incorporated. For all h, most MIDAS models for at least some L_X are able to produce a small gain at around 10% when compared to the autoregression and a larger 25% against the unconditional mean.[11] The smaller sample greatly amplifies these results. Note that for each h, there is a flat aggregation model that performed very well and, at times, even better than standard MIDAS models, but overall, there is not a single MIDAS model that systematically outperforms the rest. The forecast combination tested does not seem to improve over any particular MIDAS model.

Figures (2)-(4) condense the main findings of the second stage. Forecasts for $h=1$ display a clear trend towards better accuracy as the high frequency variable reaches a full low frequency period. 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

---

[10] A comprehensive description about this package can be found in Nash and Varadhan (2011).

[11] Tables with RMSFE ratios are presented in Appendix B.

by up to 15% at most for some days. The improved performance, when evaluated in the subsample, suggests that it is even possible to obtain even better results as the inflation rate stabilizes. Similar behavior, although less evident, is observed for forecasts for period $h = 3$ in the case of MIDAS-DL models. Forecasts for horizon $h = 2$ display a rather erratic pattern excepting the flat aggregation MIDAS-DL and MIDAS-ADL models.

Figure 5 zooms in on the evolution of all intra-period forecasts for selected models, either $h = 1$, $h = 2$ or $h = 3$. Despite the 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. Additionally, note that forecasting further into the future yields a dynamic closer to the unconditional mean of the whole process. In the future, these results could be used as a training sample from which to compute inverse mean square error weights and perform forecast combinations, which could prove to be effective in mitigating intra-period forecast volatility.

Although the results look promising, they should be interpreted with caution. The predictive ability of the models was tested with the methodology by Giacomini and White (2006)[12] and both the unconditional and the conditional versions of the test were examined. The MIDAS models were evaluated against the two naïve benchmarks, modeling the difference in forecast accuracy as a constant (unconditional) and also as a first order autoregression (conditional). The results do not indicate that the difference in forecast accuracy is significant (at 0.05) for most MIDAS models. However, since the "large" out-of-sample evaluation set actually constitutes a small sample by literature standards, the result of the tests cannot be taken as final. As more observations become available, the tests could be updated with a larger sample to arrive at a more robust conclusion.

[12] This is similar to the standard test by Diebold and Mariano (1995). The key difference lies in that the estimation sample size is kept fixed instead of ever expanding, as this allows to better incorporate estimation uncertainty and to compare nested models.

Table 1

OUT-OF-SAMPLE PREDICTIVE PERFORMANCE, RMSFE

	$h=1$			$h=2$			$h=3$		
	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$
Almon (q = 2)									
MIDAS-DL	0.630	0.627	**0.564**	0.790	*0.717*	***0.712***	*0.740*	***0.721***	*0.751*
MIDAS-ADL	*0.578*	*0.589*	**0.564**	*0.775*	0.729	**0.724**	0.760	**0.741**	0.765
MIDAS-ADL-CF	0.592	0.625	***0.561***	0.785	0.722	**0.719**	**0.748**	0.733	0.768
Almon (q = 3)									
MIDAS-DL	0.660	0.623	***0.571***	0.819	*0.731*	***0.720***	*0.755*	***0.724***	*0.757*
MIDAS-ADL	*0.609*	*0.609*	**0.574**	*0.805*	0.745	**0.731**	0.770	**0.739**	0.770
MIDAS-ADL-CF	0.617	0.636	**0.576**	0.827	0.741	***0.719***	0.762	**0.734**	0.777
Exp. Almon (q = 2)									
MIDAS-DL	0.705	0.646	***0.566***	0.816	0.755	***0.749***	***0.803***	0.863	0.837
MIDAS-ADL	***0.627***	0.633	0.632	0.775	0.768	**0.766**	0.839	0.840	***0.835***
MIDAS-ADL-CF	0.639	*0.629*	***0.557***	*0.765*	***0.745***	0.885	0.875	***0.831***	0.860
Exp. Almon (q = 3)									
MIDAS-DL	0.731	**0.648**	0.661	0.834	*0.826*	***0.742***	*0.809*	*0.785*	***0.778***
MIDAS-ADL	***0.628***	*0.633*	0.645	0.822	0.834	**0.814**	0.827	**0.792**	0.807
MIDAS-ADL-CF	0.663	0.661	***0.563***	***0.812***	0.863	0.866	0.824	**0.800**	0.824

Beta									
MIDAS-DL	0.668	0.624	**0.574**	0.768	***0.694***	*0.701*	0.747	***0.730***	0.746
MIDAS-ADL	*0.571*	*0.622*	**0.568**	*0.728*	**0.707**	0.716	***0.732***	0.750	0.748
MIDAS-ADL-CF	0.614	*0.619*	***0.558***	0.739	**0.697**	0.704	0.740	**0.736**	*0.737*
Flat									
MIDAS-DL	1.158	0.617	***0.568***	0.745	***0.673***	*0.690*	***0.713***	*0.736*	*0.745*
MIDAS-ADL	*0.944*	*0.592*	***0.568***	*0.733*	**0.694**	0.746	**0.729**	0.766	0.777
Nonparametric									
MIDAS-DL	0.623	0.629	**0.567**	0.782	**0.717**	**0.717**	**0.718**	0.721	0.755
EW Forecast Combination									
MIDAS-DL	0.657	0.621	**0.569**	0.780	*0.719*	***0.710***	***0.738***	*0.741*	*0.750*
MIDAS-ADL	*0.585*	*0.595*	**0.570**	*0.765*	0.729	**0.728**	0.763	**0.760**	0.764
MIDAS-ADL-CF	0.609	0.627	***0.563***	0.768	**0.734**	0.754	0.770	**0.756**	0.770
Autoregression									
$p = 1$	0.619	0.619	0.619	0.757	0.757	0.757	0.757	0.757	0.757
Unconditional Mean									
$\overline{y}$	0.790	0.790	0.790	0.800	0.800	0.800	0.806	0.806	0.806

Notes: The evaluation sample comprises 36 data points, from January 2015 to December 2017. Characters in **bold** indicate the best number of variables, L_X, for each model and forecast horizon, h. Characters in *italics* indicate the best model for each number of variables, L_X, and forecast horizon, h.

