Estimating Brazilian Monthly GDP: a State-Space Approach*

João Victor Issler, Hilton Hostalacio Notiniz

Novembro de 2015

URL: http://hdl.handle.net/10438/14273
Os artigos publicados são de inteira responsabilidade de seus autores. As opiniões neles emitidas não exprimem, necessariamente, o ponto de vista da Fundação Getulio Vargas.

ESCOLA DE PÓS-GRADUAÇÃO EM ECONOMIA
Diretor Geral: Rubens Penha Cysne
Vice-Diretor: Aloisio Araujo
Diretor de Ensino: Carlos Eugênio da Costa
Diretor de Pesquisa: Humberto Moreira
Vice-Diretores de Graduação: André Arruda Villela & Luis Henrique Bertolino Braido

Victor Issler, João
26p. - (Ensaios Econômicos; 774)
Inclui bibliografia.

CDD-330
Estimating Brazilian Monthly GDP: a State-Space Approach*

João Victor Issler† Hilton Hostalacio Notini‡

November 24, 2015

Abstract

The first contribution of this paper is to employ a superior interpolation method that enables to estimate, nowcast and forecast monthly Brazilian GDP for 1980-2012 in an integrated way; see Bernanke, Gertler and Watson (1997, Brookings Papers on Economic Activity). The second contribution, along the spirit of Mariano and Murasawa (2003, Journal of Applied Econometrics), is to propose and test a myriad of interpolation models and interpolation auxiliary series – all coincident with GDP from a business-cycle dating point of view. Based on these results, we finally choose the most appropriate monthly indicator for Brazilian GDP. Third, this monthly GDP estimate is compared to an economic activity indicator widely used by practitioners in Brazil - the Brazilian Economic Activity Index - (IBC-Br). We found that our monthly GDP tracks economic activity better than IBC-Br. This happens by construction, since our state-space approach imposes the restriction (discipline) that our monthly estimate must add up to the quarterly observed series in any given quarter, which may not hold regarding IBC-Br. Moreover, our method has the advantage to be easily implemented: it only requires conditioning on two observed series for estimation, while estimating IBC-Br requires the availability of hundreds of monthly series. The third contribution is to illustrate, in a nowcasting and forecasting exercise, the advantages of our integrated approach. Finally, we compare the chronology of recessions of our monthly estimate with those done elsewhere.

JEL codes: C32, E32, E37

Keywords: GDP Interpolation, State-space Representation, Kalman Filter, Composite and Leading Indicators, Nowcasting, Forecasting.

*We gratefully acknowledge the comments of participants at CIRET 2008 in Santiago, Chile. We also thank Regis Bonelli and Silvia Matos for allowing us to reproduce the results of their nowcast exercise. All errors are ours. We thank Marcia Waleria Machado, Marcia Marcos, and Rafael Burjack for excellent research assistance and thank CNPq-Brazil, FAPERJ and INCT for financial support.

†Corresponding author: Graduate School of Economics – EPGE, Getulio Vargas Foundation, Praia de Botafogo 190, s. 1100, Rio de Janeiro, RJ 22250-900, Brazil.

‡Finance Manager at Itaú-Unibanco Holding SA
1 Introduction

Any modern society is concerned with its current “state” of economic activity and what should be that state in the near future. Entrepreneurs and individuals are interested in the question because their profits and welfare are a function of it. Governments also have an interest in the subject for budgetary and welfare issues.

For the U.S., the most educated estimate of economic-activity turning points is embodied in the binary variable announced by the NBER Business Cycle Dating Committee: recession vs. expansion for any given month. These announcements are based on the consensus of a panel of experts, and they are made some time (usually six months to one year) after a turning point in the business cycle has occurred. So, today, we only possess old dating’s for the state of the economy. An alternative to infer business-cycle conditions would be to use Gross Domestic Product (GDP). Indeed, Stock and Watson (1989) argue that, if we were to choose one variable to best represent the state of the economy, this variable would be GDP. They claim that “[...] fluctuations in aggregate output are at the core of the business cycle so the cyclical component of real GDP is a useful proxy for the overall business cycle [...]”.

Unfortunately, GDP is also not timely available to continuously infer what is the current state of the economy. First, as far as we know, all countries compute GDP on a quarterly and/or on an annual frequency, but not on a monthly frequency. Second, there are delays in the release of quarterly GDP. Of course, delays vary across countries. In Brazil the delay in the initial release of quarterly GDP is bigger than three months and can reach up to six months at times. Moreover, this initial release is subject to major revisions in six months’ time.

The fact that a given statistic is only available with some delay does not invalidate its later use for inference if we possess a long-span time-series of it. Indeed, referring to the decisions of the NBER Business Cycle Dating Committee, Issler and Vahid (2006) argue that:

“Suppose that we are asked to construct an index of the health status of a patient. Also, suppose that we know that the best indicator of the health of the patient is the results of a blood test. However, blood samples cannot be taken too frequently, and test results are only available with a lag, sometimes too long to be useful. Our index therefore must be a function of variables such as blood pressure, pulse rate and body temperature that are readily available at regular frequencies. In order to estimate the best way to combine these variables into an index, would we (i) use the historical data on these variables only, or, (ii) use the historical blood test results as well? The answer is, obviously, the latter.”

Here, GDP and NBER dating are the equivalent to the results of the blood test. What should be the equivalent to blood pressure, pulse rate and body temperature? Since the pioneering work of Burns and Mitchell (1946), we have been endowed with coincident and leading indicators of economic activity, which are timely proxies, respectively, of the current state of economic activity and what will it be in the near future. This suggests that a superior strategy for having timely and more frequent estimates of economic activity is to combine GDP (or NBER dating) with information on these timely proxies. Issler and Vahid focused on
combining NBER dating information with coincident and leading series of economic activity. Our focus here will be to combine the latter data for Brazil with Brazilian GDP.

As argued above, this strategy can improve our knowledge of GDP in two dimensions. The first is that we will possess more frequent business-cycle estimates, i.e., monthly instead of quarterly. This solves an interpolation problem that was proposed by Bernanke, Gertler and Watson (1997) with an econometric model that encompassed most interpolation models that had been previously proposed in the literature; see also the extension in Mönch and Uhlig (2005). The second is that we will be able to nowcast and forecast GDP using information on coincident and leading series of economic activity. This solves a nowcasting problem, to say the least, and has been proposed by Mariano and Murasawa (2003).

