Applying the daily inflation to forecast the Broad Consumer Price Index (IPCA)

Pedro Costa Ferreira
Brazilian Institute of Economics – Getulio Vargas Foundation (FGV/IBRE)
Barão de Itambi, 60 – Botafogo, Rio de Janeiro - Brazil
pedro.guilherme@fgv.br

Juliana Carneiro
Brazilian Institute of Economics – Getulio Vargas Foundation (FGV/IBRE)
Barão de Itambi, 60 – Botafogo, Rio de Janeiro - Brazil
julianaccp@gmail.com

André Furtado Braz
Brazilian Institute of Economics – Getulio Vargas Foundation (FGV/IBRE)
Barão de Itambi, 60 – Botafogo, Rio de Janeiro - Brazil
andre.braz@fgv.br

Abstract: Since 2006, the Getulio Vargas Foundation (FGV) calculates a daily version of the Broad Consumer Price Index (IPCA), the official inflation index, calculated under the responsibility of the IBGE, the federal statistics agency in Brazil. Ardeo *et. al.* (2013) showed the importance of this indicator and how this daily information can be useful to a country that had high level of inflation. Despite the fact that this measure is a fair antecedent variable for inflation, due to some peculiarities concerning the collection period, the initial daily rating may not anticipate some effects, such as seasonal factors and the increase in prices controlled by the Brazilian Government. Hence, by taking into account the Monitor´s daily time series, this paper intends to forecast the IPCA for the first six days of data collection. The results showed up that the proposal technic improved the IPCA forecast in the beginning of data collection.

Key-words: IPCA, daily inflation, Monitor, Time Series, SARIMA

JEL code: E3; C22

1. INTRODUCTION

The predictability of financial and economic phenomena is a matter of the utmost importance for all actors in the market, including stakeholders and households. Accordingly, the more accurate processes for forecasting these events (as well as micro and macroeconomic variables) are, the more efficient the financial market will be. This is, it will be possible for the country to outperform, boosting not only its economy, but also its growing pace. (DANTHINE & DONALDSON, 2011)

The Consumer Price Index is one of the macroeconomic variables that attracts major attention from analysts since it measures inflation. For this reason, many Brazilian and overseas entities study this issue. For instance, since 1947, the Getulio Vargas Foundation estimates the Consumer Price Index (IPC) (IBRE/FGV, 2015).

In 1979, the Brazilian Institute of Geography and Statistics (IBGE) started to disclose the Broad Consumer Price Index, IPCA. IBGE is the agency responsible for statistical information in Brazil, which includes data collection and the IPCA release.

The IPCA is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services (BUREAU OF LABOR STATISTICS, 2015). Moreover, it is used by the Brazilian Central Bank as the guideline for achieving its inflation targets policy (SOUZA JÚNIOR & LAMEIRAS, 2013).

The index represents all goods and services purchased for consumption by the reference population. The sample is composed by urban inhabitants whose monthly incomes range from one to forty minimum wages. Each month, IBGE data collectors visit or call thousands of retail stores, service establishments, rental units, doctor´s offices and concessionaires of public services to obtain information on the prices of items used to track and measure price changes in the CPI.

Each category of goods receives different weights depending on the importance the households prioritize when purchasing their basket of goods. This is, the weights change over time. Major groups and examples of categories are food and beverages, transportation and housing, which one corresponding approximately to 23%, 20% and 14% of the basket of goods respectively (IBGE, 2012).

Due to the importance of the subject, there are a plethora of works trying to do "good" prices forecasting, as it can be seen in (SAZ, 2011), in which the authors analyze the efficacy of using the models SARIMA (Seasonal Autoregressive Integrated Moving Average) in order to forecast inflation rates in Turkey. The researcher found a capable, parsimonious, accurate and appropriate SARIMA time-series model of forecasting for inflation in Turkey between 2003 and 2009.

Another study using SARIMA models for inflation forecasting in the short term can be found in (PUFNIK & KUNOVAC, 2006). The research is done by the Croatian

National Bank since the entity recognizes the importance of inflation forecasting as an essential component for the monetary policy projection. Furthermore, by using seasonal processes ARIMA, it is possible to understand not only the CPI as a whole, but also its elements and weights so that the Bank can afford better views concerning detailed sources of future either inflationary or deflationary pressures in the Croatian economy.

In Brazil, owing to the hyperinflation episode in the 1980s, there have been made many studies on this issue. Moreover, the Brazilian economy faced the most extreme inflation phenomenon, with yearly price increases of three-digit percentage points and an explosive acceleration (RESENDE, 1989). Besides this, it is still a recurrent issue even after the monetary stabilization brought about by the Real Plan (Plano Real). This was a set of measures taken to stabilize the Brazilian economy in 1994, during the presidency of Itamar Franco (GIAMBIAGI, VILLELA, DE CASTRO, & HERMANN, 2011).

