Forecasting Brazilian inflation using machine learning techniques in mixed frequency models

* Diogo de Prince[a] and † Thiago Adriel da Cruz Alves Santos[b]

[1]Federal University of Sao Paulo and researcher at MacroLab-EESP/Getulio Vargas Foundation
[2]Federal University of Sao Paulo

Resumo

O estudo destaca a importância de previsões econômicas precisas, especialmente no Brasil, onde a inflação é historicamente volátil. O foco é prever a inflação brasileira, medida pelo Índice de Preços ao Consumidor Amplo (IPCA), utilizando um modelo de séries temporais de alta dimensão e frequência mista com regularização sparse-group LASSO (sg-LASSO) baseado em Babii et al. (2022). O período analisado vai de fevereiro de 2003 a abril de 2023. Para lidar com a alta dimensionalidade, são usados target factors para resumir as informações mais relevantes, reduzindo a complexidade computacional. O estudo compara quatro modelos e conclui que o modelo LASSO-U-MIDAS, que combina target factors e variáveis de inflação, apresenta a maior precisão nas previsões, especialmente para horizontes acima de um mês. A inclusão de variáveis brasileiras de inflação melhora a acurácia, enquanto a redução de dimensão antes da aplicação do LASSO pode prejudicar o desempenho das previsões.

Palavras chave: Previsão, LASSO, frequencia mista.

Abstract

The study emphasizes the importance of accurate economic forecasts, particularly in Brazil, where inflation is historically volatile. It focuses on predicting Brazilian inflation, measured by the Extended National Consumer Price Index (IPCA), using a high-dimensional mixed-frequency time series model with sparse-group LASSO (sg-LASSO) regularization based on Babii et al. (2022). The analysis covers the period from February 2003 to April 2023. To manage high dimensionality, we use target factors to summarize the most relevant information, reducing computational complexity. The study compares four models and concludes that the LASSO-U-MIDAS model, which combines target factors and inflation variables, shows the highest forecast accuracy, especially for horizons beyond one month. The inclusion of inflation variables improves accuracy, while dimension reduction before applying LASSO can negatively impact forecasting performance.

Key words: Forecasting, LASSO, mixed frequency.
Classificação JEL: C53, E37, C22
Área Anpec 4: Macroeconomia, Economia Monetária e Finanças

*Federal University of Sao Paulo and researcher at MacroLab-EESP/Getulio Vargas Foundation
E-mail diogo.mendonca@unifesp.br
†Federal University of Sao Paulo
E-mail thiago.adriel152@gmail.com

1 Introduction

Forecasting is one of the major aims of economic and econometric analysis. There is an effort among agents to improve the accuracy of the models, mainly by the monetary authorities - which depend on accurate forecasts to carry out the process of anchoring inflationary expectations (Medeiros et al., 2021). Emerging markets tends to have higher and highly volatile inflation. This issue places greater uncertainty on the decision-making process leading to mistaken decisions (Garcia et al., 2017). Due to its recent history of hyperinflation and indexed economy, more accurate forecasts are needed to mitigate risks in Brazil (de Prince et al., 2022). The relevance of this study is in line with efforts of institutions to forecast Brazilian inflation which is a more complex exercise than in advanced economies.

More accurate forecast decreases uncertainty in the decision-making process and helps to make an optimal choice about consumption and investment, which encourages the creation of models capable of accurately predicting economic variables (Araujo and Gaglianone, 2020). Time series models are typically estimated at same frequency. However there are cases in which the dependent variable is set close at lower frequency (Ghysels et al., 2004). Ghysels et al. (2004) have built a regression model capable of mix variable at different frequencies at same equation. This model allows covariates at higher frequencies to predict the dependent variable at a lower frequency, called as Mixed Data Sampling (MIDAS).

Our aim is to forecast brazilian inflation measured by the Extended National Consumer Price Index (IPCA). The approach is based on high dimension mixed frequency time series using *sparse-group* Least Absolute Shrinkage and Selection Operator (*sg*-LASSO) Regularization by Babii et al. (2022).

Several papers aim to increase the inflation accuracy from regression models as described in Vicente and Valls Pereira (2022), Aastveit et al. (2014), Medeiros and Vasconcelos (2016) and Bańbura et al. (2013). Our contribution is considering high dimensional mixed frequency models using *sg*-LASSO regularization to forecast Brazilian inflation that are little addressed in the literature. Furthermore, *sg*-LASSO is not a standard in Brazilian studies on inflation forecasting. Also, the sparse-group LASSO estimator can take advantage of high-dimensional time series data sampled at different frequencies and outperforms the classical LASSO estimator (Babii et al., 2022).

Mixed frequency models have a history of having accurate forecasts of inflation based on Garcia et al. (2017), Monteforte and Moretti (2013), Funke et al. (2015), Carlo and Marçal (2016), and Chen et al. (2023). However, the availability of a large number of variables leads to the curse of dimensionality, in which there are a large number of variables in the model compared to the number of observations. Such models are feasible by adding variable selection.

The dataset spans from February 2003 to April 2023, encompassing a robust range of time series for analysis. A total of 225 series were utilized, including 145 monthly series and 80 daily series. Our forecasting exercise is based on an expanding window. The forecast sample covers the period from January 2013 to April 2023. We have 124 observations for the one-period-ahead forecast sample. The number of observations for model estimation is at least 119 observations (period between February 2003 and December 2012).

We use target factors to manage the high dimensionality of monthly regressors in the MIDAS model. Given that including a vast number of monthly variables increases computational demands significantly, we opt to reduce this complexity by focusing on target factors to summarize monthly variables. Target factors serve as a preliminary step to identify and condense the most relevant information from a large set of potential predictors. This approach is grounded in the work of Bai and Ng (2008), who showed that pre-testing can enhance forecast accuracy by selecting the most pertinent variables for constructing factors. Relevant variables are then used to compute factors through PCA. We use target factors based on the adaptation from Medeiros and Vasconcelos (2016). Finally, the number of factors is determined using the criteria established by Bai and Ng (2002). We create two datasets: one incorporating only the target factors and another combining these factors with Brazilian monthly inflation variables, allowing us to effectively manage the dimensionality.

We use four type of models: Least Absolute Shrinkage and Selection Operator (LASSO), LASSO-U-MIDAS, LASSO-MIDAS, sparse group LASSO-MIDAS (sg-LASSO-MIDAS). To process this data, we employ Legendre polynomials of degree three to aggregate twelve lags of monthly macroeconomic indicators, as in Babii et al. (2022). LASSO is the standard autoregressive distributed lag (ARDL) model using classical LASSO estimator. LASSO-U-MIDAS keeps the weighting function unconstrained and we estimate 12 coefficients per high frequency covariate using the unstructured LASSO estimator. LASSO-MIDAS uses MIDAS weights together with the unstructured LASSO estimator. sg-LASSO-MIDAS applies the sg-LASSO estimator with MIDAS weights.