The main contributions of this paper are twofold. First, we propose a joint model for Brazilian GDP and Brazilian coincident series that can be used to interpolate and nowcast GDP. Our model is based on the encompassing methods proposed by Bernanke, Gertler and Watson (1997) and in Mönch and Uhlig (2005). Second, for forecasting, we will propose an alternative model for Brazilian GDP and Brazilian leading series. It is essentially an extension of the model proposed by Mariano and Murasawa (2003) – who are mostly concerned with nowcasting – but will be used for forecasting purposes since we will switch the roles of coincident and leading series in implementing the model. As is clear, our main contribution is empirical, although our motivation to use and combine currently available methods is somewhat original.


The econometric models used for interpolation (also in nowcasting and forecasting) in this paper employ a state-space approach. They have three advantages over other interpolation methods: first, they allow the estimation of the unobserved monthly GDP with aggregation consistency, i.e., they ensure that the sum of three-months unobserved GDP data in a given quarter is equal to the respective quarterly observed GDP data. Second it encompasses a wide range of models in the literature – allows different ways to treat non-stationarity and different assumptions about the monthly residuals. Third, nowcasting and forecasting are computed imposing the restriction (discipline) that monthly estimates should add up to quarterly observations in-sample, which yields a superior behavior out-of-sample, as our empirical results confirm.

Finally, with the estimated monthly GDP series on hand, we (re)establish a chronology of recessions of the Brazilian economy, comparing our results with that of the recent literature and with those of the Brazilian Business-Cycle Dating Committee (CODACE). Our turning points are determined using a standard method applied to our interpolated GDP figures – the Bry and Boschan (1971) algorithm.

The paper is organized as follows. In section 2, we review the state-space models used
to interpolate quarterly GDP. In section 3, we discuss and analyze the related series used in the interpolation procedure. In section 4 we present our main empirical results. In section 5 we present a dating chronology for Brazil Business Cycles. Section 6 concludes.

2 State-Space Model and the Kalman Filter for Interpolation

2.1 State-Space Representation

In this section, we give a brief review of the use of state-space models and the Kalman filter. More detailed descriptions can be found in Hamilton (1994) and Harvey (1989). The Kalman filter provides an efficient computational (recursive) mean to estimate the state of a stochastic process – usually put in state-space form. The latter encompasses relationships between observable and non-observable series and a key issue is the estimation of non-observables using past information and/or all information available.

The state-vector representation is given by a system of two vector equations. First, the state or transition equation describes the dynamics of the state vector of possibly unobserved series we want to estimate. The second vector equation represents the observation or measurement equation linking the state vector to the vector containing the observed series \( y_t \) and possibly some related series \( x_t \) which are either exogenous or pre-determined.

The state-space representation of the dynamics of \( y_t \) for \( t = 1, \ldots, T \), where \( T \) is the number of observations in our sample, is given by the following system of two vector equations:

\[
\begin{align*}
\xi_{t+1} &= F\xi_t + v_{t+1} \\
y_t &= A' x_t + H' \xi_t + w_t
\end{align*}
\]

where \( F, A', \) and \( H' \), are matrices of parameters of dimensions \( r \times r, n \times k, \) and \( n \times r, \) respectively. So far, these matrices do not vary across time, but that can be relaxed in a more general setup. Indeed, this will be the case for interpolation, where matrix \( H \) will vary within a given year but is held fixed for any given month across different years.

Both vector equations have unpredictable error terms\(^1\) which, for estimation purposes, are assumed to have a multivariate Normal distribution as follows:

\[
\begin{pmatrix}
v_t \\
w_t
\end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} \right).
\]

The coefficients matrices \( F, A', \) and \( H' \), and the two variance-covariance matrices \( Q \) and \( R \) can be estimated consistently by maximizing the conditional log-likelihood function of the system, given initial conditions \( \xi_{1|0} \) and on its variance-covariance matrix, labelled \( P_{1|0} \); we discuss that below. For the time being we assume that these parameters are known.

\(^1\)Multivariate white-noise processes.
2.2 Filtering and Smoothing

Suppose we observe $y_1, x_1, y_2, x_2, \cdots, y_T, x_T$, but want to estimate the series in the unobserved state-vector $\xi_t$. Denote by $\mathcal{I}_t$ the information set using the observed series $y_1, x_1, y_2, x_2, \cdots, y_t, x_t$ and the conditional forecast of $\xi_{t+1}$ using information up to $t$ as:

$$\xi_{t+1|t} = \mathbb{E} \left( \xi_{t+1} | \mathcal{I}_t \right).$$

Notice that we have assumed that (1) does not depend on $x_t$, which yields:

$$\mathbb{E} \left( \xi_t | x_t, \mathcal{I}_{t-1} \right) = \mathbb{E} \left( \xi_t | \mathcal{I}_{t-1} \right) = \xi_{t|t-1}. \quad (4)$$

Also, from (1):

$$\xi_{t+1|t} = F \xi_{t|t}. \quad (5)$$

To link that with $\xi_{t|t-1}$, we note that the structure in (1) and (2) is linear, and recall the usual structure for updating a linear projection:

$$\xi_t = \xi_{t|t-1} + \mathbb{E} \left( (\xi_t - \xi_{t|t-1}) (y_t - y_{t|t-1})' \right) \times \left\{ \mathbb{E} \left[ (y_t - y_{t|t-1}) (y_t - y_{t|t-1})' \right] \right\}^{-1} (y_t - y_{t|t-1}). \quad (6)$$

Combining (1) and (6), and using standard formulas for recursive mean-squared error matrices, we obtain:

$$\xi_{t+1|t} = F \xi_{t|t-1} + F \mathbb{E} \left[ (\xi_t - \xi_{t|t-1}) (y_t - y_{t|t-1})' \right] \times \left\{ \mathbb{E} \left[ (y_t - y_{t|t-1}) (y_t - y_{t|t-1})' \right] \right\}^{-1} (y_t - y_{t|t-1}). \quad (7)$$

Consider now forecasting $y_t$ using $x_t$ and past information on $\mathcal{I}_{t-1}$:

$$y_{t|t-1} = \mathbb{E} \left( y_t | x_t, \mathcal{I}_{t-1} \right) = A'x_t + H'\xi_{t|t-1}. \quad (8)$$

Combining (8) and (9) we get:

$$\xi_{t+1|t} = F \xi_{t|t-1} + F P_{t|t-1} H \left( H' P_{t|t-1} H + R \right)^{-1} \times (y_t - A' x_t - H' \xi_{t|t-1}). \quad (9)$$