Accordingly, theoretical frameworks have been constructed so that they are able to explain the origin and the duration of this economic anomaly (BARBOSA & SALLUM, 2002). Besides this, many researchers cast light on the inflation tax effect caused by high level of prices, mainly because Brazil is a country composed eminently by lower class. Therefore, it becomes an awkward kind of tax that corrodes the incomes of the less well-off (BARBOSA, 2014). The author estimates the curves for the inflation tax related to hyperinflationary processes occurred both in Germany and in Brazil. It is worth noting that the issue also concerns policy makers and the financial market.

In view of that, in 2006, FGV created a new methodology aimed at measuring daily prices variations. As showed in (ARDEO, QUADROS, & PICCHETTI, 2013), the Inflation Monitor makes daily estimations based on data from prices of the last 30 days, as a *proxy* for the official inflation calculated by IBGE. Despite the fact that this measure is a fair antecedent variable for inflation, due to some peculiarities concerning the collection period, the initial daily rating may not anticipate some effects, such as seasonal factors and the increase in prices controlled by Brazilian Government.

Hence, by taking into account the Monitor´s daily time series, this paper intends to forecast the IPCA for the first six days of data collection. Furthermore, in order to bring it about the time series was divided into six other as follows. The first one was built taking the first day of the index collection (made by IBGE); the second series took into consideration the two ensuing days of the collection start; the third data set considered three days and so on until the sixth time series.

The IPCA prevision will be made in conformity with the methodology created by Box & Jenkins, that is, the SARIMA models (Seasonal Autoregressive Integrated Moving Average). SARIMA (p,d,q) $(P,D,Q)_s$ is used when seasonal (hence nonstationary) behavior is present in the time series (BOX & JENKINS, 1970).

The present paper is relevant since it aims at doing a daily forecast for the Brazilian CPI (IPCA), as well as acting as a complement for the Inflation Monitor. It is to say that the prediction will be more robust for the first days in each month.

Besides this first, the present work is composed by three more sections. The second one deals with the Inflation Monitor, which are the main source of the data base used for the models; the third section deals with the methodology, as well as the proposed model; the forth part shows the results achieved by this work; and, then, final conclusions.

2. INFLATION MONITOR

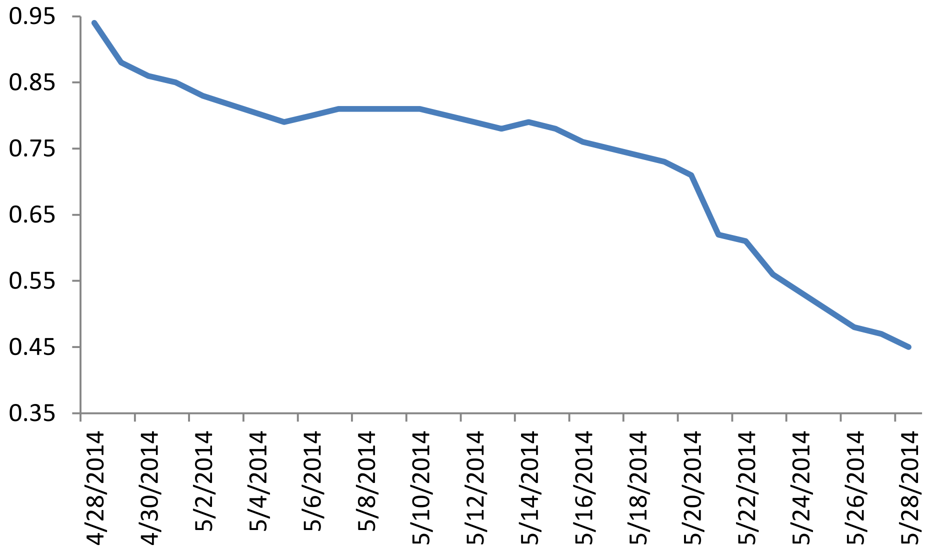
Since 2006, the Getulio Vargas Foundation (FGV) calculates a daily proxy of the Broad Consumer Price Index (IPCA) for 30 days, ending in the date of computation. It is measured in harmony with the Laspeyres Index, whereby the weights are monthly adjusted according to changes in the relative prices. The daily appraisal combines both price collection under the responsibility of FGV and the calculation procedures followed by IBGE. This data set is called Inflation Monitor (ARDEO, QUADROS, & PICCHETTI, 2013).