Basically, we have the highest forecast accuracy by LASSO-U-MIDAS using target factors and inflation variables from one to twelve periods ahead. The superiority of the forecast performance of the LASSO-U-MIDAS model compared to other models is considerable from the horizon of two periods ahead.

The dataset of target factors and inflation variables leads to better forecast accuracy than using only target factors. Therefore, dimension reduction before applying LASSO leads to losses from a forecasting perspective.

This paper is divided into five sections, besides this introduction. We present the literature review on machine learning for time series in the second section. The third section presents the methodology used to forecast Brazilian inflation. Next we present data in both monthly and daily frequency. In the fifth section, we present the results and forecasting performance of Brazilian inflation. Finally, we present our concluding remarks.

2 Literature review

Producing accurate inflation forecasts is a challenge for monetary policy makers and economic agents. An inflationary spiral considerably increases uncertainty about decisions shortening your horizon. In particular, emerging markets suffer greater consequences in a deteriorated inflationary scenario - due to recent past of high prices - which imposes value on building more accurate forecasts for this variable.

Emerging markets historically exhibit more volatile inflation than developed countries, which makes short-term inflation forecasting more critical than in advanced economies (Garcia et al., 2017). Accurate inflation predictions are essential for various economic sectors and hold significant value in fiscal policy, wage negotiations, and financial markets as described in (Carlo and Marçal, 2016). Therefore, there is a need to enhance the accuracy of inflation forecasting from now on.

Monetary policy decisions contribute to the formation of long-term expectations of economic agents and are based on information about current and future economic conditions. To substantiate their decisions, the regulator employs models to forecast economic variables, which rely on statistics provided by public and private institutions. However, economic variables may be released with delays and undergo constant fluctuations. This scenario necessitates models that take into account changes in the dependent variable and covariates not only for the future but also for the present and recent past. The exercise of predicting economic variables in the near future, present, and reconstructing the recent past has become popularly known as *Nowcasting* - or *real-time* (Giannone et al., 2006).

Real-time are models broadly studied in statistical and economic literature, however there are few studies that cover applications for inflation. Although inflation projection using nowcasting models is recent, the literature - Anesti et al. (2018), Aastveit et al. (2014), Bańbura et al. (2013), Kuzin et al. (2011), for example - deals with real-time GDP. GDP is a quarterly variable released with a delay by Brazilian Institute of Geography and Statistics (IBGE), approximately 60 days, which justifies the search for more accurate forecasts using models for this real-time variable. Thus, given that nowcasting is the forecast of the recent past, present, and very near future, we can use these models as an alternative to this issue.

Regression models involve variables always at the same frequency. However, there are cases where the dependent variable is at a lower frequency than the independent variables. In these cases, the simplest solution is to work with the variable at the lower frequency, which requires temporal aggregation of the variable that is at the higher frequency (Schumacher et al., 2012). This approach led to loss of relevant information and convolution of the dynamic relationship between the variables, resulting in unrealistic estimates - due to the large number of estimated parameters - and multicollinearity problems.(Breitung and Roling, 2015). To avoid these issues, an alternative would be to use *Mixed Data Sampling* (MIDAS) *models* proposed by Ghysels et al. (2004).

A time series is stationary when the statistical properties of the series do not depend on the time at which the series is observed, that is, constant mean and variance. A time series with a unit root is non-stationary, as its mean and variance can change over time. A unit root appears when the root of the characteristic equation is equal to 1, indicating the presence of a nonstationary trend in the series.

In the presence of nonstationary variables, there might be the so called spurious regression that happened when R^2 is high and *t-statistics* appear to be significant, but the results are without any economic meaning (Granger and Newbold, 1974).

To solve this issue we can outline a simple procedure called Dickey-Fuller Test.

MIDAS models allow for regressions using information at different frequencies parsimoniously. The main characteristic of these models is that they offer a more generic structure for the information, which prevents the loss of relevant information (Alper et al., 2008). In other words, MIDAS models allow for working with statistics at different frequencies, enabling the forecasting of the dependent variable - usually at a lower frequency - using explanatory variables at a higher frequency, thereby eliminating the problem of variable interpolation.

To build more accurate models for inflation may require a framework that can incorporate relevant variables and exclude irrelevant information. One way to address this issue would be to incorporate techniques of *machine Learning* able to identify non-linear patterns in the database (Araujo and Gaglianone, 2020). The LASSO (*Least Absolute Shrinkage and Selection Operator*) It is a method of variable selection and reduction that has the property of reducing the value of parameter estimates but is capable of producing estimates equal to zero for model parameters, as in the best subset selection method, thus generating interpretable models (Tibshirani, 1996). So, instead of imposing a quadratic penalty on the coefficient, LASSO penalizes its absolute value, which allows the shrinking of irrelevant variables towards zero (Medeiros and Vasconcelos, 2016). LASSO is essentially an optimization problem that consists of the usual minimization of squared errors plus a penalty on the regression model parameters.

Nowcasting is essentially a mixed frequency problem - as the dependent variable is at a lower frequency than the explanatory variables - and the LASSO estimator provides a solution to the difficulty of selecting the best subset of regressors, according Babii et al. (2022). Nowcasting with applications of *machine learning* techniques - especially LASSO - allows for increased forecasting accuracy through the selection of sparse and parsimonious models. Therefore, LASSO tends to solve variable selection problems in time series, covering a large volume of data. Therefore, Nowcasting models using LASSO can be an interesting solution for forecasting inflation due to the exclusion of interpolation of variables in databases with a large volume of data.

The first contribution of Babii et al. (2022) is to explain how the *sparse-group* LASSO estimator does not provide the same performance as the *unstructured* LASSO in small samples. They conclude that incorporating more specific information structures can be valuable in several applications. Furthermore, the authors reveals that employing *sparse-group-LASSO-MIDAS* solely based on macroeconomic data delivers superior statistical performance compared to forecasts from the New York Fed, particularly over a two-month horizon, and consistently shows strengh during the initial two months and at quarter-end. Babii et al. (2022) find that machine learning models are superior to state space dynamic factor models for this particular case.

Theory suggests that the feature of financial asset prices must contain information about the future path of the economy and should be considered as relevant for macroeconomic forecasting Andreou et al. (2013). The main goal to use readily available daily series is do not foregoes the possibility of providing *real-time* daily, weekly, or monthly updates of forecasts. Another reason to use it is to avoid loss of information through temporal aggregation Andreou et al. (2011). Monteforte and Moretti (2013), demonstrates that the inclusion of these variables reduces forecast errors and improves predictive performance, particularly when combined with monthly core inflation data. This is further supported by Sunon (2018), who finds that certain financial variables outperform a simple autoregressive benchmark model in forecasting inflation in Norway, Sweden, and Finland.

Time series models are essentially in same frequency. As long as inflation is available at monthly frequency, the main approach to analyse inflation in the future is to build models based on monthly variables correlated with future inflation. An alternative way is to look at financial indicators - such as yield curve - that are available on a daily basis (Monteforte and Moretti, 2013). This approach have the advantage to follow-up in sense to avoiding losses of information and providing *real-time* updates of forecasts.