One can also show that there is a recursion for the variance-covariance matrix $P_{t+1|t}$:

$$P_{t+1|t} = F \left[ P_{t|t-1} - P_{t|t-1} H \left( H' P_{t|t-1} H + R \right)^{-1} H' P_{t|t-1} \right] F' + Q. \quad (10)$$

The recursive structure in (9), (10) and (11) allows us to state the main results in computing Kalman-filter estimates. First, start the recursion with the unconditional mean and variance-covariance matrix of $\xi_1$:

$$\xi_{1|0} = \mathbb{E} (\xi_1) \quad \text{and} \quad P_{1|0} = \mathbb{E} \left[ (\xi_1 - \mathbb{E} (\xi_1)) (\xi_1 - \mathbb{E} (\xi_1))' \right],$$

and then iterate on (11), (10) and (9) to obtain $P_{t+1|t}$, $\xi_{t+1|t}$ and $y_{t|t-1}$, respectively, for $t = 2, \cdots , T$. 

5
Suppose we are interested in estimating the unobserved state variable $-\xi_t$. There are two sets of forecasts commonly employed in a Kalman-filter setup: using the full set of observations available $(y_1, x_1, y_2, x_2, \cdots, y_T, x_T)$, which is called the *smoothed* estimate of $\xi_t$, or, we can forecast $\xi_t$ using only observations available up to period $t-1$, $(y_1, x_1, y_2, x_2, \cdots, y_{t-1}, x_{t-1})$, which is called the *filtered* estimate. Both are presented, respectively, below:

$$
\xi_{t|T} = \mathbb{E}\left(\xi_t | y_1, x_1, \cdots, y_T, x_T\right), \quad (12)
$$

$$
\xi_{t|t-1} = \mathbb{E}\left(\xi_t | y_1, x_1, \cdots, y_{t-1}, x_{t-1}\right). \quad (13)
$$

### 2.3 Estimation

Under the assumption that the errors $\left(\begin{array}{c} v_t \\ w_t \end{array}\right)$ have a multivariate Normal distribution, i.e., equation (3), consistent and fully-efficient estimation of the parameters in $F$, $A'$, and $H'$, and the two variance-covariance matrices $Q$ and $R$, can be performed by maximum likelihood. This assumption implies that the conditional distribution of $y_t | x_t, I_{t-1}$ is Gaussian as well, which can then be used to form the sample log-likelihood function with elements $\ln \left( f_t \right)$:

$$
\sum_{t=1}^{T} \ln \left( f_t \right) = \sum_{t=1}^{T} \ln \left[ f_{y_t|x_t,I_{t-1}} \left( \cdot \right) \right]
$$

$$
= \sum_{t=1}^{T} \left( -\frac{n}{2} \ln (2\pi) - \frac{1}{2} \ln \left| H'P_{t|t-1}H + R \right| - \frac{1}{2} \left( y_t - A'x_t - H'\xi_{t|t-1} \right)' \left( H'P_{t|t-1}H + R \right)^{-1} \left( y_t - A'x_t - H'\xi_{t|t-1} \right) \right)
$$

where $f_{y_t|x_t,I_{t-1}} \left( \cdot \right)$ denotes the (conditional) density function of $y_t | x_t, I_{t-1}$, and the expressions for $\xi_{t|t-1}$ and $P_{t|t-1}$ are given in (10) and (11), respectively.

### 2.4 Nowcasting and Forecasting

Suppose we possess observations ranging from $t = 1, 2, \cdots, T$, and estimate the parameters in $F$, $A'$, and $H'$, and the two variance-covariance matrices $Q$ and $R$, by maximizing the sample log-likelihood (14). As we argued before, there is some delay in making GDP readings available timely and continuously. Thus, when we have the reading for quarterly GDP on $T$ – comprising the months $T-2$, $T-1$, and $T$ – calendar time already ranges somewhere from $T+3$ through $T+6$, and we already possess coincident-series monthly observations somewhere from $T+1$ through $T+4$. So although we refer to this problem as *nowcasting* because of the status of calendar-time, from an econometric point-of-view this is indeed an out-of-sample *forecasting* problem with the added twist that we possess already the realizations of the series in $x_t$.

Our starting point is the *smoothed* state at the end of the sample: $\xi_{T|T}$. With that, using (5), we can then produce:

$$
\xi_{T+1|T} = F\xi_{T|T}. \quad (15)
$$
We are now in a position to forecast \( y_{T+1} \) using the available information on \( x_{T+1} \) and our previous forecast \( \xi_{T+1|T} \) given in (15). From (9):

\[
\mathbb{E}(y_{T+1}|x_{T+1}, \mathcal{I}_T) = A'x_{T+1} + H\xi_{T+1|T}.
\] (16)

This completes the one-step-ahead forecasts \( \mathbb{E}(y_{T+1}|x_{T+1}, \mathcal{I}_T) \) and \( \xi_{T+1|T} \). From then on, we will still have observed \( x \)'s to condition on, but our information on \( y \)'s will be kept at \( t = T \). The recursive structure in (9), (10) and (11) can then be used to produce nowcasts for \( T + 2 \):

\[
\mathbb{E}(y_{T+2}|x_{T+2}, \mathcal{I}_T) = A'x_{T+2} + H\xi_{T+2|T},
\] (17)

\[
\xi_{T+2|T+1} = F\xi_{T+1|T} + FP_{T+1|T}H'P_{T+1|T}H + R)^{-1} \times
\]

\[
[\mathbb{E}(y_{T+2}|x_{T+2}, \mathcal{I}_T) - A'x_{T+2} - H\xi_{T+2|T}],
\] (18)

where it should be noted that \( P_{T+1|T} \) can be obtained by using (11):

\[
P_{t+1|t} = F \left[ P_{t|t-1} - P_{t|t-1}H'P_{t|t-1}H + R \right]^{-1} H'P_{t|t-1} \right] F' + Q,
\] (19)

which shows that, given the end-of-sample estimate \( P_{T|T-1} \), one obtains immediately \( P_{T+1|T} \) since we possess estimates of \( F, H', Q \) and \( R \). The procedure for \( T + h, h = 3, 4, \ldots, \) etc., can replicate that of \( T + 2 \), as long as we possess observations on \( x \) up until \( T + h \), where \( h \) is the number of additional observations we have on \( x \) vis-a-vis monthly GDP. This completes the nowcasting problem.