Table 2

OUT-OF-SAMPLE PREDICTIVE PERFORMANCE, RMSFE

	$h=1$			$h=2$			$h=3$		
	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$
Almon ($q=2$)									
MIDAS-DL	0.555	0.521	**0.427**	0.646	*0.541*	***0.538***	*0.568*	*0.548*	***0.547***
MIDAS-ADL	*0.460*	*0.482*	**0.430**	*0.608*	0.547	**0.544**	0.579	**0.562**	0.569
MIDAS-ADL-CF	0.508	0.528	***0.422***	0.625	0.548	**0.543**	0.586	0.574	**0.573**
Almon ($q=3$)									
MIDAS-DL	0.564	0.560	***0.430***	0.683	*0.572*	***0.546***	*0.580*	***0.547***	*0.559*
MIDAS-ADL	*0.465*	*0.523*	**0.435**	*0.642*	0.581	**0.551**	0.578	**0.561**	0.584
MIDAS-ADL-CF	0.533	0.569	**0.433**	0.663	0.581	**0.549**	0.596	**0.581**	0.601
Exp. Almon ($q=2$)									
MIDAS-DL	0.568	0.523	**0.432**	0.656	0.540	**0.539**	0.574	***0.561***	***0.561***
MIDAS-ADL	*0.470*	***0.453***	**0.453**	*0.600*	0.538	***0.535***	***0.568***	0.572	0.578
MIDAS-ADL-CF	0.581	0.531	***0.425***	0.620	***0.533***	0.821	0.669	0.585	**0.575**
Exp. Almon (q $=3$)									
MIDAS-DL	0.620	0.549	**0.430**	*0.566*	*0.551*	***0.533***	***0.555***	*0.559*	*0.559*
MIDAS-ADL	*0.469*	***0.460***	0.456	0.640	0.651	**0.540**	0.565	**0.561**	0.565
MIDAS-ADL-CF	0.613	0.545	***0.424***	0.720	0.657	**0.634**	**0.573**	0.579	0.579

Beta									
MIDAS-DL	0.598	0.491	**0.444**	0.592	***0.506***	*0.523*	0.587	***0.570***	0.583
MIDAS-ADL	*0.489*	***0.437***	*0.441*	*0.550*	**0.513**	0.529	*0.578*	0.585	***0.569***
MIDAS-ADL-CF	0.548	0.492	***0.432***	0.568	**0.511**	0.525	0.580	***0.570***	0.571
Flat									
MIDAS-DL	0.542	0.515	**0.417**	0.610	***0.501***	*0.526*	***0.539***	*0.577*	*0.594*
MIDAS-ADL	*0.540*	*0.479*	***0.405***	*0.577*	**0.515**	0.547	**0.544**	0.580	0.604
Nonparametric									
MIDAS-DL	0.551	0.539	**0.432**	0.633	0.543	**0.540**	0.554	**0.548**	0.557
EW Forecast Combination									
MIDAS-DL	0.488	0.522	**0.429**	0.621	***0.533***	***0.533***	*0.562*	*0.555*	***0.561***
MIDAS-ADL	***0.417***	*0.462*	*0.426*	*0.597*	0.544	**0.534**	**0.565**	0.566	0.573
MIDAS-ADL-CF	0.541	0.527	***0.426***	0.624	**0.551**	0.592	0.595	**0.573**	0.575
Autoregression									
$p=1$	0.552	0.552	0.552	0.675	0.675	0.675	0.692	0.692	0.692
Unconditional Mean									
$\overline{y}$	0.697	0.697	0.697	0.691	0.691	0.691	0.685	0.685	0.685

Notes: The evaluation sample comprises 18 data points, from July 2016 to December 2017. Characters in **bold** indicate the best number of variables, L_X, for each model and forecast horizon, h. Characters in *italics* indicate the best model for each number of variables, L_X, and forecast horizon, h.