For calculating the daily proxy for IPCA it is necessary to consider that the sample prices follow a uniform distribution over time as the new prices are constantly added to the time series and processed in the same day.

The announcement dates of IPCA (by IBGE) provide the parameters required by the Monitor to carry out the estimations. After issuing the publication, the weights become known and prepared to be used in the coming month. Moreover, such piece of information is straightaway embodied by FGV.

One of the various advantages that the daily indicator affords is calculating the changes in the monthly rate of the official index day by day. As can be seen in graphic 1, there is a decreasing tendency for the IPCA over may 2014. Furthermore, the second fortnight reveals a sharp decline in price index. In using these results, the financial market operators and the monetary authority are able to catch a glimpse of the index behavior as well as its tendency for the ensuing month so that they can steer their decision making. It is of the utmost importance for the former to have access to this kind of daily information since they use them to beacon their decision on portfolio diversification. Besides this, as the Central Bank of Brazil deliberates on monetary policy, and as it is aimed at the short run, the daily information plays a key role also for the latter.

Graphic 1: Daily IPCA - Monitor



Source: FGV. By the authors.

In terms of disadvantage, the Monitor uses data from the previous month when doing the estimation of the index in the current month. The Monitor does a kind of moving average in order to calculate the daily IPCA, however, as it uses data from other month, its prediction can be biased and impaired. The day the Monitor is more accurate coincides with the last day of collection of IPCA done by IBGE. Thus, on this day, the margin of error is minimized (very short). Therefore, the closer it is from this date, the more precise the Monitor is when estimating the price index since it uses data that come from the same month, as shows graphic 2.

Graphic 2: Monthly IPCA and Monitor on the day of shutting

Source: FGV. By the authors

The proposed model intends to minimize that flaw by using SARIMA models. Capturing seasonal and tendency effects is a huge advantage offered by this model as the Monitor is neither capable of performing this way, nor able to amend other irregularities.

3. METODOLOGY BOX & JENKINS AND PROPOSED METHOD

3.1. BOX & JENKINS Models

Time series can be either stationary or nonstationary; either stochastic or deterministic process. A stochastic process that has a Gaussian distribution can presents weak stationarity. That is the mean and the variance of a stochastic process do not depend on t (that is they are constant) and the autocovariance between X_t and $X_{t+\tau}$ only can depend on the lag τ (τ is an integer, the quantities also need to be finite), which is the temporal distance between the observations (HAMILTON, 1994); (NASON, 2008).

The Box & Jenkins models are used to deal with time series (TS) originally stationary or made stationary by differencing – that is computing the differences between consecutive observations. On one hand, transformations such as logarithms may stabilize the variance of a time series; on the other hand, differencing can stabilize the mean of a time series. Generally, economic time series are non-stationary, thus they need to be differenced until they become stationary.

The Box and Jenkins methodology to stationary time series and to ARIMA time series forecasting follows an iterative cycle composed by five parts (GRANGER & NEWBOLD, 1976):

1- Specification: the general class of the structures (p,d,q) is analyzed.
2- Identification: based on sample ACF and PACF and other criteria. If the autocorrelation function (ACF) plot shows a very slow decay, then the time series is supposed to be non-stationary. Thus, one must do unit root tests in order to confirm statistically the graphic hypothesis. If the null hypothesis is not rejected, then diffencing is required so that the time series can become stationary eventually.
3- Estimation: the parameters of the identified model are estimated and tests are made to determine their statistical significance.
4- Diagnosis: The residuals are analyzed and must be white noise. Ljung-Box test is also necessary to verify the model fitting. Afterwards, it is necessary to verify which models have the smallest values for Akaike information criterion (AIC) and Bayesian information criterion (BIC) tests. Should the diagnosis phase show problems, one must go back to identification phase.
5- Definitive model: for forecasting or control. One must verify which models have the best RMSE and MAPE (it is worth noting that the latter cannot be applied for values close to zero; in this case, it is recommended that one use other method to analyze the errors).

An ARIMA (p,d,q) process is an ARMA with *d* differencing (differencing should be done until the process become stationary). The SARIMA models are used with series which shows a periodic behavior over time (*s* times). That is, when similar performances are find time after time (with periodicity *s*) (BOX & JENKINS, 1970). This is the case of the time series that this paper deals with.

3.2. PROPOSED METHOD

The method used in this paper is the one step ahead forecast by employing SARIMA models (Seasonal Autoregressive Integrated Moving Average), in accordance with Box & Jenkins methodology as explained above.