The results on inflation forecasting in the international literature are diverse. Medeiros and Vasconcelos (2016) showed - for United States of America (USA) inflation - that high-dimensional models lead to lower forecast errors for some macroeconomic variables, including inflation. Funke et al. (2015) indicate the potential that models using mixed frequencies offer for *nowcast* for Chinese inflation. The authors obtain that models using macroeconomic variables as dynamic factors can be more accurate for projecting and *nowcast* by combining disaggregated variables.

For the Brazilian case and emerging countries, there are few studies that seek new ways to forecast inflation. Garcia et al. (2017) compare various econometric models to forecast inflation in real-time using a large number of predictors. They found that for a five-day horizon, the LASSO estimator and the market expectations captured by the FOCUS survey organized by the Central Bank of Brazil yield virtually the same result and good forecasts. However, LASSO outperforms all other methods for horizons beyond five-day horizon.

In terms of cumulative inflation, Garcia et al. (2017) explain that the Complete Subset Regression (CSR) is the best model. However, for the majority of forecasts from various models, they are not statistically different according to the model's confidence interval. Therefore, they consider the final projection as the average of the results of the models within the confidence interval. Additionally, the Bayesian Vector Autoregression (BVAR) model produces accurate forecasts, but not as good as these high-dimensional models. Thus, models from the LASSO family are better for one and two months ahead, while CSR models are better for longer horizons.

Carlos (2021) finds evidence that MIDAS models improve the accuracy of predictions when compared to naive models, such as random walk and AR(1) - first-order autoregressive - and are better than forecasts made by professionals in the FOCUS report. Additionally, MIDAS models tend to perform slightly better than the top 5 from FOCUS for up to 11 days. Therefore, the author finds that MIDAS models are better for anticipating the dynamics of inflation for the current month. Furthermore, Carlos (2021) finds that MIDAS models using daily information showed slightly more accurate predictions than models with weekly information. Finally, they show that the addition of ARIMA components to the MIDAS model tends to make the predictions even more accurate.

Carlo and Marçal (2016) compare inflation forecasts for Brazil over a horizon longer than 12 months, considering disaggregated and aggregated information. They found that the aggregation of models appears to show a significant improvement in forecast performance, which was not as evident for disaggregated models.

Vicente and Valls Pereira (2022) propose to forecast the inflation trajectory using prices from online purchases. They explain that online prices can be collected at a higher frequency than the frequency at which official inflation is reported. This allows for incorporating this variable into the mixed-frequency model in order to generate more accurate forecasts. The findings suggests that online inflation serves as a strong predictor of offline inflation trends. Notably, forecasts that maintain online inflation at its original frequency, rather than aggregating it to a lower frequency, outperform models that rely on a single frequency approach.

Borges and Portugal analyze whether the inclusion of monthly data increases the accuracy of a mixed-frequency Vector Autoregression (MF-VAR) model that uses mixed-frequency data - monthly and quarterly. The overall results indicate that using monthly observations within the quarter increases the accuracy of short-term forecasts for the Brazilian inflation.

3 Methodology

Next, we present the methodology used to forecast the Broad Consumer Price Index (IPCA). We present a high-dimensional mixed frequency time series using *sparse-group*-LASSO regularization as described in Babii et al. (2022), comparing it with naive models such as Random Walk.

3.1 High-dimensional MIDAS Regression

3.1.1 MIDAS regression

Consider the nowcast target y_t, $t \in [T]$, that we observe once in every period t. There are K potential covariates $x_{t-\frac{j-1}{m},k}$, $j \in [m]$, $k \in [K]$, that we observe m times in each period t. We can express the MIDAS Model from for nowcasting as

$$y_t = \sum_{k=1}^{K} \psi(L^{1/m};\beta_k)x_{t,k} + \varepsilon_t \tag{1}$$

where ε_t is the error term. We can write the lag polynomial as

$$\psi(L^{1/m};\beta_k)x_{t,k} = \frac{1}{m}\sum_{j=1}^{m}\omega\left(\frac{j-1}{m};\beta_k\right)x_{t-\frac{j-1}{m},k} \tag{2}$$

where β_k is a L-dimensional vector of coefficients with $L \leq m$ and $\omega : [0, 1]$ is a weighting function. The key element in MIDAS modeling is this weighting function.

One of the most used parametrizations is the exponential Almon lag, for example. In this case, we parametrize as

$$\omega(g;\vartheta) = \frac{exp(\vartheta_1 g + \vartheta_2 g^2)}{\sum_0^G exp(\vartheta_1 g + \vartheta_2 g^2))} \tag{3}$$

considering two parameters as suggested by Ghysels et al. (2007). Hyper-parameterized weighting scheme solves parameter proliferation in this scenario.

3.1.2 High-dimensional MIDAS Regression: general problem

The regression model integrates the target series denoted as $y_t : t \in [T]$, where t represents low-frequency time points (e.g., monthly) and high-frequency covariates, denoted as $x_{t-\frac{(j-1)}{m,k}} : j \in [m], t \in [T], k \in [K]$, where m represents the higher frequency (e.g., daily), and K is the number of covariates. The challenge is to effectively incorporate these high-frequency covariates into a model that accurately predicts the low-frequency target series.

$$\phi(L)y_t = \rho_0 + \sum_{k=1}^{K} \psi(L^{\frac{1}{m}}; \beta_k)x_{t,k} + u_t, t \in [T] \tag{4}$$

where $\phi(L)$ = I - $\rho_1 L$ - $\rho_2 L^2$ - .... - $\rho_j L^j$ is the low-frequency polynomial where L is the lag operator and ρ_i is the coefficient associated with the i lag of the variable. I is the identity matrix. $\psi(L^{\frac{1}{m}}; \beta_k)x_{t,k} = \frac{1}{m}\sum_{j=1}^{m} \omega\left(\frac{j-1}{m}; \beta_k\right) x_{t-\frac{j-1}{m},k}$ is the high-frequency polynomial, where $L^{\frac{1}{m}}$ represents the fractional lag operator, accommodating higher frequency and $\beta_{j,k}$ are coefficients for the high-frequency covariates.

When m = 1, the model simplifies to a standard ARDL (Autoregressive distributed Lag) model. With $J + 1 + m \times K$ parameters, the model can become highly complex, leading to overfitting, especially when the number of covariates K is large, and traditional estimation methods may fail. For example, in the case of state space models, implementing this data can be challenging as most of these models do not support more than 40 series.

To address this problem, regularization techniques like LASSO can be used to manage high-dimensional data. LASSO imposes a penalty on the size of coefficients, shrinking some to zero to achieve sparsity.

However, adding a penalty function can make the optimization problem non-convex. This is an issue when specifying MIDAS, given that machine learning typically involves convex optimization problems.