Figure 2

EVOLUTION OF THE RMSFE FOR HORIZON $h = 1$ WITHIN A MONTH FOR SELECTED MODELS WITH $L_X = 3$

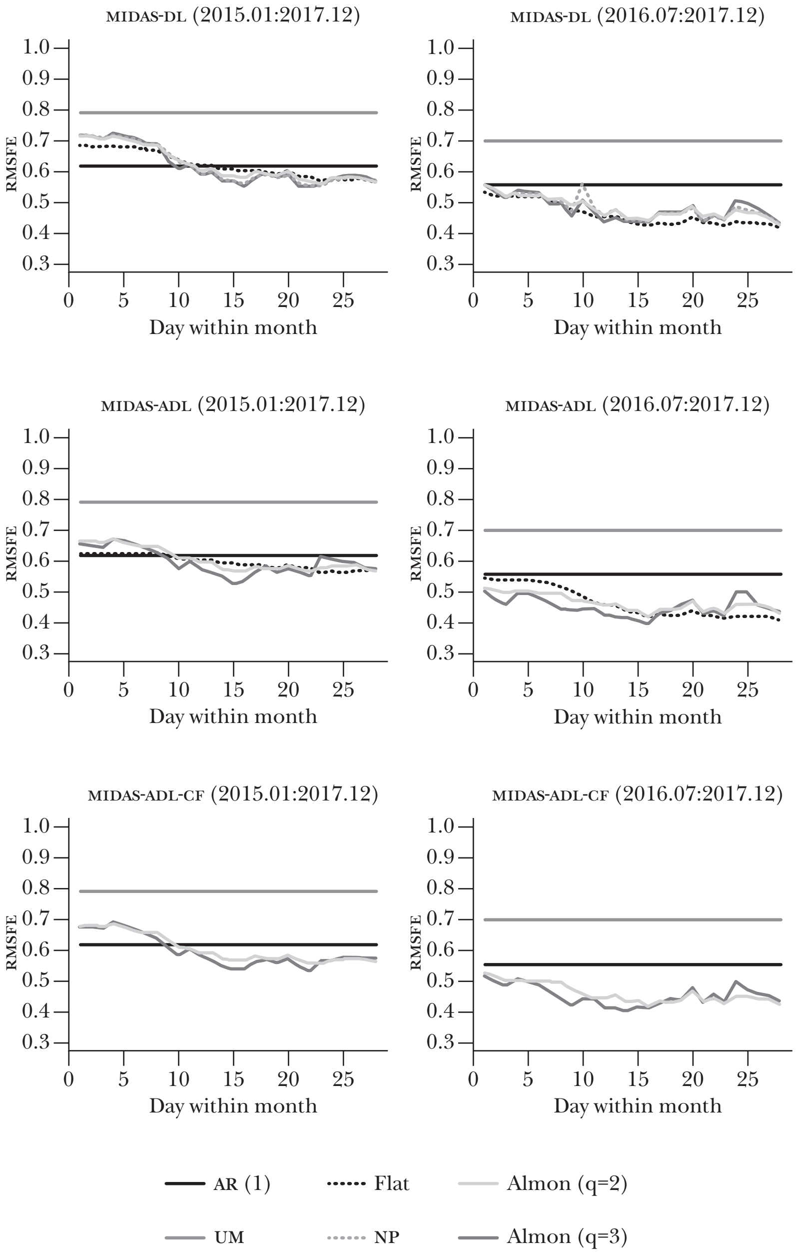


Figure 3

EVOLUTION OF THE RMSFE FOR HORIZON $h=2$ WITHIN A MONTH FOR SELECTED MODELS WITH $L_X = 3$

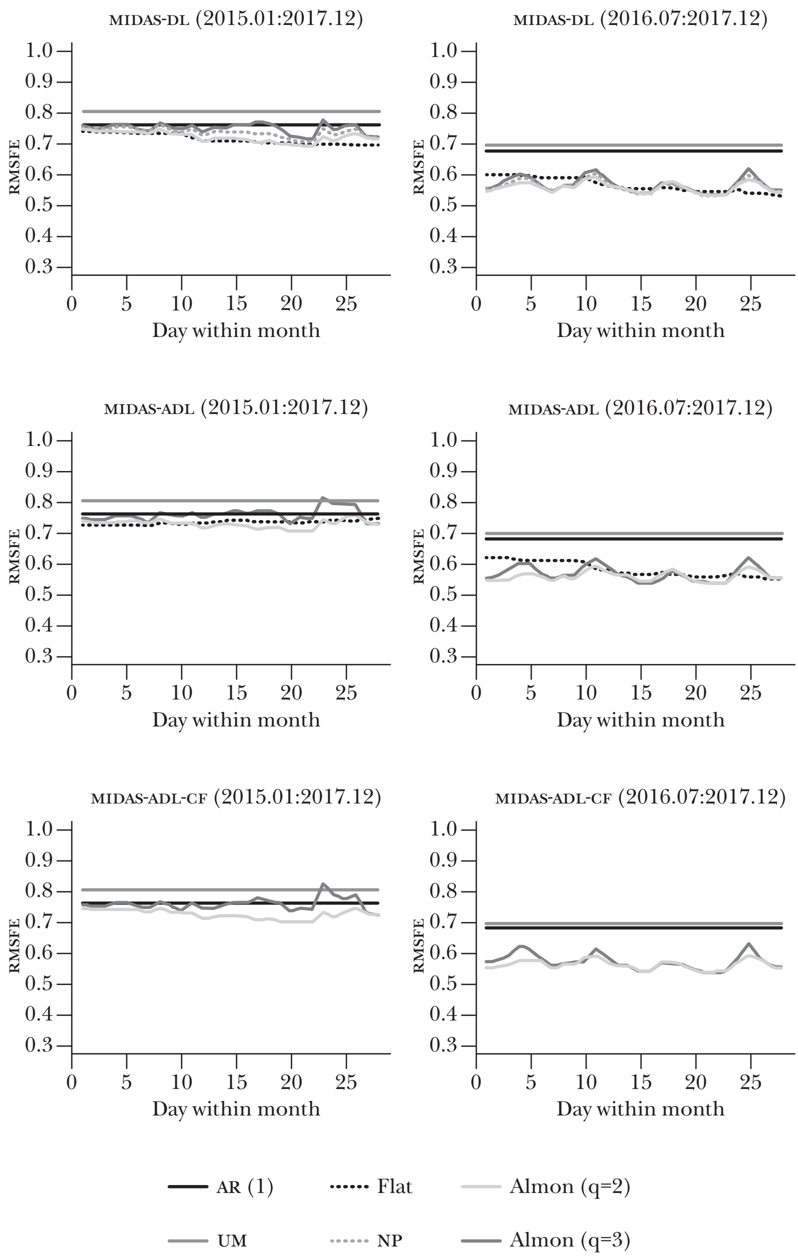


Figure 4

EVOLUTION OF THE RMSFE FOR HORIZON $h = 3$ WITHIN A MONTH FOR SELECTED MODELS WITH $L_X = 2$

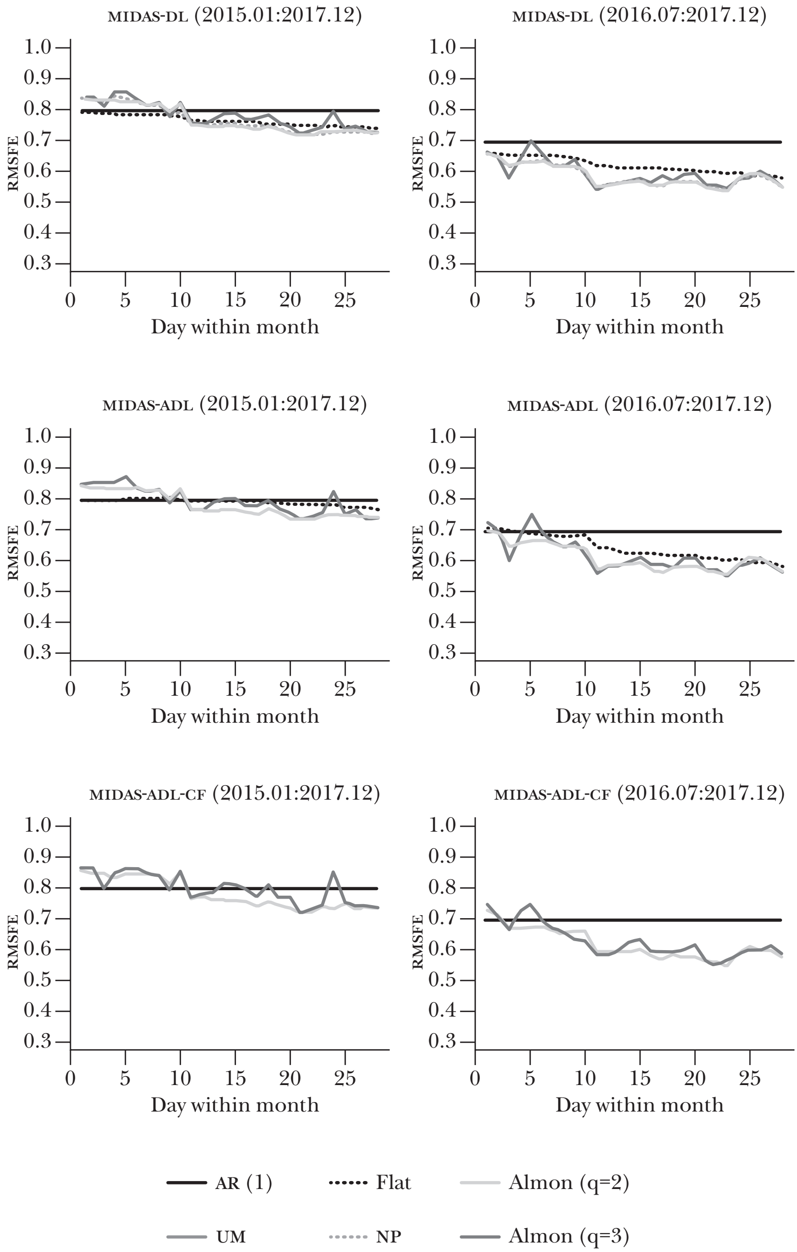
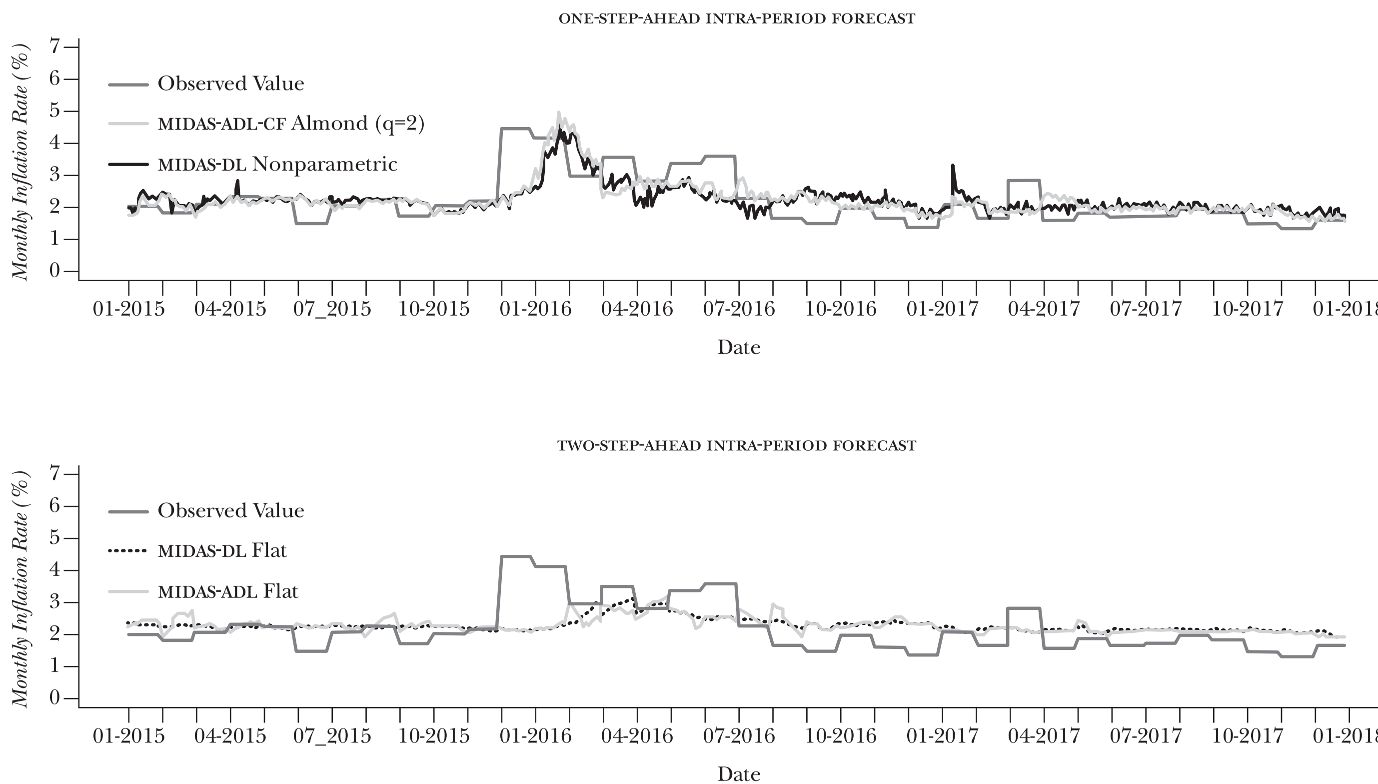
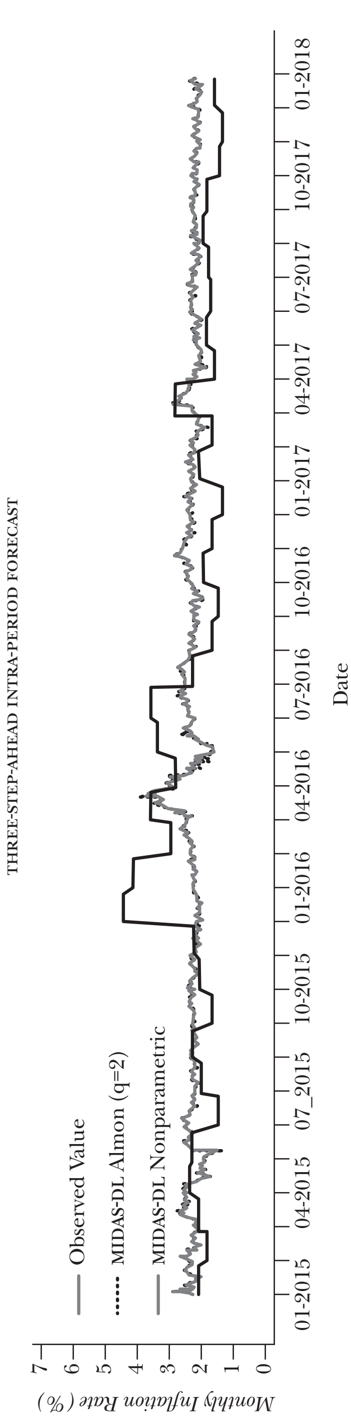


Figure 5

EVOLUTION OF INTRA-PERIOD FORECASTS FOR SELECTED MODELS AND FORECAST HORIZONS




THREE-STEP-AHEAD INTRA-PERIOD FORECAST
Observed Value
MIDAS-DL Almon (q=2)
MIDAS-DL Nonparametric
Monthly Inflation Rate (%)
0
1
2
3
4
5
6
7
01-2015
04-2015
07_2015
10-2015
01-2016
04-2016
07-2016
10-2016
01-2017
04-2017
07-2017
10-2017
01-2018
Date

5. CONCLUSION

For some particular MIDAS specifications, there is a slight improvement compared to the low-frequency benchmark autoregression and the unconditional mean. In principle, this would imply that high-frequency online price indices have a good potential to forecast future behavior of consumer inflation for immediate horizons in Argentina, but these results are still not robust. This could serve as a useful complementary tool to assess the out-of-sample performance of perhaps more sophisticated models. Future research could focus on building an alternative variable such as a daily financial factor as suggested by Monteforte and Moretti (2013) or comparing with measures of market expectations in order to further validate the findings of this paper.