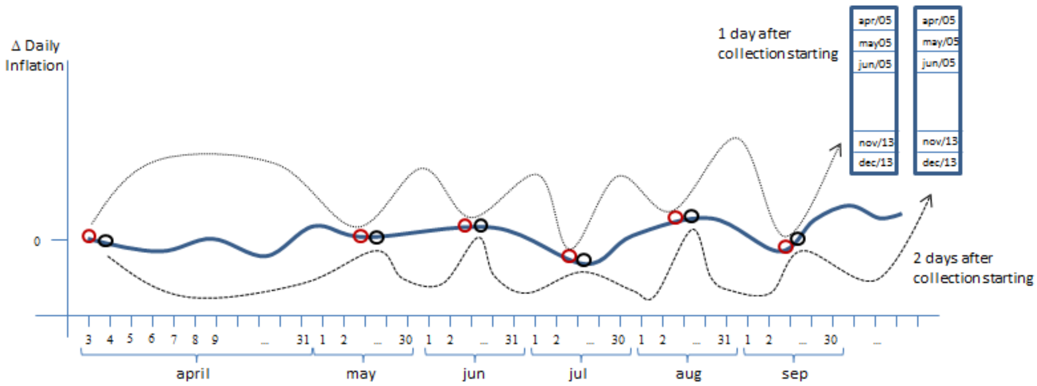
The time series were constructed from a data set from the Inflation Monitor. The Monitor series are built by using moving averages concerning previous periods of time. For instance, the IPCA released on May 3 is estimated with data starting on April 3, and so on. Therefore, the closer it is from the collection beginning, the greater the sort of data will be related to the antecedent month. Thus, the forecasting error will be also greater (Monitor).

As discussed in the first section, there were built six time series (TS) from daily data set for this work. It is to say that the first one is created with parts of information concerning to the first day of data collection. By following the example showed in graphic 3, one can imagine that the collection always starts the third day of every month. In connection with that, there is a monthly TS with information related to one day after the collection starting and with “too much” information dating back the previous month. To build the second time series, daily data from the second day after the collection starting were used, and so on until the sixth TS. These six time series were used in this paper in order to forecast daily IPCA.

Ultimately, it is worth remembering that the aforesaid hypotheses, though useful, do not retreat the reality since the collection period changes every month. Besides this, the days for the series can be Saturday, Sunday or Holiday, when the Monitor does not estimate price index.

By using the six monthly time series and SARIMA models, it is possible to do one step ahead forecasting for the IPCA. Furthermore, the purpose of this article is using these subterfuges (construction of six time series + SARIMA models) in order to boost the forecasting performed by the Inflation Monitor since its inflation estimations put together too much information from the last month.

Graphic 3 – Draft of the time series built in this article



Source: By the authors.

4. RESULTS

The six time series used in the models were made up of 87 observations between February 2007 and April 2014. Then, it was created a one-step-ahead forecasting for May 2014. Moreover, the forecasts were built for one, two, three until six days (in the same month) after the last day that IPCA is collected by the IBGE.

The Box & Jenkins methodology for SARIMA models were used. In analyzing the ACF and PACF, the IPCA time series could be considered stationary in the level part and non-stationary in the seasonal part. Then, it was necessary differencing in the seasonal part so that the non-stationarity could be corrected. The unit root test (DICKEY & FULLER, 1979) rejected $H_{0,}$ revealing that the series were eventually stationary, as it can be seen in the table 1 below.

Table 1: ADF Test (Augmented Dickey-Fuller)

Series	Lags	t -Statistic (ADF)	Critical Value 1%	Critical Value 5%
1	5	-4,9899	-4,04	-3,45
2	5	-4,8807	-4,04	-3,45
3	5	-4,9842	-4,04	-3,45
4	5	-4,8832	-4,04	-3,45
5	5	-4,9017	-4,04	-3,45
6	5	-4,9527	-4,04	-3,45

Source: By authors (software R)

There were made tests for the normality of data (Jarque-Bera), whose null hypothesis of skewness and curtosis equal to zero was not rejected (JARQUE & BERA, 1987). Some models for every single time series were tested and the best-fit models are showed in the table 2 below.

Table 2: Adjusted models

Series	Chosen models
1	$(0,0,2)(0,1,1)_{12}$
2	$(0,0,1)(0,1,2)_{12}$
3	$(0,0,2)(0,1,1)_{12}$
4	$(0,0,2)(0,1,1)_{12}$
5	$(0,0,2)(0,1,1)_{12}$
6	$(0,0,2)(0,1,1)_{12}$

Source: By authors (software R)

The ACFs of the residual confirmed normality , in other words, they are statistically equal to zero. These were the expected results and corroborate the models fitting. Afterwards, there were made Jarque Bera test for normality and Ljung Box test for serial autocorrelation (G.M.LJUNG & G.E.P.BOX, 1978).