The first insight of Babii et al. (2022) is to solve the convexity problem. They approximate the weight function as

$$\omega(s; \beta_k) \approx \sum_{l=1}^{L} \beta_{k,l} f_l(s) \tag{5}$$

where $u \in [0, 1]$, and $\{f_l : l = 1, ..., L\}$ is a collection of approximating functions, denominated as dictionary.

Using (5) in Equation (4):

$$\phi(L)y_t = \rho_0 + \sum_{k=1}^{K} \frac{1}{m} \sum_{j=1}^{m} \sum_{l=1}^{L} \beta_{k,l} f_l \left(\frac{j-1}{m}\right) x_{t-\frac{j-1}{m},k} + u_t \tag{6}$$

and we can rewrite as

$$\phi(L)y_t = \rho_0 + \sum_{k=1}^{K} \frac{1}{m} \sum_{l=1}^{L} \beta_{k,l} z_{t,l,m} + u_t \tag{7}$$

where $z_{t,l,m} = f_l\left(\frac{j-1}{m}\right) x_{t-\frac{j-1}{m},k}$. Thus, the non-linear MIDAS model becomes the linear regression with transformed covariates. Babii et al. (2022) recommend Legendre polynomials.[1] These polynomials pre-filter high-frequency information and determine the number of basis functions needed to estimate the models.

Based on linear regression (7), we can use LASSO to select covariates and L. One problem with using LASSO - among others[2] - is that Lasso does not recognize time series structure (Tibshirani, 1996). Model

[1] Legendre Polynomials are beneficial because: i) they are orthogonal, which stabilizes coefficients and reduces multicollinearity, ii) they approximate arbitrary MIDAS weight functions well in $L_2[0, 1]$, encompassing all continuous MIDAS and beyond. For example, Legendre polynomials are orthogonal, while exponential Almon weights are not and are suject to multicollinearity problem increasing the dimension.

[2] The standard Lasso literature assumes that: 1) The data are not only independently and identicaly distributed (i.i.d)

matrices in each group are assumed to be orthonormal. Depending on the value of λ, an entire group of predictors may be excluded from the model.

But we have a group structure. For a single covariate, a set of all its high-frequency lags is a group. Lags of y_t is also a group. We could use Group Lasso alone. However, Group Lasso does not achieve sparsity within a group. If a group of parameters is non-zero, all its members will be non-zero. This presents a challenge in selecting the degree of the polynomial. This issue can be resolved with a structured sparsity solution.[3] We incorporate such structure into the estimation procedure using the sparse-group LASSO (sg-LASSO). The sg-LASSO considers sparsity between and within groups, and we can capture predictive information from each group, such as approximating functions from the dictionary or specific covariates from each group.

Consider $\mathbf{y} = (y_1, ..., y_T)^\top$ a vector of dependent variable and let $\mathbf{X} = (\iota, \mathbf{y}_1, ..., \mathbf{y}_J, Z_1F, ..., Z_KF)$, where $\iota = (1, ..., 1)^\top$ is a vector of ones, $\mathbf{y}_j = (y_{1-j}, ..., y_{T-j})^\top$, $Z_k = (x_{k,t-\frac{j-1}{m}})$ is a $T \times m$ matrix of the covariate with $t \in [T]$ and $j \in [m]$, $k \in [K]$, and $F = \left(f_l\left(\frac{j-1}{m}\right)/m\right)$ is an $m \times L$ matrix of weights with $j \in [m]$ and $l \in [L]$. Let $\beta = (\beta_0^\top, ..., \beta_K^\top)^\top$, where $\beta_0 = (\rho_0, \rho 1, ..., \rho_J)^\top$ is a vector encompassing intercept and autoregressive coefficients. β_k represents the parameters of the high-frequency lag polynomial associated with the covariate $k \geq 1$. The sg-LASSO estimator is based on the penalized least-squares problem

$$min_{b \in \mathbb{R}^p} ||\mathbf{y} - \mathbf{X}b||_T^2 + 2\lambda\Omega(b) \tag{8}$$

with a penalty function

$$\Omega(b) = \alpha|b|_1 + (1-\alpha)\|b\|_{2,1} \tag{9}$$

where $\|b\|_{2,1} = \sum_{G \in \mathcal{G}} |b_G|_2$ is the group LASSO norm and $\mathcal{G}$ is a group structure that we specify. Sparse-group Lasso (sg-Lasso) (Simon et al., 2013) addresses this by combining the L^1 norm for the entire parameter space $(\alpha|b|_1)$ and the L^2 norm for each group separately $(1-\alpha)\|b\|_{2,1})$. This approach assumes that covariates appear in groups rather than individually over time, making the model sparse between groups of covariates and encouraging sparsity at the group level. We use the LASSO estimator if $\alpha = 1$ and the group LASSO estimator if $\alpha = 0$.

Therefore, we use sg-Lasso for three reasons. First, it recognizes data structures. Groups are all (high frequency) lags of a specific covariate. Second, within groups penalty allows for polynomial lag model selection. Third, LASSO and group LASSO are special cases and single group resembles the elastic net.

We select the value of tuning parameters λ and α using the 10-fold cross-validation. Folds are adjacent blocks over the time dimension that take into account the time series dependence.

In short, we have four methods for selecting variables: LASSO, LASSO-U-MIDAS, LASSO-MIDAS, sg-LASSO-MIDAS. LASSO is the standard autoregressive distributed lag (ARDL) model using classical LASSO estimator. LASSO-U-MIDAS keeps the weighting function unconstrained and we estimate 12 coefficients per high frequency covariate using the unstructured LASSO estimator. LASSO-MIDAS uses MIDAS weights together with the unstructured LASSO estimator. In this case, we estimate $L+1$ coefficients per high-frequency covariate, where L is the degree of Legendre polynomials. We use $L = 3$ as Babii et al. (2022) so that is 4 coefficients per high-frequency covariate. sg-LASSO-MIDAS applies the sg-LASSO estimator with MIDAS weights. High-frequency covariates lags are flow-aggregated.

but also sub-Gaussian, meaning they follow a normal distribution. 2) Sparcity over high-frequency Lags $j = 0, ..., m-1$ is dubious. 3) Lasso does not recognize time series structure.

[3] Structured sparsity involves: 1) Defining a group as a set of all lags in high-frequency for a single covariate. 2) High-dimensional model selection considers both the size of the dictionary and the selection of covariates. 3) Performing selection at two levels: within groups—learning the weight type of MIDAS regression—and between groups—identifying the most relevant covariates.

3.2 Target Factors

Since the dimension of monthly regressors is too large in the MIDAS family model and the computational time becomes too large, we replace this high dimensionality of monthly variables with their factors. Our choice is to target factors as a pre-test step to reduce the dimension of monthly variables.

Bai and Ng (2008) demonstrated that when faced with a large set of potentially relevant variables, we can enhance our forecasts by pre-testing to identify which variables are most pertinent for predicting the target variable when constructing factors.