ANNEX

Appendix A: MIDAS Specifications

The full set of specifications of the models is detailed below. All models were estimated with $L_X \in \{1,2,3\}$, $h \in \{1,2,3\}$ and d as explained in subsection 4.2. The subscript (d,h) on parameter estimates denoting dependence on d and h has been suppressed for simplicity.

MIDAS-DL:

A.1

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\lambda} \sum_{j=0}^{m \times L_X - 1} \left(\sum_{s=1}^{2} \gamma_s j^s \right) \pi_{T-d-j/m}^{SSPS}$$

A.2

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\lambda} \sum_{j=0}^{m \times L_X - 1} \left(\sum_{s=1}^{3} \hat{\gamma}_s j^s \right) \pi_{T-d-j/m}^{SSPS}$$

A.3

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\lambda} \sum_{j=0}^{m \times L_X - 1} \left(\frac{e^{\sum_{s=1}^{2} \gamma_s j^s}}{\sum_{j=0}^{m \times L_X - 1} e^{\sum_{s=1}^{2} \gamma_s j^s}} \right) \pi_{T-d-j/m}^{SSPS}$$

A.4

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\lambda} \sum_{j=0}^{m \times L_X - 1} \left(\frac{e^{\sum_{s=1}^{3} \gamma_s j^s}}{\sum_{j=0}^{m \times L_X - 1} e^{\sum_{s=1}^{3} \gamma_s j^s}} \right) \pi_{T-d-j/m}^{SSPS}$$

A.5

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\lambda} \sum_{j=0}^{m \times L_x - 1} \left(\frac{z_j^{\gamma_1 - 1} \left(1 - z_j\right)^{\gamma_2 - 1}}{\sum_{j=0}^{m \times L_x - 1} z_j^{\gamma_1 - 1} \left(1 - z_j\right)^{\gamma_2 - 1}} \right) \pi_{T-d-j/m}^{SSPS}$$

A.6

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\lambda} \sum_{j=0}^{m \times L_x - 1} \pi_{T-d-j/m}^{SSPS}$$

A.7

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \sum_{j=0}^{m \times L_x - 1} \hat{\theta}_j^{NP} \pi_{T-d-j/m}^{SSPS}$$

MIDAS-ADL:

A.8

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \phi \hat{\pi}_T^{rIPCBA} + \hat{\lambda} \sum_{j=0}^{m \times L_X - 1} \left(\sum_{s=1}^{2} \hat{\gamma}_s j^s \right) \pi_{T-d-j/m}^{SSPS}$$

A.9

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi} \pi_T^{rIPCBA} + \hat{\lambda} \sum_{j=0}^{m \times L_X - 1} \left(\sum_{s=1}^{3} \hat{\gamma}_s j^s \right) \pi_{T-d-j/m}^{SSPS}$$

A.10

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi} \pi_T^{rIPCBA} + \hat{\lambda} \sum_{j=0}^{m \times L_x - 1} \left(\frac{e^{\sum_{s=1}^{2} \hat{\gamma}_s j^s}}{\sum_{j=0}^{m \times L_x - 1} e^{\sum_{s=1}^{2} \hat{\gamma}_s j^s}} \right) \pi_{T-d-j/m}^{SSPS}$$

A.11

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi} \pi_T^{rIPCBA} + \hat{\lambda} \sum_{j=0}^{m \times L_x - 1} \left(\frac{e^{\sum_{s=1}^{3} \hat{\gamma}_s j^s}}{\sum_{j=0}^{m \times L_x - 1} e^{\sum_{s=1}^{3} \hat{\gamma}_s j^s}} \right) \pi_{T-d-j/m}^{SSPS}$$

A.12

$$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi} \pi_T^{rIPCBA} + \hat{\lambda} \sum_{j=0}^{m \times L_x - 1} \left(\frac{z_j^{\hat{\gamma}_1 - 1} \left(1 - z_j\right)^{\hat{\gamma}_2 - 1}}{\sum_{j=0}^{m \times L_x - 1} z_j^{\hat{\gamma}_1 - 1} \left(1 - z_j\right)^{\hat{\gamma}_2 - 1}} \right) \pi_{T-d-j/m}^{SSPS}$$

A.13 $$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi}\pi_T^{rIPCBA} + \hat{\lambda}\sum_{j=0}^{m\times L_x-1}\pi_{T-d-j/m}^{SSPS}$$

MIDAS-ADL-CF:

A.14 $$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi}\pi_T^{rIPCBA} + \hat{\lambda}\left(1+\hat{\phi}L^h\right)\sum_{j=0}^{m\times L_X-1}\left(\sum_{s=1}^{2}\hat{\gamma}_s j^s\right)\pi_{T-d-j/m}^{SSPS}$$

A.15 $$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi}\pi_T^{rIPCBA} + \hat{\lambda}\left(1+\hat{\phi}L^h\right)\sum_{j=0}^{m\times L_X-1}\left(\sum_{s=1}^{3}\hat{\gamma}_s j^s\right)\pi_{T-d-j/m}^{SSPS}$$

A.16 $$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi}\pi_T^{rIPCBA} + \hat{\lambda}\left(1+\hat{\phi}L^h\right)\sum_{j=0}^{m\times L_x-1}\left(\frac{e^{\sum_{s=1}^{2}\hat{\gamma}_s j^s}}{\sum_{j=0}^{m\times L_x-1}e^{\sum_{s=1}^{2}\hat{\gamma}_s j^s}}\right)\pi_{T-d-j/m}^{SSPS}$$

A.17 $$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi}\pi_T^{rIPCBA} + \hat{\lambda}\left(1+\hat{\phi}L^h\right)\sum_{j=0}^{m\times L_x-1}\left(\frac{e^{\sum_{s=1}^{3}\hat{\gamma}_s j^s}}{\sum_{j=0}^{m\times L_x-1}e^{\sum_{s=1}^{3}\hat{\gamma}_s j^s}}\right)\pi_{T-d-j/m}^{SSPS}$$

