rbe

Estimating Brazilian Monthly GDP: a State-Space Approach[*]

João Victor Issler,[†] Hilton Hostalacio Notini[‡]

Contents: 1. Introduction; 2. State-Space Model and the Kalman Filter for Interpolation; 3. Empirical Results; 4. Conclusions; Appendix. Detailed Interpolation Models.

This paper has several contributions. The first 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, & Watson (1997) [*Systematic monetary policy and the effects of oil price shocks* (Brookings Papers in Economic Activity No.1)]. Second, along the spirit of Mariano & Murasawa (2003) [A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4), 427–443], we 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. Third, in a nowcasting and forecasting exercise, we illustrate the advantages of our integrated approach. Finally, we compare the chronology of recessions of our monthly estimate with those done elsewhere.

Esse artigo tem três contribuições originais. A primeira é o emprego de uma técnica superior de interpolação para o PIB brasileiro em bases trimestrais para a frequência mensal, método este que auxilia no nowcast e na previsão fora da amostra do crescimento do PIB de forma integrada Bernanke, Gertler, & Watson (1997) [Systematic

[†]Corresponding author: Escola Brasileira de Economia e Finanças / Fundação Getúlio Vargas (EPGE/FGV). Praia de Botafogo 190, sala 1100, Rio de Janeiro – RJ, Brasil. CEP 22250-900.E-mail: joao.issler@fgv.br

[‡]Finance Manager at Itaú-Unibanco Holding SA.E-mail: hilton@fgvmail.br

monetary policy and the effects of oil price shocks *(Brookings Papers in Economic Activity No.1)]. A segunda, no espírito de* Mariano & Murasawa (2003) *[A new coincident index of business cycles based on monthly and quarterly series.* Journal of Applied Econometrics, 18(4), 427–443], *é a de propor e testar uma miríade de modelos de interpolação e de séries auxiliares para interpolar o PIB, sendo as últimas todas coincidentes ao PIB do ponto de vista dos ciclos de negócios da economia brasileira. Baseados nesses resultados, escolhemos o melhor modelo para estimar o PIB brasileiro em bases mensais. A terceira é a comparação entre as dentro e fora da amostra entre a nossa série de PIB mensal e um indicador do PIB mensal popularmente utilizado pelo mercado financeiro — o Índice de Atividade Econômica do Banco Central (IBC-Br). Nessas comparações, notamos que o nosso indicador do PIB mensal não só segue mais de perto as variações do PIB dentro da amostra, como também o prevê de forma mais precisa fora da amostra. O primeiro resultado é esperado, pois impomos a disciplina de que as variações trimestrais do PIB mensal devam ser iguais às do PIB trimestral. Porém, impor essa disciplina ao PIB interpolado também melhora a previsão fora da amostra do modelo que o contém, algo que a técnica de interpolação em si não poderia garantir, mas que observamos na prática. Esse resultado aponta na direção de que talvez a construção da série mensal do IBC-Br não seja muito eficiente, pois requer o armazenamento e acompanhamento de centenas de séries mensais, enquanto a construção do nosso PIB mensal requer apenas o armazenamento e acompanhamento de duas séries auxiliares.*

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 & 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 & 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 & 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.
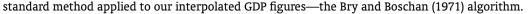
As argued above, this strategy can improve our knowledge of GDP in two dimensions. The first is that we will possess more frequent business-cycle estimates, i.e., monthly instead of quarterly. This solves an interpolation problem that was proposed by Bernanke, Gertler, & Watson (1997) with an econometric model that encompassed most interpolation models that had been previously proposed in the literature; see also the extension in Mönch & 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 & 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 et al. (1997) and in Mönch & 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 & 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.

In the business-cycle dimension, the ideas and models discussed in this paper are also related to the work of Stock & Watson (1988, 1991, 1993a, 1993b, 2002a, 2002b), Chauvet (1998), Kim & Nelson (1999), Cuche & Hess (2000), Liu & Hall (2001), and Mariano & Murasawa (2003). Regarding the literature on Brazilian business cycles, our paper is related to the work of Contador (1977), Cardoso (1981), Contador & Santos Filho (1987), Nakane (1994), Contador & Ferraz (1999), Spacov (2001), Issler & Spacov (2000), Chauvet (1998, 2001, 2002), Picchetti & Toledo (2002), Duarte, Issler, & Spacov (2004), Hollauer, Issler, & Notini (2009), BCB/Copom (2010), Issler, Notini, & Rodrigues (2013), Issler, Notini, & Rodrigues (2013) and Issler, Notini, Rodrigues, & Soares (2013).