The truly out-of-sample problem starts when we do not possess anymore observations on the \( x \)'s, which occurs after \( T + h \). Here, our proposed strategy is to keep the same state-space representation given in the previous section, switching only the series in \( x \): in the nowcasting exercise these essentially are coincident series of economic activity. In the out-of-sample exercise, these will be leading series of economic activity, i.e., series that lead economic activity and therefore are useful for forecasting.

### 2.5 The Encompassing Model for Interpolation

First, a word of caution about using the term interpolation here. GDP is a flow variable for which we possess quarterly observations we want to distribute within the months in the quarter. Stock and Watson (2010) cite Harvey (1989) to note that the problem of allocating a quarterly flow as such is referred to as distribution, whereas interpolation estimates monthly values of stock variables from quarterly values. Despite this technical distinction for flow and stock variables, several authors still refer to interpolated GDP, a term that is now ingrained in the literature, being the reason why we employ it here.

Bernanke, Gertler and Watson (1997) proposed a general state-space model that encompassed several competing models used in the literature for interpolation: Chow and Lin (1971), Fernandez (1981), and Mitchell and Jones (2005), for example. They assume that unobserved monthly GDP (labelled as \( y^* \) here) follows an AR(p) process explained by regressors \( x_t \) and an AR(1) error term. The variables in \( x_t \) are co-variates (pre-determined or exogenous), which should have a high correlation with the series being interpolated: much
of the contemporaneous behavior of the interpolated series comes from them. Also in $x_t$ are deterministic series such as a constant and/or seasonal dummies. The model proposed for unobserved monthly GDP $y_t^+$ is:

$$(1 - \phi_1 L - \cdots - \phi_p L^p) y_t^+ = x_t \beta + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t. \quad (20)$$

Observed quarterly GDP (labelled as $y_t$ here) is the variable being interpolated from quarterly to monthly frequency. It relates to the interpoland series $y_{t}^{+}$ and obeys:

$$y_{t} = \sum_{i=0}^{2} y_{t-i}^{+}, \quad t = 3, 6, 9, 12, \ldots T \quad (21)$$

$$y_{t} = 0, \quad \text{otherwise}. \quad (22)$$

Hence, quarterly GDP, which we can only observe on months $t = 3, 6, 9, 12, \ldots T$, is the sum of the corresponding monthly GDPs in that quarter\(^2\). Otherwise, it is just set to a fictional value of zero. Notice that setting $y_{t} = 0$ for the months we do not observe GDP is a way of making quarterly GDP observable at the monthly frequency; see Mönch and Uhlig (2005), Appendix, just above equation (1). In the Kalman-filter literature for mixed frequency models (see, for example, Giannone, Reichlin and Small, 2008) a fictional value is usually assumed for missing observations (zero is the most frequent choice). The crucial step is to impose that the fictional observed data has a very large variance, so that the zero value is discounted and overwritten by the Kalman-filter technique. This is exactly how Bernanke, Gertler and Watson and Mönch and Uhlig proceed.

A second issue is how to initialize the values of $\beta$, $\rho$, and the variance of $u_t$ in the Kalman-filter procedure. Mönch and Uhlig aggregate the covariates in $x_t$ from high (monthly) to low (quarterly) frequency, and run an OLS regression of $y_t$ on its lags and $x_t$ (regression (20)) at the quarterly frequency. This yields estimates of $\beta$ and the $\phi$’s, as well as an estimate for the variance of $u_t$. With the latter, an estimate of $\rho$ is obtained from an OLS regression of $u_t$ on $u_{t-1}$.

If we assume that the polynomial $(1 - \phi_1 L - \cdots - \phi_p L^p)$ is of order one, i.e., $p = 1$, with coefficient $\phi$, our state-space form is the following:

$$\xi_t = \left(\begin{array}{c}
y_t^+ \\
y_{t-1}^+ \\
y_{t-2}^+ \\
u_t
\end{array}\right) = \left(\begin{array}{cccc}
\phi & 0 & 0 & \rho \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & \rho
\end{array}\right) \left(\begin{array}{c}
y_{t-1}^+ \\
y_{t-2}^+ \\
y_{t-3}^+ \\
u_{t-1}
\end{array}\right) + \left(\begin{array}{c}
x_t \beta \\
0 \\
0 \\
0
\end{array}\right) + \left(\begin{array}{c}
\varepsilon_t \\
0 \\
0 \\
\varepsilon_t
\end{array}\right) \quad (23)$$

$$y_t = H_t^\xi, \quad (24)$$

\(^2\)Note that the aggregation of monthly GDP can also be made averaging the $y_{t}^{+}$’s, i.e., as $y_t = \frac{1}{3} \sum_{i=0}^{2} y_{t-i}^{+}$. Regardless of how this is done, notice that, up until $T$, all estimates of monthly unobserved GDP are obtained with the restriction that they add up to quarterly observed GDP data. However, although the parameter estimates embody this restriction, it will not hold for the nowcasting and forecasting period, since we do not possess yet quarterly GDP for this period.
where (23) and (24) are respectively the state and the observation equations and the matrix \( H_t \) is time-varying, with the following format:

\[
H_t = \begin{cases} 
1 & 1 & 0 \\
0 & 0 & 0 
\end{cases}
, \quad t = 3, 6, 9, 12, \ldots T
\]

(25)

As mentioned before, one interesting characteristic of using the Bernanke, Gertler and Watson (1997) approach is that the model described in (23) and (24) encompasses several data interpolation models. They are summarized in Table 1 below. Also, the Appendix contains some discussion of their features and properties:

| Table 1 – Resulting Model as a Function of \( \phi \) and \( \rho \) in (23) |
|-----------------|---|---|
| Model | \( \phi \) | \( \rho \) |
| \( M_1 \): Static model in levels with IID residuals | 0 | 0 |
| \( M_2 \): Static model in levels with AR(1) residuals (Chow and Lin, 1971) | 0 | free |
| \( M_3 \): Static model in 1st differences with IID residuals (Fernandez, 1981) | 0 | 1 |
| \( M_4 \): Dynamic model in levels with IID residuals (Mitchell et al., 2005) | free | 0 |
| \( M_5 \): Dynamic model in 1st differences with IID residuals | free | 1 |
| \( M_6 \): Dynamic model in levels with AR(1) residuals | free | free |