A.18 $$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi}\pi_T^{rIPCBA} + \hat{\lambda}\left(1+\hat{\phi}L^h\right)\sum_{j=0}^{m\times L_x-1}\left(\frac{z_j^{\hat{\gamma}_1-1}\left(1-z_j\right)^{\hat{\gamma}_2-1}}{\sum_{j=0}^{m\times L_x-1}z_j^{\hat{\gamma}_1-1}\left(1-z_j\right)^{\hat{\gamma}_2-1}}\right)\pi_{T-d-j/m}^{SSPS}$$

Other:

A.19 $$\hat{\pi}_{T+h}^{rIPCBA} = \hat{\alpha} + \hat{\phi}\pi_T^{rIPCBA}$$

A.20 $$\hat{\pi}_{T+h}^{rIPCBA} = \left(1/T\right)\sum_{t=1}^{T}\pi_t^{rIPCBA}$$

Appendix B: Additional Tables

Table B.1

OUT-OF-SAMPLE PREDICTIVE PERFORMANCE, RATIO TO RMSFE OF AUTOREGRESSION ×100

	$h=1$			$h=2$			$h=3$		
	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$
Almon ($q=2$)									
MIDAS-DL	101.8	101.4	**91.2**	104.2	*94.7*	***93.9***	*93.1*	***90.8***	*94.5*
MIDAS-ADL	*93.4*	*95.1*	**91.1**	*102.3*	96.2	**95.6**	95.6	**93.2**	96.2
MIDAS-ADL-CF	95.7	101.0	***90.6***	103.6	95.3	**94.9**	94.1	**92.2**	96.6
Almon ($q=3$)									
MIDAS-DL	106.6	100.7	***92.2***	108.1	*96.5*	**95.0**	*95.0*	***91.2***	*95.2*
MIDAS-ADL	*98.3*	*98.4*	**92.7**	*106.3*	98.4	**96.5**	96.9	**93.0**	96.9
MIDAS-ADL-CF	99.8	102.7	**93.0**	109.2	97.8	***94.9***	95.8	**92.3**	97.7
Exp. Almon ($q=2$)									
MIDAS-DL	113.8	104.4	**91.5**	107.7	99.7	***98.8***	***101.0***	108.5	105.3
MIDAS-ADL	***101.3***	102.2	102.1	102.4	101.4	**101.1**	105.5	105.7	***105.1***
MIDAS-ADL-CF	103.2	*101.6*	***90.0***	*100.9*	***98.4***	116.8	110.1	***104.6***	108.2
Exp. Almon (q $=3$)									
MIDAS-DL	118.2	**104.7**	106.8	110.1	*109.1*	***97.9***	*101.8*	*98.8*	***97.9***

MIDAS-ADL	101.4	*102.3*	104.2	108.5	110.1	**107.4**	104.0	99.7	**101.5**
MIDAS-ADL-CF	107.1	106.8	***90.9***	*107.3*	114.0	**114.4**	103.7	**100.6**	103.7
Beta									
MIDAS-DL	108.0	100.8	**92.7**	101.4	***91.6***	*92.6*	*94.0*	***91.8***	*93.8*
MIDAS-ADL	*92.3*	100.5	**91.7**	*96.1*	**93.4**	94.5	**92.1**	94.4	94.2
MIDAS-ADL-CF	99.2	*100.1*	***90.2***	97.6	**92.0**	92.9	*93.1*	**92.6**	92.7
Flat									
MIDAS-DL	187.1	99.8	***91.8***	98.3	***88.8***	*91.1*	***89.7***	*92.6*	*93.8*
MIDAS-ADL	*152.6*	*95.6*	***91.8***	*96.7*	**91.6**	98.5	**91.7**	96.4	97.8
Nonparametric									
MIDAS-DL	100.6	101.6	**91.7**	103.3	**94.6**	**94.6**	**90.4**	90.7	95.0
EW Forecast Combination									
MIDAS-DL	106.2	100.4	**92.0**	102.9	*95.0*	***93.7***	***92.9***	*93.3*	*94.3*
MIDAS-ADL	*94.5*	*96.1*	**92.2**	*100.9*	96.3	**96.1**	96.0	**95.6**	96.2
MIDAS-ADL-CF	98.4	101.4	***90.9***	101.4	**96.9**	99.5	96.9	**95.1**	96.9
Unconditional Mean									
$\overline{y}$	127.7	127.7	127.7	105.6	105.6	105.6	101.4	101.4	101.4

Notes: The evaluation sample comprises 36 data points, from January 2015 to December 2017. Characters in **bold** indicate the best number of variables, L_X, for each model and forecast horizon, h. Characters in *italics* indicate the best model for each number of variables, L_X, and forecast horizon, h.

Table B.2

OUT-OF-SAMPLE PREDICTIVE PERFORMANCE, RATIO TO RMSFE OF AUTOREGRESSION×100

	$h=1$			$h=2$			$h=3$		
	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$
Almon ($q=2$)									
MIDAS-DL	100.6	94.4	***77.4***	95.7	*80.1*	***79.6***	*82.2*	*79.2*	***79.0***
MIDAS-ADL	*83.4*	*87.4*	**77.9**	*90.1*	81.0	**80.6**	83.8	**81.2**	82.4
MIDAS-ADL-CF	92.0	95.7	**76.5**	92.6	81.2	**80.5**	84.8	82.9	**82.8**
Almon ($q=3$)									
MIDAS-DL	102.3	101.5	***78.0***	101.2	*84.8*	***80.8***	83.9	***79.2***	*80.8*
MIDAS-ADL	84.2	*94.7*	**78.9**	*95.1*	86.0	**81.6**	*83.6*	**81.1**	84.4
MIDAS-ADL-CF	96.6	103.2	**78.5**	98.3	86.0	**81.3**	86.1	**84.1**	86.9
Exp. Almon ($q=2$)									
MIDAS-DL	103.0	94.8	***78.3***	97.1	80.0	**79.9**	83.1	***81.2***	***81.2***
MIDAS-ADL	*85.3*	***82.1***	**82.1**	*88.9*	79.7	***79.2***	***82.1***	82.8	83.5
MIDAS-ADL-CF	105.4	96.2	**77.1**	91.9	***79.0***	121.6	96.8	84.6	**83.2**
Exp. Almon (q $=3$)									
MIDAS-DL	112.4	99.5	**77.9**	*83.8*	*81.6*	***79.0***	***80.3***	80.8	80.8