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 present our main empirical results including a dating chronology for Brazil business cycles. Section 4 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 $r \times 1$ state vector of possibly unobserved series ξ_t we want to estimate. The second vector equation represents the *observation or measurement equation* linking the state vector to the $n \times 1$ 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:

$$\xi_{t+1} = F\xi_t + v_{t+1}, \tag{1}$$

$$y_t = A'x_t + H'\xi_t + w_t, \tag{2}$$

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:

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} \right). \tag{3}$$

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.

2.2. Filtering and Smoothing

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

$$\xi_{t+1|t} = \mathbb{E}\left[\xi_{t+1} \mid \mathcal{I}_t \right].$$

---

[1]Multivariate white-noise processes.

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

$$\mathbb{E}\left[\xi_t|x_t,\mathcal{I}_{t-1}\right] = \mathbb{E}\left[\xi_t|\mathcal{I}_{t-1}\right] = \xi_{t|t-1}. \tag{4}$$

Also, from (1),

$$\xi_{t+1|t} = F\xi_{t|t}. \tag{5}$$

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 projections:

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

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

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

$$= F\xi_{t|t-1} + FP_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1} \times \left(y_t - y_{t|t-1}\right). \tag{8}$$

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

$$y_{t|t-1} = \mathbb{E}\left[y_t|x_t,\mathcal{I}_{t-1}\right] = A'x_t + H'\xi_{t|t-1}. \tag{9}$$

Combining (8) and (9) we get

$$\xi_{t+1|t} = F\xi_{t|t-1} + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1} \times (y_t - A'x_t - H'\xi_{t|t-1}). \tag{10}$$

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

$$P_{t+1|t} = F\left[P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1}\right]F' + Q. \tag{11}$$

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 ξ_1,

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

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 $t = 2,\ldots,T$.

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

$$\xi_{t|T} = \mathbb{E}\left[\xi_t\middle|y_1,x_1,\ldots,y_T,x_T\right], \tag{12}$$

$$\xi_{t|t-1} = \mathbb{E}\left[\xi_t\middle|y_1,x_1,\ldots,y_{t-1},x_{t-1}\right]. \tag{13}$$

2.3. Estimation

Under the assumption that the errors $\begin{pmatrix} v_t \\ w_t \end{pmatrix}$ 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 $y_t | x_t, \mathcal{I}_{t-1}$ is Gaussian as well, which can then be used to form the sample log-likelihood function with elements $\ln(f_t)$:

$$
\begin{aligned}
\sum_{t=1}^{T} \ln(f_t) &= \sum_{t=1}^{T} \ln\left[f_{y_t | x_t, \mathcal{I}_{t-1}}(\cdot) \right] \\
&= \sum_{t=1}^{T} -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln\left| H' P_{t|t-1} H + R \right| \\
&\quad - \frac{1}{2} (y_t - A'x_t - H'\xi_{t|t-1})' (H' P_{t|t-1} H + R)^{-1} (y_t - A'x_t - H'\xi_{t|t-1}),
\end{aligned}
\tag{14}
$$

where $f_{y_t | x_t, \mathcal{I}_{t-1}}(\cdot)$ denotes the (conditional) density function of $y_t | x_t, \mathcal{I}_{t-1}$, 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 $t = 1, 2, \ldots, T$, 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

$$
\xi_{T+1|T} = F\xi_{T|T}.
\tag{15}
$$

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),

$$
\mathbb{E}\left[y_{T+1} \big| x_{T+1}, \mathcal{I}_T \right] = A'x_{T+1} + H\xi_{T+1|T}.
\tag{16}
$$

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 $T+2$:

$$
\mathbb{E}\left[y_{T+2} \big| x_{T+2}, \mathcal{I}_T \right] = A'x_{T+2} + H\xi_{T+2|T},
\tag{17}
$$

$$
\begin{aligned}
\xi_{T+2|T+1} &= F\xi_{T+1|T} + FP_{T+1|T}H(H'P_{T+1|T}H + R)^{-1} \\
&\quad \times \left(\mathbb{E}\left[y_{T+2} \big| x_{T+2}, \mathcal{I}_T \right] - A'x_{T+2} - H\xi_{T+2|T} \right),
\end{aligned}
\tag{18}
$$