Our target is to forecast y_t. Let h the forecasting horizon, and W_t a set of control variables. W_t includes twelve lags of y_t. Additionally, let X_{it}, for $i = 1, \ldots, N$, represent the candidate variables. Based on Medeiros and Vasconcelos (2016), we follow the procedure as follows:

1. For each $i = 1, \ldots, N$, fit a regression of y_{t+h} on W_t and X_{it}, and record the t-statistics for all X_{it}.
2. Sort the absolute values of the t-statistics in descending order.
3. Select a significance level α. The variables considered relevant will be those with $|t_i|$ larger than the critical value of 1.96.
4. Let $x_t(\alpha)$ denote the selected relevant variables. Compute factors F_t from $x_t(\alpha)$ using principal component analysis.
5. Fit a regression of y_{t+h} on W_t and $f_t \subset F_t$.

We select the number of factors using the criteria from Bai and Ng (2002).

4 Data

Our main goal is to nowcast IPCA (Brazilian consumer price index) using machine learning algorithms. We have a dataset with two frequencies - monthly and daily - spanning from January 1, 2002 to May 01, 2023. The initial dataset contained 930 variables: 658 daily and 272 monthly.

Data processing involves collecting observations separately for each variable rather than having the entire dataset available at once. In the database, variables had been discontinued. We excluded variables that contained excess missings and this process was carried out in two steps. First, we removed all variables that contained missings in the last 3 months to exclude series that were discontinued during this period. Our second step was removing series whose missings were above the median of missings from the database at each frequency. The median of missings for each frequency was 2232 for the daily and 62 for the monthly variables. After this data cleaning, we have 225 variables. Our database contains: (i) 145 monthly variables with 256 observations, and (ii) 80 daily variables with 4582 observations.

We can divide the dataset into seven groups: I) Prices; II) Money and Finance; III) Production and Sales; IV) Foreign Sector; V) Public Sector; VI) Labor; Employment, Income and Consumption; and VII) Expectations; divided into 128 brazilian variables and 128 are global one. We have international variables to analyze whether the inflation nowcast becomes more accurate. We consider that the global economy impacts Brazilian prices according to Dées and Galesi (2021) which states that changes in US monetary policy roughly effects international equity prices, capital flows, and global growth.

Once the data is organized, we test the stationary properties in our time series data. For this purpose, we use the Augmented Dickey–Fuller(ADF) test which the null hypothesis is that there is a unit root which imples that the data is non-stationary. The alternative hyphotesis is presence of stationarity that could assume three diferent forms: i) with drift and trend, ii) with drift and iii) without drift. We specify 15 higher-order lags used to capture the higher-order autocorrelation in the monthly dataset and 20 higher-order lags in the daily one. Following Stock and Watson (2004), the best model has been chosen based on Akaike Information Criteria (AIC) in that models with smaller values are preferred. After determining the type of stationarity of the time series, we verify from the most general - with drift and trend - case whether the time series has an unit root and - if the answer is positive - it is necessary to take first difference. Otherwise, we analyze the second more general case - with drift - to check the existence of a unit root in the series and to take the fist difference. Finally, we analyze the last case without constant to check whether the series has a unit root and perform the first difference if the test is positive and - otherwise - verify that the time series does not in fact have a unit root. With the tests completed, we obtain the final database with the series specified in level or variation.

5 Results

We nowcast Brazilian inflation using macroeconomic and financial data, with details provided in Appendix A. Inspired by Medeiros and Vasconcelos (2016), we use target factors to reduce the number of monthly covariates in the model. After we define which monthly variables are relevant, we extract the factors from these remaining variables. The computational time to estimate using sg-LASSO-MIDAS with expanding window including all the monthly variables is unfeasible, for example.

Then, from all the monthly variables, we build two monthly datasets. The first is to extract target factors from all the 145 monthly variables so that the only monthly covariates are the target factors tp be selected in the model. The second is to consider the 17 Brazilian monthly inflation variables separately and extract target factors from the 128 other monthly variables. Thus, the monthly regressors are the target factors plus the Brazilian monthly inflation variables.

Predictions are generated through a expanding window. The initial nowcast is for January 2013, using data spanning eleven years (132 months). Forecasts are made from January 2013 (1-month horizon) through December 2013 (12-month horizon), continuing until the data sample ends in April 2023. We present results for forecast horizons ranging from 1 to 12 months.

We use four type of models: LASSO, LASSO-U-MIDAS, LASSO-MIDAS, sg-LASSO-MIDAS. To process this data, we employ Legendre polynomials of degree three to aggregate twelve lags of monthly macroeconomic indicators, as in Babii et al. (2022). We fine-tune the sg-LASSO-MIDAS regularization parameters, λ and α, using 10-fold cross-validation, with time-adjacent blocks as folds to account for the time series nature of the data. We compare our predictions against the Random Walk model as a benchmark, a standard baseline for short-term Brazilian inflation forecasts.

Table 1 displays the RMSE for one to twelve horizons of different models. We have two types of monthly variables as covariates: (i) target factors from all 145 monthly variables (denominated here as target factors), and (ii) a group of 17 monthly variables related to Brazilian inflation and target factors of the remaining 128 monthly variables (denominated as target factors and inflation). The values in Table 1 represent the RMSE multiplied by 10. This metric measures the average difference between the predicted and actual values, providing a proxy of the model's predictive accuracy.

Table 1: Root mean square error forecasting performance for 1 to 12 months ahead for different models

Horizon		Target Factors and Inflation				Target Factors			
	Random Walk	LASSO	LASSO-U-MIDAS	LASSO-MIDAS	sg-LASSO-MIDAS	LASSO	LASSO-U-MIDAS	LASSO-MIDAS	sg-LASSO-MIDAS
1	3.67	3.12	2.48	3.37	3.24	3.36	4.21	5.29	5.14
2	4.50	3.09	1.15	3.21	3.21	3.25	3.99	4.95	4.86
3	4.95	3.02	1.93	3.33	3.32	3.25	3.99	4.95	4.86
4	5.43	3.11	1.24	3.21	3.14	3.44	3.52	4.91	4.80
5	5.72	3.07	1.25	3.07	3.01	3.34	3.51	4.94	4.90
6	5.65	3.06	1.10	3.11	3.01	3.33	3.34	5.00	4.89
7	5.54	3.07	1.04	3.07	3.04	3.41	3.34	5.27	5.02
8	5.33	3.06	1.09	2.82	2.77	3.40	3.29	4.86	4.39
9	5.09	3.07	1.13	2.98	2.93	3.40	3.44	4.82	4.59
10	4.93	3.00	1.12	2.94	2.81	3.33	3.28	4.80	4.58
11	4.99	3.04	1.06	3.09	2.94	3.35	3.27	4.71	4.61
12	5.07	3.05	1.05	3.00	2.94	3.37	3.33	5.05	4.70

The results presented in Table 1 highlight the LASSO-U-MIDAS method with target factors and inflation variables as the best, having a lower RMSE compared to the other models for all forecast horizons analyzed. The forecasting performance of LASSO-U-MIDAS with target factors and inflation variables is closer to that of the other models for one month ahead, although it is superior. In the case of forecast horizons above one, the forecast accuracy of LASSO-U-MIDAS is considerably better than that of the others.