MIDAS-ADL	*85.0*	*83.4*	**82.7**	94.8	96.4	**80.0**	81.7	**81.2**	81.7
MIDAS-ADL-CF	111.1	98.8	***77.0***	106.6	97.3	**94.0**	**82.9**	83.7	83.7
Beta									
MIDAS-DL	108.5	89.1	**80.5**	87.7	***74.9***	*77.5*	84.9	***82.4***	84.3
MIDAS-ADL	*88.6*	***79.2***	80.0	*81.5*	**75.9**	78.4	*83.5*	84.6	***82.3***
MIDAS-ADL-CF	99.4	89.3	***78.3***	84.1	**75.7**	77.7	83.8	***82.4***	82.5
Flat									
MIDAS-DL	98.3	93.4	**75.6**	90.3	***74.2***	*77.9*	***77.9***	*83.4*	*85.9*
MIDAS-ADL	*98.0*	*86.8*	***73.4***	*85.5*	**76.3**	81.0	**78.7**	83.8	87.3
Nonparametric									
MIDAS-DL	100.0	97.8	**78.3**	93.8	**80.4**	80.0	80.1	**79.2**	80.6
EW Forecast Combination									
MIDAS-DL	88.4	94.7	**77.7**	91.9	***78.9***	***78.9***	*81.2*	***80.2***	*81.2*
MIDAS-ADL	***75.5***	*83.9*	*77.3*	*88.4*	80.6	**79.1**	**81.8**	81.9	82.8
MIDAS-ADL-CF	98.1	95.5	***77.3***	92.4	**81.6**	87.7	86.0	**82.9**	83.1
Unconditional Mean									
$\bar{y}$	126.3	126.3	126.3	102.4	102.4	102.4	99.0	99.0	99.0

Notes: The evaluation sample comprises 18 data points, from July 2016 to December 2017. Characters in **bold** indicate the best number of variables, L_X, for each model and forecast horizon, h. Characters in *italics* indicate the best model for each number of variables, L_X, and forecast horizon, h.

Table B.3

OUT-OF-SAMPLE PREDICTIVE PERFORMANCE, RATIO TO RMSFE OF UNCONDITIONAL MEAN×100

	$h=1$			$h=2$			$h=3$		
	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$
Almon ($q=2$)									
MIDAS-DL	79.7	79.4	**71.4**	98.7	*89.6*	**88.9**	*91.8*	***89.5***	*93.2*
MIDAS-ADL	*73.1*	*74.5*	**71.4**	*96.9*	91.1	**90.5**	94.2	**91.9**	94.9
MIDAS-ADL-CF	75.0	79.1	***71.0***	98.1	90.3	***89.8***	92.8	**90.9**	95.3
Almon ($q=3$)									
MIDAS-DL	83.5	78.9	***72.2***	102.4	*91.3*	**89.9**	*93.7*	***89.9***	*93.9*
MIDAS-ADL	*77.0*	*77.1*	**72.6**	*100.6*	93.2	**91.3**	95.5	**91.7**	95.6
MIDAS-ADL-CF	78.1	80.4	**72.9**	103.4	92.6	***89.8***	94.5	**91.0**	96.3
Exp. Almon ($q=2$)									
MIDAS-DL	89.1	81.7	**71.7**	101.9	94.4	***93.6***	***99.6***	107.0	103.8
MIDAS-ADL	***79.4***	80.1	79.9	96.9	96.0	**95.7**	104.0	104.2	***103.6***
MIDAS-ADL-CF	80.8	*79.5*	***70.5***	*95.6*	***93.1***	110.6	108.5	***103.1***	106.7
Exp. Almon (q = 3)									
MIDAS-DL	92.5	**82.0**	83.6	104.3	*103.3*	***92.7***	*100.3*	*97.4*	***96.6***
MIDAS-ADL	***79.4***	*80.1*	81.6	102.7	**104.2**	101.7	102.6	**98.3**	100.1

MIDAS-ADL-CF	83.8	83.7	***71.2***	***101.6***	107.9	108.3	102.2	**99.2**	102.2
Beta									
MIDAS-DL	84.5	79.0	**72.6**	96.0	***86.7***	*87.6*	92.6	***90.5***	92.5
MIDAS-ADL	*72.3*	78.7	**71.8**	*91.0*	**88.4**	89.4	***90.8***	93.0	92.8
MIDAS-ADL-CF	77.6	*78.4*	***70.6***	92.4	**87.1**	88.0	91.8	**91.3**	***91.4***
Flat									
MIDAS-DL	146.5	78.1	**71.9**	93.1	***84.1***	*86.2*	***88.4***	*91.3*	*92.5*
MIDAS-ADL	*119.5*	*74.9*	***71.9***	*91.6*	**86.7**	93.3	**90.5**	95.0	96.4
Nonparametric									
MIDAS-DL	78.8	79.6	**71.8**	97.8	**89.6**	**89.6**	**89.1**	89.4	93.7
EW Forecast Combination									
MIDAS-DL	83.2	78.6	**72.1**	97.5	*89.9*	***88.7***	***91.6***	*92.0*	*93.0*
MIDAS-ADL	*74.0*	*75.3*	**72.2**	*95.6*	91.2	**90.9**	94.6	**94.3**	94.8
MIDAS-ADL-CF	77.1	79.4	***71.2***	96.0	**91.7**	94.2	95.5	**93.7**	95.5
Autoregression									
$p=1$	78.3	78.3	78.3	94.7	94.7	94.7	98.6	98.6	98.6

Notes: The evaluation sample comprises 36 data points, from January 2015 to December 2017. Characters in **bold** indicate the best number of variables, L_X, for each model and forecast horizon, h. Characters in *italics* indicate the best model for each number of variables, L_X, and forecast horizon, h.