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

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

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 $T + h$, $h = 3, 4, \ldots$, 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 & 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 et al. (1997) proposed a general state-space model that encompassed several competing models used in the literature for interpolation: Chow & Lin (1971), Fernández (1981), and Mitchell & Jones (2005), for example. They assume that unobserved monthly GDP (labelled as y_t^+ 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

$$\left(1 - \phi_1 L - \cdots - \phi_p L^p\right) y_t^+ = x_t \beta + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t. \tag{20}$$

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

$$y_t = \begin{cases} \displaystyle\sum_{i=0}^{2} y_{t-i}^+, & t = 3, 6, 9, 12, \ldots, T \tag{21} \\[2ex] 0, & \text{otherwise.} \tag{22} \end{cases}$$

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 & Uhlig (2005, Appendix, just above equation (1)). In the Kalman-filter literature for mixed frequency models (see, for example, Giannone, Reichlin, & 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 et al. (1997) and Mönch & Uhlig proceed.

---

[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.

A second issue is how to initialize the values of β, ρ, and the variance of u_t in the Kalman-filter procedure. Mönch & Uhlig (2005) 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 β and the ϕ's, as well as an estimate for the variance of u_t. With the latter, an estimate of ρ 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 ϕ, our state-space form is the following:

$$\xi_t = \begin{pmatrix} y_t^+ \\ y_{t-1}^+ \\ y_{t-2}^+ \\ u_t \end{pmatrix} = \begin{pmatrix} \phi & 0 & 0 & \rho \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} y_{t-1}^+ \\ y_{t-2}^+ \\ y_{t-3}^+ \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} x_t \beta \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \varepsilon_t \end{pmatrix} \tag{23}$$

$$y_t = \mathbf{H}_t' \xi_t, \tag{24}$$

where (23) and (24) are respectively the state and the observation equations and the matrix $\mathbf{H}_t'$ is time-varying, with the following format:

$$\mathbf{H}_t' = \begin{cases} \begin{bmatrix} 1 & 1 & 1 & 0 \end{bmatrix}, & t = 3, 6, 9, 12, \ldots, T \\ \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}, & \text{otherwise.} \end{cases} \tag{25}$$

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

2.6. Goodness of Fit Statistics for Interpolated Models

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

$$R_{\text{level}}^2 = \frac{\text{VAR}\left(\widehat{y}_{t|T}^+\right)}{\text{VAR}\left(\widehat{y}_{t|T}^+\right) + \text{VAR}\left(\widehat{u}_{t|T}\right)} \quad \text{and} \quad R_{\text{diff}}^2 = \frac{\text{VAR}\left(\widehat{\Delta y}_{t|T}^+\right)}{\text{VAR}\left(\widehat{\Delta y}_{t|T}^+\right) + \text{VAR}\left(\widehat{\Delta u}_{t|T}\right)}.$$

Bernanke et al. (1997) 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^2s, picking the one with the best fit.

Table 1. Resulting Model as a Function of ϕ and ρ in (23).

Model	ϕ	ρ
M_1: Static model in levels with IID residuals	0	0
M_2: Static model in levels with AR(1) residuals (Chow & Lin, 1971)	0	free
M_3: Static model in 1st differences with IID residuals (Fernández, 1981)	0	1
M_4: Dynamic model in levels with IID residuals (Mitchell & Jones, 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

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 ξ_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:

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

(ii) 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 & 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, & Rodrigues (2013) solved this problem by proposing a back-cast algorithm similar to the interpolation method in Bernanke et al. (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 & Watson (1989) 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.

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.[3] All the analyzed series show a high correlation with the GDP series once both are first-differenced, although

---

[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.

Table 2. ADF Unit Root Test.

Variables	$H_0 : z$ is $I(1)$	$H_0 : z$ is $I(2)$
GDP*	-0.898	-4.098
	[0.95]	[0.01]
Steel	-0.792	-5.514
	[0.82]	[0.00]
Vehicles	0.522	-6.208
	[0.99]	[0.00]
Cement	-0.220	-3.392
	[0.93]	[0.01]
Energy	-0.208	-26.672
	[0.93]	[0.00]
IP	0.690	-6.270
	[0.99]	[0.00]
Sales	-0.118	-6.156
	[0.95]	[0.00]

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.