This best LASSO-U-MIDAS result is different from Babii et al. (2022) who report sg-LASSO-MIDAS with greater accuracy among the different models for forecasting the quarterly GDP growth rate from monthly and quarterly variables. The sg-LASSO-MIDAS method presents a worse forecasting performance in the case of having only the target factors.

Furthermore, the forecast accuracy of models using target factors is lower compared to target factor and Brazilian monthly inflation variables separately. So, a dimensionality reduction before the LASSO method negatively affects the forecast performance.

Also, any method using target factors and inflation variables has higher forecast accuracy than random walk. This is not the case using only target factors.

6 Conclusion

In this paper, we set out to improve the accuracy of inflation forecasts for Brazil by employing a high-dimensional mixed-frequency approach, specifically using LASSO regularization technique, as in Babii et al. (2022). Our aim is to evaluate the effectiveness of mixed frequency models in predicting monthly Brazilian inflation, measured by the IPCA. We analyze the forecasting performance of the models from one to twelve months ahead based on the root mean square error. We use expanding window with forecasts from January 2013 to April 2023.

We use four type of models: LASSO, LASSO-U-MIDAS, LASSO-MIDAS, sg-LASSO-MIDAS. The dataset contains 81 daily variables and 145 monthly covariates. Due to the large computational time required to estimate the models, we choose to reduce the dimension of monthly covariates using target factors of these monthly variables. We have two possibilities of monthly databases: (i) target factors from the 145 monthly variables, and (ii) 17 Brazilian monthly inflation variables and target factors from the remaining 128 monthly variables.

The results indicate that the LASSO-U-MIDAS method, incorporating target factors and inflation variables, outperforms other models in terms of lower RMSE across all forecast horizons. While its advantage is modest for one-month-ahead forecasts, it becomes significantly superior for longer horizons.

This finding contrasts with previous research by Babii et al. (2022), which found that the sg-LASSO-MIDAS model was more accurate for forecasting quarterly GDP growth. In our study, the sg-LASSO-MIDAS method performs worse when only target factors are used.

Moreover, models using target factors alone have lower forecast accuracy than those combining target factors with inflation variables. This suggests that dimensionality reduction prior to applying the LASSO method negatively impacts forecast performance.

Additionally, any model incorporating both target factors and inflation variables performs better than a random walk model, a result not observed when using target factors alone.

References

K. A. Aastveit, K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud. Nowcasting gdp in real time: A density combination approach. *Journal of Business & Economic Statistics*, 32(1):48–68, 2014.

C. E. Alper, S. Fendoglu, and B. Saltoglu. Forecasting stock market volatilities using midas regressions: An application to the emerging markets. 2008.

E. Andreou, E. Ghysels, and A. Kourtellos. Forecasting with mixed-frequency data. 2011.

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

N. Anesti, A. B. Galvão, and S. Miranda-Agrippino. Uncertain kingdom: nowcasting gdp and its revisions. 2018.

G. S. Araujo and W. P. Gaglianone. Machine learning methods for inflation forecasting in brazil: New contenders versus classical models. Technical report, Mimeo, 2020.

A. Babii, E. Ghysels, and J. Striaukas. Machine learning time series regressions with an application to nowcasting. *Journal of Business & Economic Statistics*, 40(3):1094–1106, 2022.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1): 191–221, 2002.

J. Bai and S. Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

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

B. K. Borges and M. S. Portugal. Previsão macroeconômica para o brasil utilizando o modelo var com dados mistos.

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

T. C. Carlo and E. F. Marçal. Forecasting brazilian inflation by its aggregate and disaggregated data: a test of predictive power by forecast horizon. *Applied Economics*, 48(50):4846–4860, 2016.

T. C. Carlos. *Forecasting inflation with mixed frequency data sampling: an application for Latin America.* PhD thesis, 2021.

C. Chen, Y. Sun, and Y. Rao. Threshold midas forecasting of inflation rate. Technical report, 2023.

D. de Prince, P. L. Valls Pereira, and E. F. Marçal. Are professional forecasters rational? what is the role of instability and what variables affect them? *What is the role of instability and what variables affect them*, 2022.

S. Dées and A. Galesi. The global financial cycle and us monetary policy in an interconnected world. *Journal of International Money and Finance*, 115:102395, 2021.

M. Funke, A. Mehrotra, and H. Yu. Tracking chinese cpi inflation in real time. *Empirical Economics*, 48 (4):1619–1641, 2015.

M. G. Garcia, M. C. Medeiros, and G. F. Vasconcelos. Real-time inflation forecasting with high-dimensional models: The case of brazil. *International Journal of Forecasting*, 33(3):679–693, 2017.

E. Ghysels, P. Santa-Clara, and R. Valkanov. The midas touch: Mixed data sampling regression models. 2004.

E. Ghysels, A. Sinko, and R. Valkanov. Midas regressions: Further results and new directions. *Econometric reviews*, 26(1):53–90, 2007.

D. Giannone, L. Reichlin, and D. H. Small. Nowcasting gdp and inflation: the real-time informational content of macroeconomic data releases. 2006.

C. W. Granger and P. Newbold. Spurious regressions in econometrics. *Journal of econometrics*, 2(2): 111–120, 1974.

V. Kuzin, M. Marcellino, and C. Schumacher. Midas vs. mixed-frequency var: Nowcasting gdp in the euro area. *International Journal of Forecasting*, 27(2):529–542, 2011.

M. C. Medeiros and G. F. Vasconcelos. Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, 138:50–52, 2016.

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

L. Monteforte and G. Moretti. Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, 32(1):51–61, 2013.

C. Schumacher, M. Marcellino, and C. Foroni. U-midas: Midas regressions with unrestricted lag polynomials. 2012.

N. Simon, J. Friedman, T. Hastie, and R. Tibshirani. A sparse-group lasso. *Journal of computational and graphical statistics*, 22(2):231–245, 2013.

J. H. Stock and M. W. Watson. Combination forecasts of output growth in a seven-country data set. *Journal of forecasting*, 23(6):405–430, 2004.

K. H. Sunon. Forecasting inflation: The predictive ability and stability of financial variables. *Journal of Applied Business and Economics*, 20(7):169, 2018.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

H. d. O. R. Vicente and P. L. Valls Pereira. Forecasting inflation using online daily prices: a midas approach for brazil. *Available at SSRN 4044997*, 2022.