2.6 Goodness of Fit Statistics for Interpolated Models

To assess the quality of interpolation, Bernanke, Gertler, and Watson (1997) propose the use two \( R^2 \) measures of fit. Denoting by \( \hat{y}_{t|T} \) the smoothed estimate of monthly GDP, and by \( \hat{u}_{t|T} \) the same estimate of the error term \( u_t \), they consider:

\[
R^2_{\text{level}} = \frac{\text{VAR}(\hat{y}_{t|T})}{\text{VAR}(\hat{y}_{t|T}) + \text{VAR}(\hat{u}_{t|T})}, \quad \text{and,}
\]

\[
R^2_{\text{diff}} = \frac{\text{VAR}(\Delta \hat{y}_{t|T})}{\text{VAR}(\Delta \hat{y}_{t|T}) + \text{VAR}(\Delta \hat{u}_{t|T})}.
\]

Bernanke, Gertler, and Watson claim that it is more informative to report the \( R^2 \) in first differences since the same statistic in levels will always be close to unity. Thus, we will compute all models listed in Table 1 and compare them regarding their \( R^2 \)s, picking the one with the best fit.

Given the best models found in empirical tests, all listed in Table 1 above, we can check which of their properties best fit the data: does monthly GDP \( (y_t) \) follows an autoregressive (AR) process? What does the structure of the monthly residuals look like? In our interpolation exercise, we estimate the six models summarized in Table 1 via Maximum Likelihood using the Kalman filter. For each model, the smoothed estimate of \( \xi_t \) serves as our final interpolated monthly GDP sequence.
3 Empirical Results

3.1 Data

One of the most important factors in the interpolation procedure adopted in this paper is the signal extraction from related series $x_t$. They represent the main information source for interpolation and must fulfill two requirements:

1. They have to be available in the desired higher frequency of interpolated GDP (monthly in our case) in a timely fashion.

2. They need to have a high correlation with GDP.

The first condition imposes a strong restriction for Brazilian data, since there are only a few time-series that cover the entire estimation period (1980-2012) and are available on a monthly frequency. Following Mariano and Murasawa (2003), natural candidates for series in $x_t$ would be the coincident series used in business-cycle analysis – Industrial Production, Sales, Income, and Employment. Unfortunately, Income and Employment are only available in Brazil after 2002, since a major change in the Monthly Employment Survey (Pesquisa Mensal do Emprego) made it virtually impossible to chain pre- and post-2002 data. Issler, Notini and Rodrigues (2013) solved this problem by proposing a back-cast algorithm similar to the interpolation method in Bernanke, Gertler and Watson (1997), where pre-2002 data is back-cast on the basis of long-span covariates which are highly correlated with the back-cast series.

Our first step is to get as many series as possible that satisfy the first condition, later checking which of them satisfy the second condition. As a starting point, we considered the first two series that represent monthly economic activity – Industrial Production, labelled $IP$, and Sales, labelled $Sales$. For these variables we have long-span data, and they are believed to have cycles that are concurrent with the latent “business cycles.” Since Stock and Watson (1999) argue that, if we were to choose one variable to best represent the state of business cycles, this variable would be the GDP, a natural empirical strategy to interpolate GDP would be to use not only the coincident series used in business-cycle analysis but also additional coincident series as well. The latter can be verified by standard methods – e.g., the Bry and Boschan (1971) algorithm. In order to increase the number of coincident series, we follow Cardoso (1981), from which we have selected additional four candidate series: Energy Demand – labelled $Energy$, Steel Production – labelled $Steel$, Cement Production – labelled $Cement$, and Vehicle Production – labelled $Vehicles$.

Our next step was to test if the potential candidates satisfy some desired statistics requirements (second condition). First, we test if the variables have the same integration order of GDP. Results are summarized in Table 2.
Table 2 - ADF Unit Root Test

<table>
<thead>
<tr>
<th>Variables</th>
<th>( H_0: z \text{ is } I(1) )</th>
<th>( H_0: z \text{ is } I(2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP*</td>
<td>-0.898</td>
<td>-4.098</td>
</tr>
<tr>
<td></td>
<td>[0.95]</td>
<td>[0.01]</td>
</tr>
<tr>
<td>Steel</td>
<td>-0.792</td>
<td>-5.514</td>
</tr>
<tr>
<td></td>
<td>[0.82]</td>
<td>[0.00]</td>
</tr>
<tr>
<td>Vehicles</td>
<td>0.522</td>
<td>-6.208</td>
</tr>
<tr>
<td></td>
<td>[0.99]</td>
<td>[0.00]</td>
</tr>
<tr>
<td>Cement</td>
<td>-0.220</td>
<td>-3.392</td>
</tr>
<tr>
<td></td>
<td>[0.93]</td>
<td>[0.01]</td>
</tr>
<tr>
<td>Energy</td>
<td>-0.208</td>
<td>-26.672</td>
</tr>
<tr>
<td></td>
<td>[0.93]</td>
<td>[0.00]</td>
</tr>
<tr>
<td>IP</td>
<td>0.690</td>
<td>-6.270</td>
</tr>
<tr>
<td></td>
<td>[0.99]</td>
<td>[0.00]</td>
</tr>
<tr>
<td>Sales</td>
<td>-0.118</td>
<td>-6.156</td>
</tr>
<tr>
<td></td>
<td>[0.95]</td>
<td>[0.00]</td>
</tr>
</tbody>
</table>

Notes: (i) All series are in logs. The specification of the test equation was chosen on the basis of the Schwarz Information Criterion; (ii) the asterisk (*) indicates that a linear trend was included in the test equation; (iii) figures in brackets are p-values.

Based on the results of the Augmented Dickey-Fuller (ADF) test in Table 2, we do not reject the null hypothesis that the series are \( I(1) \), which makes us conclude that all analyzed time series have the same integration order of GDP. Next, we apply first differences to all these series and compute the correlation between each first-differenced series and first-differenced GDP, which is reported in Table 3. All the analyzed series show a high correlation with the GDP series once both are first-differenced, although the correlation of GDP with Industrial Production and Sales is higher than that of other series. This concludes the cyclical analysis of the auxiliary series.

Table 3 - Correlation Coefficients – GDP and Coincident Series

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cement</th>
<th>Vehicles</th>
<th>Steel</th>
<th>Energy</th>
<th>IP</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>0.49</td>
<td>0.33</td>
<td>0.29</td>
<td>0.21</td>
<td>0.66</td>
<td>0.57</td>
</tr>
</tbody>
</table>

We also investigate the trend pattern between each candidate auxiliary series and GDP. First, we run Johansen’s (1991) cointegration test in order to verify whether there is a long-run relationship between GDP and candidate auxiliary series. As expected, we found that all tested series have one common trend with GDP.