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.[4]

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.[5]

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 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.

Table 3. Correlation Coefficients – GDP and Coincident Series.

Variables	Cement	Vehicles	Steel	Energy	IP	Sales
GDP	0.49	0.33	0.29	0.21	0.66	0.57

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.

---

[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.

Table 4. Interpolation Results.

Parameter	M_1	M_2	M_3	M_4	M_5	M_6
ϕ				0.94	-0.67	-0.60
Intercept	11.03	33.51	51.60	-0.029	62.01	45.77
I. Production	0.693	0.198	0.212	0.0480	0.349	0.36
Sales	0.0002	0.0001	0.0001	0.0000	0.0001	0.0001
ρ		1.0041			1.0045	
σ	6.5881	0.8955	0.9439	1.1463	1.2570	1.3259
Log likelihood	205.6269	88.3283	93.7001	109.0752	84.9365	92.7201

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.

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 ρ 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 ε_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.

Table 5. R^2_{diffs} Measures of Fit.

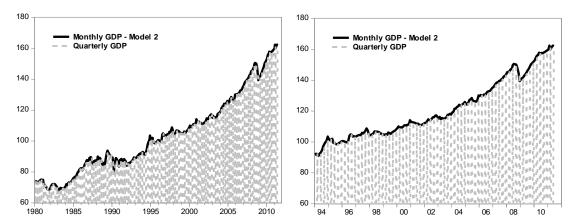|	R^2_{diffs}	
Model	GDP_filtered	GDP_smoothed
M_1	0.0273	0.0346
M_2	0.4652	0.8097
M_3	0.4382	0.7923
M_4	0.3609	0.7098
M_5	0.4011	0.5806
M_6	0.3766	0.5736

Figure 1. Monthly (Interpolated) and Quarterly (Observed) GDP.

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 BCB/Copom (2010), for example.[6]

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 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

---

[6]A detailed methodological description and the complete list of component time series could be found at Brazil's Central Bank homepage: http://www.bc.gov.br

[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.

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.
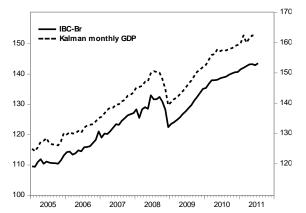

Figure 2. IBC-Br and Kalman-Filter Monthly GDP (levels).

Figure 3. IBC-Br and Kalman monthly GDP in first differences.
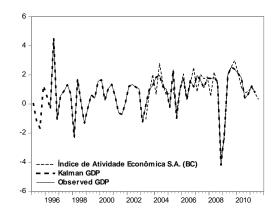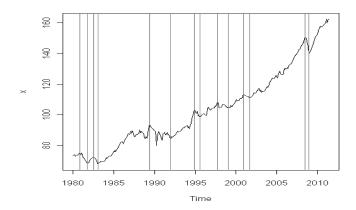
Figure 4. Quarterly Growth Rates of GDP, IBC-Br and Kalman-filter monthly GDP.

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.

Results in Table 6 show that the dating using our GDP monthly indicator yields results close to that of Duarte et al. (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

Figure 5. Monthly GDP – shaded by Bry Boshan recessions periods.

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

Table 6. Turning-Point Comparisons (Peaks and Throughs).

Dated Peaks			Dated Throughs		
Brazilian Monthly GDP	Duarte et al.	CODACE	Brazilian Monthly GDP	Duarte et al.	CODACE
1980:11	NA	1980:10	1981:10	NA	
1982:07			1983:02		1983:02
–	1987:04	1987:02	–	1989:02	1988:10
1989:06	1989:08	1989:06	1991:12	1991:12	1991:12
1994:12	1994:12	1994:12	1995:08	1995:09	1995:09
1997:10	1997:10	1997:10	1999:02	1999:02	1999:02
2000:12	2000:12	2000:12	2001:09	2001:09	2001:09
–	NA	2002:10	–	NA	2003:06
2008:07	NA	2008:07	2008:12	NA	2009:01

1994 affected Brazil and other emerging markets. In 1997–98, the Asian, the Russian, and the Brazilian 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 2000-02 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 et al. (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 Brasil—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.
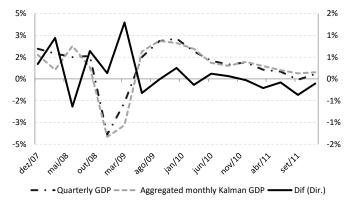
Figure 6. Nowcast Errors of the Monthly GDP model.