Appendices

A Dataset

Index	Name of series	Frequency
1	Brazil CPI IPCA MoM	M
2	Brazil CPI IPCA IBGE Food Inflation MoM	M
3	Brazil CPI IPCA IBGE Transportation Inflation MoM	M
4	Brazil CPI IPCA IBGE House Inflation MoM	M
5	Brazil CPI IPCA IBGE Health Inflation MoM	M
6	Brazil IPCA-15 CPI Extended National MoM	M
7	Brazil CPI INPC MoM	M
8	FGV Brazil General Prices IGP-10 MoM	M
9	FGV Brazil General IGP-M MoM	M
10	FGV Brazil General Prices IGP-DI MoM	M
11	FGV Brazil Wholesale Prices IPA-10 MoM	M
12	FGV Brazil IGP-M Wholesale Prices IPA-M MoM	M
13	FGV Brazil Wholesale Prices IPA-DI MoM	M

14	FGV Brazil CPI IPC-10 MoM	M
15	FGV Brazil IGP-M CPI IPC-M MoM	M
16	FGV Brazil IGP-M Construction Prices INCC-M MoM	M
17	FGV Brazil Construction Prices INCC-DI MoM	M
18	US Inflation Indexed CPI Ratio 30 Yr Bonds Issued April 1998	D
19	US Inflation Indexed CPI Ratio 30 Yr Bonds Issued April 1999	D
20	Pictet LPP 2000 - 60 Daily Index Switzerland	D
21	Sonia Deposit Rates Swap 1 Month Daily Compounding	D
22	India Net Foreign Equity Investment 12 Month Rolling Sum USD	D
23	Sonia Deposit Rates Swap 3 Month Daily Compounding	D
24	Brazil CETIP DI Rate Accumulated	D
25	Brazil Financial Index	D
26	US Continental Gas Heating Degree Days Historical	D
27	Brazil Reference Interest Rate TR	D
28	BRL Interest Rate Return	D
29	USD-BRL Int Rate Spread Curncy	D
30	Brazil Savings Accounts Deposit -1 Day Yield Rate	D
31	Brazil Cetip DI Interbank Deposit Rate	D
32	Brazil Selic Target Rate	D
33	EURBRL Spot Exchange Rate - Price of 1 EUR in BRL	D
34	Brazil Central Bank SDR Avg Rate	D
35	Brazilian Real Spot	D
36	Ibovespa Brasil São Paulo Stock Exchange Index	D
37	Brazil Money Market CDB DI Floating - Daily Return	D
38	Brazil Selic Average Overnight Daily Rate	D
39	Crude Oil Futures Front Contract in Brazilian Real	D
40	Brazil ANBIMA Estimated Index Assumption IGP-M	D
41	Banque de France Total Retail Sales Volume SWDA YoY	M
42	Italy Banks: Deposits in EUR of Dom Res	M
43	Brazil Monetary Base	M
44	Brazil Money Supply M1 Brazil M1	M
45	Brazil Money Supply M2 Brazil M2	M
46	Brazil Money Supply M3 Brazil M3	M
47	Brazil Money Supply M4 Brazil M4	M
48	Brazil Monetary Base Bank Reserves	M
49	Brazil Financial System Loans	M
50	Brazil Personal Loans More Than 90 Days Late	M
51	Brazil Nonperforming Loans Public Financial Institutions	M
52	Brazil Business Loans 15 to 90 Days Late	M
53	Brazil Total Savings Deposits	M
54	SXI Switzerland Sustainability 25	D
55	SXI Switzerland Sustainability 40	D
56	Open Interest - Gold	D
57	Open Interest - Crude Oil	D
58	Lean Hogs 60 Day Historical Volatility	D
59	BCOM Roll Yield	D
60	Crude Oil 60 Day Historical Volatility Index	D
61	Open Interest - Brent	D
62	Open Interest - Silver	D
63	NASS Cattle on Feed For All US States Reporting Data	M
64	Port of Long Beach Inbound Containers	M
65	NASS Cattle on Feed Placements Data	M
66	ISM Total Manufacturing & Non-Manufacturing New Orders	M
67	USDA Monthly Cold Storage Beef Only	M
68	Malaysian Palm Oil Board Crude Palm Oil Production Data	M
69	NASS Cattle on Feed for 7 States Placements Year over Year Percentage Data	M
70	Malaysia Palm Oil Board Total Palm Oil Inventory Data	M
71	USDA Monthly Cold Storage Pork Bellies	M
72	NASS Cattle on Feed For Total of US States Year over Year Percentage Data	M
73	Malaysian Palm Oil Board Crude Palm Oil Inventory Data	M
74	Multifamily Starts to Total Starts NSA	M
75	USDA Monthly Total Milk Production All States	M
76	Japan Monthly Semiconductor Wafer Exports Volume to Foreign Countries	M
77	USDA Commercial Cattle Number Slaughtered	M
78	Brazil Active Focus of Fires in Pantanal	M
79	NASS Cattle on Feed For 7 States Marketing Data Year over Year Percent	M
80	Brazil Industrial Production Activity Manufacturing Industry 2012 SA	M
81	Brazil Industrial Production Activity Manufacturing Industry MoM2012	M
82	Brazil Industrial Production Activity Extractive Industry MoM2012	M
83	Anfavea Brazil Vehicle Production	M
84	Brazil Retail Sales Revenue Supermarkets Index NSA	M
85	Brazil ANP Sales of Ethanol by State-National Total	M
86	Brazil ANP Sales of Ethanol by State-National Total	M
87	Brazil Crude Oil Monthly Exports	M
88	Port of Los Angeles Total Containers	M
89	Brazil Soybeans Monthly Exports	M
90	Brazil Terms of Trade Price Index	M
91	Malaysia Palm Oil Exports	M
92	Brazil Iron Ore Monthly Exports	M
93	Port of Long Beach Total Containers	M
94	Brazil Commercial Banks Foreign Exchange Position	M
95	Brazil Import Price Index	M
96	36 South Global Implied Volatility Commodities	M
97	North America Tractors 40 and Under 100 HP Retail	M
98	Japan Total Floor Area of New Housing Construction Started SA	M
99	Bloomberg Total OPEC Crude Oil Production Output	M
100	Bloomberg OPEC Crude Oil Production Output Data/Saudi Arabia	M
101	Bloomberg Crude Oil Historical Price	M
102	Semiconductor Industry Association (SIA) Global Semiconductor Sales Data	M
103	DOE EIA US Crude Oil Production (MBPD)	M
104	World Crude Oil & Liquid Fuels End of Period Inventory OECD Commercial	M
105	World Crude Oil & Liquid Fuels Production Total World Supply	M
106	Bloomberg OPEC Crude Oil Production Output Data/Iran	M
107	NOAA Climate Prediction Center Oceanic Nino Index ONI	M