Based on the series cycles high correlation and the existence of a cointegration relationship, we select Industrial Production and Sales, and possibly Cement Production as well. It

---

3 We have to aggregate the monthly series in order to compute correlation coefficients. We also compute correlations using seasonally adjusted (X12) series. The results are very similar and are available upon request.

4 We also compute the correlation coefficient between each candidate series cycles (extracted by Hodrick-Prescott filter) and GDP cycle. Results are very similar to those in Table 3.

5 Results are available upon request.
is interesting to note that the traditionally Coincident Series used in business-cycles analysis are the best choices to our interpolation method what confirms their capacity to track very well the economic activity.

3.2 Evaluation Results

In this section we present the results of the Kalman-filter estimation for each model considered. Before estimation, all series have been seasonally adjusted using the X12 procedure. In Table 4, we present the estimated parameters and coefficients of the auxiliary variables.

Table 4 - Interpolation Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$M_1$</th>
<th>$M_2$</th>
<th>$M_3$</th>
<th>$M_4$</th>
<th>$M_5$</th>
<th>$M_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>0.94</td>
<td>-0.67</td>
<td>-0.60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>11.03</td>
<td>33.51</td>
<td>51.60</td>
<td>-0.029</td>
<td>62.01</td>
<td>45.77</td>
</tr>
<tr>
<td>I. Production</td>
<td>0.693</td>
<td>0.198</td>
<td>0.212</td>
<td>0.0480</td>
<td>0.349</td>
<td>0.36</td>
</tr>
<tr>
<td>Sales</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\rho$</td>
<td>1.0041</td>
<td>1.0045</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>6.5881</td>
<td>0.8955</td>
<td>0.9439</td>
<td>1.1463</td>
<td>1.2570</td>
<td>1.3259</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>205.6269</td>
<td>88.3283</td>
<td>93.7001</td>
<td>109.0752</td>
<td>84.9365</td>
<td>92.7201</td>
</tr>
</tbody>
</table>

Notes: (1) The models are described in the text and in the Appendix. (2) Industrial Production and Sales are used as auxiliary series $x_t$.

From the results above we can see that the auxiliary coefficients and the estimated parameters change according to the model being estimated. In order to choose the model that best describe our GDP series, we make use of the $R^2$ measures of fit to gauge the accuracy of interpolation. In Table 5, we present for each model, the $R^2$ measures of fit in first difference for the filtered and smoothed GDP.

Table 5 - $R^2_{\text{diffs}}$ Measures of Fit

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2_{\text{diffs}}$ (GDP filtered)</th>
<th>$R^2_{\text{diffs}}$ (GDP smoothed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>0.0273</td>
<td>0.0346</td>
</tr>
<tr>
<td>$M_2$</td>
<td>0.4652</td>
<td>0.8097</td>
</tr>
<tr>
<td>$M_3$</td>
<td>0.4382</td>
<td>0.7923</td>
</tr>
<tr>
<td>$M_4$</td>
<td>0.3609</td>
<td>0.7098</td>
</tr>
<tr>
<td>$M_5$</td>
<td>0.4011</td>
<td>0.5806</td>
</tr>
<tr>
<td>$M_6$</td>
<td>0.3766</td>
<td>0.5736</td>
</tr>
</tbody>
</table>

With the exception of model $M_1$, all models have a high $R^2$, so they show a reasonable interpolation accuracy. The best $R^2$ measure is obtained for Model $M_2$, followed by model $M_3$. Both impose that $\rho$ is zero in (23). They only differ about the stochastic process of the error term: model $M_2$ imposes that the error is an $AR(1)$ process, whereas model $M_3$ imposes that it is an $i.i.d.$ process.

Model $M_2$ is a contemporaneously static model with an $AR(1)$ error term. Taking first quasi-difference of this model yields a restricted dynamic structure with one lag of dependent
explanatory variables – the so called *common-factor* model:

\[
(1 - \rho L) y_t^+ = (1 - \rho L) x_t \beta + (1 - \rho L) u_t \\
y_t^+ = \rho y_{t-1}^+ + x_t \beta - \rho \beta x_{t-1} + \varepsilon_t,
\]

where \( \varepsilon_t \) is white noise and there is an embedded restriction in the dynamic multipliers due to the existence of a common factor. Figure 1 depicts monthly and quarterly GDP, the former being estimated by Model \( M_2 \). The first graph shows the complete sample (1980-2012) while in the second one just the recent period (1994-2012). Both show a smooth interpolated series, which fits closely to the observed quarterly series. Next, we investigate how well interpolated GDP behaves as compared to alternative GDP proxies available in Brazil.

![Figure 1 - Monthly (Interpolated) and Quarterly (Observed) GDP](image)

**3.3 Comparing and evaluating the monthly GDP with other economic activity’s proxies**

Monthly GDP estimated in last section could be widely used by practitioners and academics alike interested in measuring in real time the state of the economic activity. It also could be used to detect turning points in the Brazilian economy. In view of that, this section compares our monthly interpolated GDP with an alternative proxy of the economic activity recently constructed by the Central Bank of Brazil - the Brazilian Economic Activity Index - (IBC-Br); see Central Bank of Brazil (2010), for example.

IBC-Br is computed by aggregating monthly time-series from economy-wide supply-side sectors: agriculture, industrial sector and services. In agriculture, the source of information is the Systematic Survey of Agricultural Production released by IBGE. Regarding the industrial sector, IBC-Br monthly index is obtained by averaging the indices for four sub-sectors in our National Accounts, which are available on a monthly basis. They are weighted by value added.

---

6 A detailed methodological description and the complete list of component time series could be found at Brazil’s Central Bank homepage: [www.bc.gov.br/](http://www.bc.gov.br/)
at basic prices of the Quarterly National Accounts System in the previous year. Finally, the services sector includes time series from the trade activities, transportation, storage and mail services, financial intermediation, insurance, pension funds and related services, real estate and rentals, management, public health and education and social security, and other services.