Table B.4

OUT-OF-SAMPLE PREDICTIVE PERFORMANCE, RATIO TO RMSFE OF UNCONDITIONAL MEAN×100

	$h=1$			$h=2$			$h=3$		
	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$	$L_X=1$	$L_X=2$	$L_X=3$
Almon ($q=2$)									
MIDAS-DL	79.6	74.8	**61.3**	93.5	*78.2*	***77.7***	*83.0*	*80.0*	***79.8***
MIDAS-ADL	*66.1*	*69.2*	**61.7**	*88.0*	79.1	**78.7**	84.6	**82.0**	83.2
MIDAS-ADL-CF	72.9	75.8	***60.6***	90.4	79.2	**78.5**	85.6	83.8	**83.7**
Almon ($q=3$)									
MIDAS-DL	81.0	80.3	***61.8***	98.8	*82.8*	***78.9***	84.7	***79.9***	*81.6*
MIDAS-ADL	*66.7*	*75.0*	**62.4**	*92.9*	*92.9*	84.0	**79.7**	*84.5*	**81.9**
MIDAS-ADL-CF	76.5	81.7	**62.2**	95.9	84.0	**79.4**	87.0	**84.9**	87.7
Exp. Almon ($q=2$)									
MIDAS-DL	81.5	75.0	**62.0**	94.8	78.1	**78.0**	83.9	***82.0***	***82.0***
MIDAS-ADL	67.5	***65.0***	**65.0**	*86.8*	77.8	***77.3***	***82.9***	83.6	84.4
MIDAS-ADL-CF	83.5	76.2	***61.0***	89.7	***77.1***	118.7	**97.7**	85.4	84.0
Exp. Almon (q = 3)									
MIDAS-DL	89.0	78.8	***61.7***	*81.8*	*79.6*	***77.2***	***81.1***	*81.6*	*81.6*
MIDAS-ADL	*67.3*	*66.0*	**65.5**	92.6	94.1	**78.1**	82.5	**82.0**	82.5

MIDAS-ADL-CF	87.9	78.2	**60.9**	104.1	**95.0**	91.8	**83.7**	84.5	84.5
Beta									
MIDAS-DL	85.9	70.5	**63.7**	85.6	***73.2***	*75.7*	85.7	***83.2***	*85.2*
MIDAS-ADL	*70.2*	***62.7***	63.3	*79.6*	**74.1**	76.5	*84.3*	85.5	**83.1**
MIDAS-ADL-CF	78.7	70.7	***62.0***	82.1	**73.9**	75.9	84.6	***83.2***	83.3
Flat									
MIDAS-DL	77.8	73.9	**59.8**	88.2	***72.4***	*76.1*	***78.7***	*84.2*	*86.8*
MIDAS-ADL	*77.6*	*68.7*	***58.1***	*83.5*	**74.5**	79.1	**79.4**	84.6	88.2
Nonparametric									
MIDAS-DL	79.1	77.4	**62.0**	91.5	78.5	**78.1**	**80.9**	80.0	81.4
EW Forecast Combination									
MIDAS-DL	70.0	75.0	**61.5**	89.7	***77.0***	***77.0***	*82.0*	***81.0***	*82.0*
MIDAS-ADL	***59.8***	*66.4*	*61.2*	*86.3*	78.7	**77.2**	**82.6**	82.7	83.6
MIDAS-ADL-CF	77.6	75.6	***61.2***	90.2	**79.7**	85.6	86.8	**83.7**	84.0
Autoregression									
$p=1$	79.2	79.2	79.2	97.6	97.6	97.6	101.0	101.0	101.0

Notes: The evaluation sample comprises 18 data points, from July 2016 to December 2017. Characters in **bold** indicate the best number of variables, L_X, for each model and forecast horizon, h. Characters in *italics* indicate the best model for each number of variables, L_X, and forecast horizon, h

References

Armesto, M.T., K.M. Engemann and M.T. Owyang (2010), "Forecasting with Mixed Frequencies," *Federal Reserve Bank of St. Louis Review*, 92(6), pp. 521–36.

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

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

Cavallo, A. (2013), "Online and Official Price Indexes: Measuring Argentina's Inflation," *Journal of Monetary Economics*, 60(2), pp. 152–165.

Clements, M.P., and A.B. Galvão (2008), "Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States," *Journal of Business & Economic Statistics*, 26(4), pp. 546–554.

Diebold, F.X., and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13(3), pp. 134–144.

Duarte, C. (2014), *Autoregressive Augmentation of MIDAS Regressions*, Banco de Portugal Working Paper 201401, Lisbon, Portugal, Banco de Portugal.

Faust, J., and J.H. Wright (2009), "*Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset*," Journal of Business & Economic Statistics, 27(4), pp. 468–479.

Forni, M. et al. (2000), "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82(4), pp. 540–554.

Foroni, C., and M. Marcellino (2013), *A Survey of Econometric Methods for Mixed-Frequency Data*, Norges Bank Working Paper 2013-06, Oslo, Norway: Norges Bank.

Foroni, C., M. Marcellino and C. Schumacher (2015), "Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1), pp. 57–82.

Ghysels, E., V. Kvedaras and V. Zemly (2016), “Mixed Frequency Data Sampling Regression Models: The R Package MIDASr,” *Journal of Statistical Software*, 72(4), pp. 1–35.

Ghysels, E., P. Santa-Clara and R. Valkanov (2004), *The MIDAS Touch: Mixed Data Sampling Regression Models*, CIRANO Working Paper 2004s-20, Montreal, Canada, Center for Interuniversity Research and Analysis of Organizations (CIRANO).

Ghysels, E., A. Sinko and R. Valkanov (2007), “MIDAS Regressions: Further Results and New Directions,” *Econometric Reviews*, 26(1), pp. 53–90.

Giacomini, R., and H. White (2006), “Tests of Conditional Predictive Ability,” *Econometrica*, 74(6), pp. 1545–1578.

Guérin, P., and M. Marcellino (2013), “Markov-Switching MIDAS Models,” *Journal of Business & Economic Statistics*, 31(1), pp. 45–56.

Judge, G.G. et al. (1985), *Theory and Practice of Econometrics. Second Edition*, New York, United States, John Wiley & Sons.

Marcellino, M., and C. Schumacher (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), pp. 518–550.

Marcellino, M., J.H. Stock and M.W. Watson (2006), “A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series,” *Journal of Econometrics*, 135(1), pp. 499–526.

Miller, J.I. (2013), “Mixed-Frequency Cointegrating Regressions with Parsimonious Distributed Lag Structures,” *Journal of Financial Econometrics*, 12(3), pp. 584–614.

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

Nash, J.C., and R. Varadhan (2011), “Unifying Optimization Algorithms to Aid Software System Users: Optimx for R,” *Journal of Statistical Software*, 43(9), pp. 1–14.

Schumacher, C. (2016), “A Comparison of MIDAS and Bridge Equations,” *International Journal of Forecasting*, 32(2), pp. 257–270.

Stock, J.H., and M.W. Watson (2007), "Why Has US Inflation Become Harder to Forecast?," *Journal of Money, Credit and Banking*, 39(1), pp. 3–33.