108	Bloomberg OPEC Crude Oil Production Output Data/Venezuela	M
109	Middle East Arab Light Crude Saudo to Asia OSP Spread vs Average Oman/Dubai FOB	M
110	Bloomberg OPEC Crude Oil Production Output Data/Nigeria	M
111	World Crude Oil & Liquid Fuels Consumption Total World Consumption	M
112	DOE Natural Gas Dry Production Data	M
113	Japan Crude Cocktail Detailed Release Monthly in USD	M
114	Japan JCC LNG (JLC) Import Price USD/MMBtu	M
115	DOE Russia Crude Oil Production Data	M
116	IISI Crude Steel Production Data/China	M
117	Total US Natural Gas Marketed Production	M
118	DOE Saudi Arabia Crude Oil Production Data	M
119	Potassium Chloride (Muriate of Potash) Standard Grade: FOB Vancouver Spot Price	M
120	Brazil International Reserves - Liquidity Concept - Total - US$ MM	M
121	Brazil Trade Balance FOB Imports NSA	M
122	Brazil Current Account Monthly	M
123	Brazil Current Account Balance on Goods and Services	M
124	Brazil Current Account Construction Credit	M
125	Brazil BOP Portfolio Investment Foreign in Fixed Income	M
126	Brazil BOP Portfolio Investment Net	M
127	Brazil Financial Account Portfolio Investment Net Incurrence of Liabilities	M
128	Brazil Financial Account Direct Investment Balance	M
129	Brazil BOP Financial Derivatives Assets	M
130	Brazil Fin Acct Portfolio Investment Equity and Inv Fund Shares Assets Bought	M
131	Brazil Fin Acct Portfolio Investment Equity and Inv Fund Shares Assets Sold	M
132	Brazil BOP Capital Account Net	M
133	Brazil BOP Financial Account Net	M
134	Brazil BOP Errors and Omissions	M
135	Brazil Fin Acct Portfolio Investment Acquisition of Financial Assets Credit	M
136	Brazil Financial Account Direct Investment Intercompany Assets	M
137	Brazil Financial Account Portfolio Investment Debt Securities Long-Term Assets	M
138	Brazil Financial Account Loans Net Incurrence of Liabilities	M
139	Brazil Financial Account Central Banks Currency and Deposits Liabilities	M
140	Brazil Importing Tax Income Nominal	M
141	Cost of Funds Federal	M
142	Brazil Financial Account Governement Currency and Deposits Assets	M
143	Brazil Total Net Debt in % of GDP	M
144	Brazilian States Debt to Foreigners in % of GDP	M
145	Brazilian Federal Government Debt to Foreigners in % of GDP	M
146	Brazilian Federal Government Debt	M
147	Brazilian Federal Government Domestic Debt	M
148	Brazilian Treasury Securitized Domestic Debt	M
149	Brazilian States Debt	M
150	Brazil Public Primary Budget Result	M
151	Brazil Public Nominal Budget Result	M
152	Brazil Federal Govt and Central Bank Primary Balance	M
153	Brazil Public Nominal Interest Payments	M
154	Brazil Social Security Values Expenditures	M
155	Brazil Social Security Values Revenues	M
156	Brazil Total Federal Revenue	M
157	Brazil Income Tax Collection Nominal	M
158	Brazil Central Government Net Revenue	M
159	Brazil Central Government Revenue from the National Treasury	M
160	Brazil Central Government Total Expenditures	M
161	Brazil Public Net Fiscal Debt	M
162	Brazil Public Net Fiscal Debt % of GDP	M
163	Brazil National Treasury Revenue from Industrialized Products Tax	M
164	Housing Turnover	M
165	Existing Housing Sales Total Home Sales	M
166	Single Family Starts Percent of Total Starts SAAR	M
167	Home Improvement Expenditures	M
168	Brazil ANP Oil Production (Barrels)	M
169	Brazil Total Electricity Consumption	M
170	Brazil Consumer Goods SA MoM	M
171	Brazil Primary Income Compensation of Employees	M
172	Brazil Secondary Income	M
173	Brazil Secondary Income From General Government	M
174	Brazil Real Minimum Wage	M
175	Brazil FGV Consumer General Price Index Market First 10 Day Period Preview	M
176	Brazil FGV Consumer General Price Index Market Second 10 Day Period Preview	M
177	1 Year Curve - Crude Oil	D
178	BarCap US Corp HY YTW - 10 Year Spread	D
179	Bloomberg USDJPY 3 Month Hedging Cost	D
180	Bloomberg EURUSD 3 Month Hedging Cost	D
181	Real 10 Year Yield Based on Core CPI US	D
182	real 10 Year Yield Based on Headline CPI US	D
183	Bloomberg USDEUR 3 Month Hedging Cost	D
184	Bloomberg EURJPY 3 Month Hedging Cost	D
185	Bloomberg USDJPY 1 Month Hedging Cost	D
186	Bloomberg USDCHF 3 Month Hedging Cost	D
187	Bloomberg USDJPY 12 Month Hedging Cost	D
188	Bloomberg AUDJPY 3 Month Hedging Cost	D
189	Bloomberg GBPUSD 3 Month Hedging Cost	D
190	Bloomberg EURGBP 3 Month Hedging Cost	D
191	Bloomberg EURCHF 3 Month Hedging Cost	D
192	Bloomberg USDJPY 6 Month Hedging Cost	D
193	Real 10 Year Yield Based on Core CPI Japan	D
194	Bloomberg GBPJPY 3 Month Hedging Cost	D
195	Real 2 Year Yield Based on Core CPI US	D
196	Real 10 Year Yield Based on Headline CPI Eurozone	D
197	Bloomberg EURJPY 12 Month Hedging Cost	D
198	Real 10 Year Yield Based on Core CPI Eurozone	D
199	Real 10 Year Yield Based on Headline Japan	D
200	1 Year Curve - Corn	D
201	Real 10 Year Yield Based on Headline CPI Germany	D
202	Bloomberg GBPUSD 1 Month Hedging Cost	D
203	Real 2 Year Yield Based on Core CPI Eurozone	D
204	Real 2 Year Yield Based on Headline CPI Eurozone	D

205	1 Year Curve - Natural Gas	D
206	Real 10 Year Yield Based on Core CPI Canada	D
207	1 Year Curve - Brent Crude Oil	D
208	1 Year Curve - Gold	D
209	1 Year Curve - Copper	D
210	1 Year Curve - Silver	D
211	Bloomberg EURGBP 1 Month Hedging Cost	D
212	Bloomberg GBPJPY 1 Month Hedging Cost	D
213	1 Year Curve - Soy Meal	D
214	Bloomberg USDJPY 2 Year Hedging Cost	D
215	Real 10 Year Yield Based on Headline CPI Canada	D
216	Real 2 Year Yield Based on Headline CPI Germany	D
217	Bloomberg EURNOK 3 Month Hedging Cost	D
218	Real 2 Year Yield Based on Headline CPI UK	D
219	BFV United States Interpolated Yield Govt 10 Year Strip 90 Day Volatility USD	D
220	1 Year Curve - Wheat	D
221	Real 2 Year Yield Based on Core CPI Japan	D
222	Real 2 Year Yield Based on Headline CPI Japan	D
223	Brazil CPI IPCA Median Market Expectation Next 12 Months YoY	D
224	Brazil CPI IGP-M Median Market Expectation Next 12 Months YoY	D
225	Peru Central Bank Expectation Survey CPI 12 Months Ahead	M

Note: D (Daily) and M (Monthly).