The first advantage of our monthly GDP over IBC-Br is computational. While putting together IBC-Br requires the aggregation of hundreds of monthly series, our interpolation method just requires the use of two covariate series in estimation. Second, IBC-Br methodology does not impose any restriction (discipline) between estimated levels vis-a-vis that of GDP or between estimated growth rates vis-a-vis that of GDP. In technical language, IBC-Br and GDP need not cointegrate and the cycles in IBC-Br and GDP growth rates need not be synchronized at the quarterly frequency. On the other hand, since our interpolated monthly GDP adds up to quarterly GDP for any given quarter, it must display the same short- and long-run behavior of observed GDP by construction, which is the great advantage of imposing discipline at the estimation level. Finally, because our estimate is based on a state-space representation, it can be immediately used to nowcast and forecast GDP\(^7\), as discussed in the previous section.

With these points in mind, we now compare these two economic activity proxies. Figure 2 describes both series in levels while in Figure 3 they are depicted in first differences. As can be seen, for the period 2005-12, the two series show a very similar trend pattern. However, this is not the case for their growth rates. For the latter, we time-aggregate both proxies of GDP at the quarterly frequency, and then compute their respective quarterly growth rates. These results are plotted in Figure 4. As expected, our estimated GDP proxy growth rate tracks that of quarterly GDP perfectly, which is not true regarding IBC-Br. We observe some larger discrepancies, especially in 2003-05 and 2007-08.

\[\text{Figure 2 - IBC-Br and Kalman-Filter Monthly GDP (levels)}\]

\(^7\)We should also mention that the IBC-Br series only begins in 2003, which prevents its use in business-cycle analysis, which requires long-span data.
3.4 Detecting Business Cycles Turning Points

In this section, we first compute the chronology of business cycles (recessions vs. expansions) that is consistent with our monthly interpolated GDP, later comparing it with alternative business-cycle dating’s, including that of the official Brazilian Business-Cycle Dating Committee (CODACE). Turning points are determined using the Bry and Boschan (1971) algorithm for detecting local minima and maxima of a time series. The dated turning points of Brazilian economic activity according to our interpolated GDP are presented below in Figure
5. In Table 6 we make turning-point comparisons regarding alternative datings of Brazilian economic activity.

![Figure 5 - Monthly GDP - shaded Bry Boshan recessions periods](image)

Results in Table 6 show that the dating using our GDP monthly indicator yields results close to that of Duarte, Issler and Spacov (2004), although the latter does not cover such a wide time span as our GDP proxy. Our dating is also close to that of CODACE: on that regard, the most striking differences appear in the dating of the 1987-88 recession and of the 2002-03 recession. Our GDP dating misses these two episodes, while it keeps a very good record vis-a-vis CODACE’s datings, if we take into account the fact that the Brazilian Business-Cycle Dating Committee seems to have merged the 1980-81 recession with that of 1982-83.
Finally, we attempt here a historical account of Brazilian recessions according to our monthly GDP proxy. The recession in 1980-81 can be linked to the increase in interest rates by the FED in the early 1980s, which was later responsible for the emerging-market debt crisis (mostly in Latin America). The 1982-83 recession is related to the Latin American debt crisis itself, where international credit to these economies came to a halt after the Mexican moratorium in 1982. In these two recessions, there was an external factor at work.

Our next dated recession (1989-1991) was “home made,” which came as the result of the Brazilian government inability to curb high inflation with many unsuccessful economic plans. These all had in common the absence of long-term fiscal discipline, with an initial sudden transitory contraction of money supply.

After the Real Plan, in July 1994, inflation finally gets fairly under control and most recessions are again related to events abroad, which generated capital flight and thus prompted a sudden rise in domestic interest rates as a reaction to keep foreign capital in domestic markets. The Mexican crisis in 1994 affected Brazil and other emerging markets. In 1997-98, the Asian and the Russian crises had similar effects here and in other emerging markets. On these occasions domestic interest rates had risen to very high levels, leading to a reduction in domestic economic activity. In 2000-12 our local energy-supply crisis – the rationing of electrical energy for consumers and industry – was responsible to an economy-wide contraction of economic activity, while in 2008-09 we had the global financial crisis with a short and limited contraction of Brazilian economic activity.

### 3.5 Nowcasting GDP

In this section we present evidence that one can indeed successfully nowcast Brazilian GDP using our interpolated monthly GDP. Here, we reproduce the results of the exercise in Notini, Issler, Rodrigues, Matos, and Bonelli (2012). Instead of resorting to the techniques outlined in Section 2.4, which rely on kalman filtering, they performed a much simpler exercise, where the conditional nowcasting model is estimated by OLS and current observations of the covariates
are used to nowcast GDP every month up to one quarter ahead. This quarterly forecast is then compared with the quarterly estimate of IBC-Br – the monthly GDP series put forth by the Central Bank of Brazil (2010) – in terms of its accuracy in forecasting actual quarterly GDP.

In their exercise, Notini et al. (2012) set the forecasting horizon to be one quarter. Hence, the GDP model forecasts three months ahead in order to nowcast quarterly GDP. Figure 6 contains the results for this exercise, where the green line is the resulting error series transformed to growth rates. As can be seen below, apart from the beginning of 2009, the monthly GDP model nowcasts GDP with high accuracy.

![Figure 6 - Nowcast Errors of the Monthly GDP Model](image)

Next, we perform an out-of-sample forecast comparison between the monthly GDP model and IBC-Br. Notice that the latter is widely used by the banking industry to nowcast Brazilian GDP. It is important to note that there is no informational gap between our GDP nowcast and that of IBC-Br. So this is a fair nowcasting competition. Notini et al. compute the one-quarter ahead Mean Square Error (MSE) to assess the forecast accuracy of both forecasts. This is computed for two different periods: 2007:4 to 2011:4 and 2010:4 to 2011:4. Table 6 summarizes the results.

### Table 7 - Nowcast Mean Squared Error

<table>
<thead>
<tr>
<th>Nowcast Period</th>
<th>IBC-Br</th>
<th>Monthly GDP Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007.4 - 2011.4</td>
<td>0.46%</td>
<td>0.40%</td>
</tr>
<tr>
<td>2010.4 - 2011.4</td>
<td>0.35%</td>
<td>0.16%</td>
</tr>
</tbody>
</table>

As can be seen from Table 7, the monthly GDP model outperforms IBC-Br, and its superior performance is even stronger in the recent past, showing its usefulness in nowcasting Brazilian GDP. Furthermore, we should expect the forecasting performance of the monthly GDP model to improve if we use kalman filtered coefficient estimates in nowcasting GDP instead of using OLS estimates.
4 Conclusions

In this paper, we propose an estimate for real monthly GDP in Brazil for the period 1980-2012, which is an interpolation of the quarterly observed series; see Bernanke, Gertler and Watson (1997) and Mönch and Uhlig (2005). Our monthly GDP proxy is based on a state-space representation which imposes the restriction (discipline) that the monthly proxy adds up to quarterly observed GDP within every single quarter. A key issue is the choice of covariates used in interpolating the quarterly series: we chose to employ Industrial Production and Sales (corrugated paper sales) as covariates in interpolation, since these two series are key in Brazilian business-cycle analysis; see Duarte, Issler and Spacov (2004) and Issler, Notini and Rodrigues (2012).