The monthly forecasts using daily data showed the following results. The six time series, that is those built from the six first days after the release of IPCA by IBGE (each of the six series correspond to the number of days after the official announcement) presented low Mean Absolute Error (MAE) when compared to the same measure related to the Monitor errors for the same period of time (table 3). The MAE increased for longer periods of time (this fact was observed by some other models constructed for forecasting day beyond the initial six) and was more significantly than the Inflation Monitor.

Table 3: Mean Absolut Error (MAE)

Series	MAE	
(days ahead)	Monitor	Model
1	0,048	0,043
2	0,044	0,031
3	0,041	0,037
4	0,044	0,038
5	0,042	0,039
6	0,039	0,038

Source: By authors (software R)

It was verified that for the first six days the forecasting proposed in this paper is more accurate than the estimation made by the Inflation Monitor. In addition to this result, after the sixth day, the latter proved to be errorless when estimating daily inflation.

5. CONCLUSIONS

The core purpose of this work was to reduce the forecast error of daily IPCA done by the Inflation Monitor as it is more accurate for the closest days to the announcement date of the official index (done by IBGE). It happens because the Monitor uses data from the previous month. In doing a kind of moving average, the Monitor absorbs the price index (IPCA) daily; thus, the closer to the end of the month it is (when the IPCA is officially announced), the greater the amount of data collected in the same month in analysis is, which conveys to more accurate daily estimations.

By using SARIMA models, the daily IPCA could be forecasted one step ahead. In order to do this, there were built six time series corresponding to the period between February 2007 and April 2014.

The achieved results corroborate the tested hypothesis by the models that the first collections done shortly after the release of the official index and that are forecasted until six days after this date can be forecasted more accurately by the models used in this paper.

The data used to create the time series stem from the Inflation Monitor, which does not take the seasonal feature of IPCA into consideration in its daily forecasting. Despite this fact, the model deals with and treats the seasonality so that it manages to improve the results showed by the Monitor for the first six days of the IPCA collection, as well as to correct further irregularities.

Should those details be corrected, the results will be more accurate. Therefore, it is proposed for future works to consider those aspects when modeling. Furthermore, daily forecasts could be done by using SARIMA models.

As Pufnik ans Kunovac (2006) have done, this article suggests the hypothesis that more precise results could be achieved by, firstly, forecasting the components of the index themselves and, afterwards, by aggregating them in an index as a whole. The authors put forward for consideration the implementation of these tests in further works on the theme.

6. REFERENCES

ARDEO, V., QUADROS, S., & PICCHETTI, P. (2013). A daily frequency inflation measure and its information content on forecasts.

BARBOSA, F. D. (2014). As Curvas do Imposto Inflacionário nas Hiperinflações da Alemanha e do Brasil. EPGE/FGV.

BARBOSA, F. D., & SALLUM, É. M. (2002, out-dez). Hiperinflação: um arcabouço teórico. *Revista Brasileira de Economia*.

BOX, G. E., & JENKINS, G. M. (1970). *Time Series Analysis forecasting and control.*

DANTHINE, J., & DONALDSON, J. (2011). *Intermediate Financial Theory*.

DICKEY, D. A., & FULLER, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, pp. 427–431.

G.M.LJUNG, & G.E.P.BOX. (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika*, pp. 297-303.

GIAMBIAGI, F., VILLELA, A., DE CASTRO, L. B., & HERMANN, J. (2011). Economia brasileira contemporânea 1945-2010. Campus.

GRANGER, C. W., & NEWBOLD, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society B 38*, pp. 189–203.

HAMILTON, J. D. (1994). *Time series analysis.*

JARQUE, C. M., & BERA, A. K. (1987). A test for normality of observations and regression residuals. pp. 163–172.

NASON, G. (2008). *Wavelet Methods in Statistics with R.* Springer.

PUFNIK, A., & KUNOVAC, D. (2006). *Short-Term Forecasting of inflation in Croacia with seasonal ARIMA processes.* Croation National Bank.

RESENDE, A. L. (1989, janeiro-março). Da inflação crônica à hiperinflação: observações sobre o quadro atual. *Revista de Economia Política*, pp. 7-20.

SAZ, G. (2011). The Efficacy of SARIMA Models for Forecasting Inflation Rates. *International Research Journal of Finance and Economics*.

SOUZA JÚNIOR, J. R., & LAMEIRAS, M. A. (2013). In *Evolução Recente das políticas monetária e cambial e do mercado de crédito no Brasil.* IPEA.