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. (2012) 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 7 summarizes the results.

Table 7. Nowcast Mean Squared Error − Monthly GDP Model Versus IBC-Br

Nowcast Period	IBC-Br	Monthly GDP Model
2007.4 − 2011.4	0.46%	0.40%
2010.4 − 2011.4	0.35%	0.16%

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 et al. (1997) and Mönch & 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 et al. (2004) and Issler, Notini, & Rodrigues (2013).

The methodology in Bernanke et al. (1997) and Mönch & Uhlig (2005), 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 an 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 et al. (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

BCB/Copom (2010). Índice de Atividade Econômica do Banco Central (IBC-Br). In *Relatório de Inflação*, Vol.12 No.1, 24–28. Brasília–DF: Banco Central do Brasil. Retrieved from http://www.bcb.gov.br/?RELINF201003

Bernanke, B. S., Gertler, M., & Watson, M. (1997). *Systematic monetary policy and the effects of oil price shocks* (Brookings Papers in Economic Activity No. 1). Washington, DC: Brookings Institution. Retrieved from http://www.brookings.edu/about/projects/bpea/papers/1997/effects-of-oil-price-shocks-bernanke

Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. New York: National Bureau of Economic Research (NBER).

Cardoso, E. A. (1981). Uma equação para a demanda de moeda no Brasil. *Pesquisa e Planejamento Econômico*, *11*(3), 617–655.

Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review*, *39*(4), 969–996. doi: 10.2307/2527348

Chauvet, M. (2001). A monthly indicator of Brazilian GDP. *Brazilian Review of Econometrics*, *21*(1), 1–48. Retrieved from http://bibliotecadigital.fgv.br/ojs/index.php/bre/article/view/3191

Chauvet, M. (2002). The Brazilian business and growth cycles. *Revista Brasileira de Economia*, *56*(1), 75–106. doi: 10.1590/S0034-71402002000100003

Chow, G. C., & Lin, A.-l. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, *53*(4), 372–375. doi: 10.2307/1928739

Contador, C. R. (1977). *Ciclos econômicos e indicadores de atividade no Brasil* (Relatório de pesquisa No. 35). Rio de Janeiro: Instituto de Planejamento Econômico e Social / Instituto de Pesquisas (IPEA/INPES).

Contador, C. R., & Ferraz, C. (1999). *Previsão com indicadores antecedentes*. Rio de Janeiro: Silcon.

Contador, C. R., & Santos Filho, W. A. C. d. (1987). Produto Interno Bruto trimestral: bases metodológicas e estimativas. *Pesquisa e Planejamento Econômico*, *17*(3), 711–742.

Cuche, N. A., & Hess, M. K. (2000). Estimating monthly GDP in a general Kalman filter framework: Evidence from Switzerland. *Economic and Financial Modelling*, *7*(4), 153–194.

Duarte, A. J. M., Issler, J. V., & Spacov, A. (2004). Indicadores coincidentes de atividade econômica e uma cronologia de recessões para o Brasil. *Pesquisa e Planejamento Econômico*, *34*(1), 1–37.

Fernández, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, *63*(3), 471–476. doi: 10.2307/1924371

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676. doi: 10.1016/j.jmoneco.2008.05.010

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press.

Hollauer, G., Issler, J. V., & Notini, H. H. (2009). Novo indicador coincidente para a atividade industrial brasileira. *Economia Aplicada*, *13*(1), 5–28. doi: 10.1590/S1413-80502009000100001

Issler, J. V., Notini, H. H., & Rodrigues, C. F. (2013). Constructing coincident and leading indices of economic activity for the Brazilian economy. *Journal of Business Cycle Measurement and Analysis*, *2012*(2), 43–65. doi: 10.1787/19952899

Issler, J. V., Notini, H. H., Rodrigues, C. F., & Soares, A. F. (2013). Constructing coincident indices of economic activity for the Latin American economy. *Revista Brasileira de Economia*, *67*(1), 67–96. doi: 10.1590/S0034-71402013000100004

Issler, J. V., & Spacov, A. D. (2000). *Usando correlações canônicas para identificar indicadores antecedentes e coincidentes da atividade econômica no Brasil* (Relatório de Pesquisa – mimeo). Ministério da Fazenda.