The methodology in Bernanke, Gertler and Watson, and Mönch and Uhlig, allows estimating six different interpolation models, some of which have a long tradition in the literature. We are able to assess the fit of different models and related series in order to get the most appropriate monthly real GDP estimate: we evaluate six competing models, and six competing coincident series (Energy Demand, Steel Production, Cement Production, Vehicles Production, Industrial production and Sales), all available at the monthly frequency for the period 1980-2012.

First, we identify the series which behavior is closer to that of GDP, selecting Industrial Production and Sales. Second, we select the state-space model that has the highest goodness-of-fit statistic vis-a-vis GDP. We chose a contemporaneously static model with an $AR(1)$ error term – which implies a restricted dynamic structure with one lag of dependent and explanatory variables. This best model is then estimated using Industrial Production and Sales as covariates yielding a smoothed estimate of Brazilian monthly GDP which serves as our GDP monthly proxy.

Next, we perform several interesting empirical exercises: we compare our GDP monthly proxy with alternative proxies available for Brazil. Our main comparison is regarding IBC-Br – the monthly GDP proxy made available by the Central Bank of Brazil. We discuss the advantages of our approach vis-a-vis theirs; we (re)establish a chronology of recessions in the recent past of the Brazilian economy using our GDP proxy. Its dating of recessions and expansions is compared with those of Duarte, Issler and Spacov (2004) and to those of the Brazilian Business-Cycle Dating Committee (CODACE), showing a close enough dating; we also present the results of an out-of-sample nowcasting exercise using the monthly GDP model. Finally, its forecast accuracy in forecasting actual quarterly GDP is compared with that of the quarterly estimate of IBC-Br. From 2007:4 to 2011:4, its mean-squared error is 13% smaller than that of IBC-Br, whereas, for the recent past, 2010:4 to 2011:4, it is more than 50% smaller, showing its usefulness in nowcasting Brazilian GDP.

References


A Appendix

A.1 Detailed Interpolation Models

The literature appoint as the main problem of the models 1a and 1b is the fact that they extract signals only from the presumed stochastic process of the original series in a way that no new information is added. We believe that we would be better off enriching the model with additional information contained in the related series.

Model M1

This is the simpler model which incorporates information from related series. So it can be used as a benchmark when we estimate the more sophisticated models. It can be stated as:

\[ y_t = x_t' \beta + \epsilon_t \]

where: \( x_t' \) is a vector of related series and \( \epsilon_t \) a iid error term with distribution \( N(0, \sigma^2) \).

Model M2

This is the Chow and Lin (1971) model. This model is very used in the literature because they are the first to show how to include related series in the interpolation procedure. They suggest a regression model without lagged dependent variables, but autoregressive errors, so we can obtain the Chow and Lin (1971) model by fixing \( \phi = 0 \) and by letting \( \rho \) to be estimated freely:

\[
\begin{pmatrix}
    y_t \\
    y_{t-1} \\
    y_{t-2} \\
    u_t
\end{pmatrix}
= \begin{pmatrix}
    0 & 0 & 0 & \rho \\
    1 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0 \\
    0 & 0 & 0 & \rho
\end{pmatrix}
\begin{pmatrix}
    y_{t-1} \\
    y_{t-2} \\
    y_{t-3} \\
    u_{t-1}
\end{pmatrix}
+ \begin{pmatrix}
    x_t' \beta \\
    0 \\
    0 \\
    0
\end{pmatrix}
+ \begin{pmatrix}
    \epsilon_t \\
    0 \\
    0 \\
    \epsilon_t
\end{pmatrix}

\]

It is important to note that in their seminal paper, Chow and Lin do not use the Kalman filter but it can be shown that the Kalman filter above and the Chow and Lin regression yield the same estimates by maximum likelihood (Cuche and Hess).

Model M3

This is a variation of the Chow and Lin model, suggested by Fernandez (1981), where instead of the regression of \( y_t \) in levels, it uses the first differences of \( y_t \) in order to account
for non-stationarity. This model is obtained by letting regression residuals to be a random walk, i.e. $\phi = 0$ and $\rho = 1$.

**Model M$_4$**

This model was suggested by Mitchell and Jones (2005) and it is a dynamic version of model M$_1$.

**Model M$_5$**

A modified version of model M4 where we impose that $\rho = 1$.

**Model M$_6$**

It is the most general version of the Mitchell and Jones (2005) model, where we let $\phi$ and $\rho$ to be estimated freely.

In our interpolation exercise, we estimate the six models summarized above via Maximum Likelihood. We then perform a set of bilateral likelihood ratio tests to discover whether the imposed restrictions are borne by the data or not and select the most appropriate model accordingly.

### A.2 Data and Sources

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Source (IPEADATA)</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>quarterly</td>
<td>IBGE</td>
<td>Índice Geral de Preços</td>
</tr>
<tr>
<td>IGP-DI</td>
<td>monthly</td>
<td>FGV</td>
<td></td>
</tr>
<tr>
<td>Industrial Production</td>
<td>monthly</td>
<td>IBGE/PIM-PF</td>
<td>Produção Industrial - Indústria Geral</td>
</tr>
<tr>
<td>Sales</td>
<td>monthly</td>
<td>ABPO</td>
<td>Expedição de caixas, acessórios e chapas - papelão ondulado</td>
</tr>
<tr>
<td>Energy Demand</td>
<td>monthly</td>
<td>Eletrobrás</td>
<td>Consumo - energia elétrica</td>
</tr>
<tr>
<td>Cement</td>
<td>monthly</td>
<td>SNIC</td>
<td>Produção - Cimento</td>
</tr>
<tr>
<td>Vehicles</td>
<td>monthly</td>
<td>ANFAVEA</td>
<td>Produção - Automóveis</td>
</tr>
<tr>
<td>Steel Production</td>
<td>monthly</td>
<td>IBS/IE</td>
<td>Produção- Aço bruto</td>
</tr>
</tbody>
</table>