Issler, J. V., & Vahid, F. (2006). The missing link: Using the NBER recession indicator to construct coincident and leading indices of economic activity. *Journal of Econometrics*, *132*(1), 281–303. doi: 10.1016/j.jeconom.2005.01.031

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, *59*(6), 1551–1580. doi: 10.2307/2938278

Kim, C.-J., & Nelson, C. (1999, Sep). *A bayesian approach to testing for Markov switching in univariate and dynamic factor models* (Discussion Paper No. 0035). University of Washington, Department of Economics.

Liu, H., & Hall, S. G. (2001). Creating high-frequency national accounts with state-space modelling: a Monte Carlo experiment. *Journal of Forecasting*, *20*(6), 441–449. doi: 10.1002/for.810

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443. doi: 10.1002/jae.695

Mitchell, T. D., & Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, *25*(6), 693–712. doi: 10.1002/joc.1181

Mönch, E., & Uhlig, H. (2005). Towards a monthly business cycle chronology for the Euro Area. *Journal of Business Cycle Measurement and Analysis*, *2005*(1), 43–69. doi: 10.1787/jbcma-2005-5km7v183t48r

Nakane, M. I. (1994). *Testes de exogeneidade fraca e de superexogeneidade para a demanda por moeda no Brasil*. Rio de Janeiro: BNDES.

Notini, H., Issler, J. V., Rodrigues, C., Matos, S., , & Bonelli, R. (2012). Nowcasting Brazilian monthly GDP: A state-space approach [paper presented at CIRET 2012, Vienna].

Picchetti, P., & Toledo, C. (2002). Estimating and interpreting a common stochastic component for the Brazilian Industrial Production Index. *Revista Brasileira de Economia*, *56*(1), 107–120. doi: 10.1590/S0034-71402002000100004

Spacov, A. D. (2001). *Índices antecedentes e coincidentes da atividade econômica brasileira: Uma aplicação da análise de correlação canônica* (Dissertação de Mestrado). Fundação Getúlio Vargas, Escola de Pós-Graduação em Economia (FGV/EPGE), Rio de Janeiro.

Stock, J. H., & Watson, M. W. (1988, November). *A probability model of the coincident economic indicators* (Working Paper No. 2772). National Bureau of Economic Research (NBER). doi: 10.3386/w2772

Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. In O. J. Blanchard & S. Fischer (Eds.), *NBER Macroeconomics Annual 1989, volume 4* (pp. 351–409). MIT Press. Retrieved from http://www.nber.org/chapters/c10968

Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident economic indicator. In K. Lahiri & G. H. Moore (Eds.), *Leading economic indicators: New approaches and forecasting records* (pp. 63–89). Cambridge University Press.

Stock, J. H., & Watson, M. W. (Eds.). (1993a). *New research on business cycles, indicators and forecasting*. University of Chicago Press.

Stock, J. H., & Watson, M. W. (1993b). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. In J. H. Stock & M. W. Watson (Eds.), *New research on business cycles, indicators and forecasting*. University of Chicago Press.

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. doi: 10.1198/016214502388618960

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. doi: 10.1198/073500102317351921

Stock, J. H., & Watson, M. W. (2010). *Distribution of quarterly values of GDP/GDI across months within the quarter* [Research Memorandum]. Retrieved from https://www.princeton.edu/~mwatson/mgdp_gdi/Monthly_GDP_GDI_Sept20.pdf

APPENDIX. DETAILED INTERPOLATION MODELS

The literature appoint as the main problem of the models M_1 and M_2 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 M_1
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 ϵ_t a iid error term with distribution $N(0, \sigma^2)$.

Model M_2
This is the Chow & 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 & Lin model by fixing $\phi = 0$ and by letting ρ to be estimated freely:

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

$$y^+ = h_t'\varepsilon_t$$

It is important to note that in their seminal paper, Chow & Lin do not use the Kalman filter but it can be shown that the Kalman filter above and the Chow & Lin regression yield the same estimates by maximum likelihood (Cuche & Hess, 2000).

Model M_3
This is a variation of the Chow & Lin model, suggested by Fernández (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 & Jones (2005) and it is a dynamic version of model M_1.

Model M_5
A modified version of model M_4 where we impose that $\rho = 1$.

Model M_6
It is the most general version of the Mitchell & Jones (2005) model, where we let ϕ and ρ 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.