FUNDAÇÃO GETULIO VARGAS
ESCOLA de PÓS-GRADUAÇÃO em ECONOMIA

Guilherme Branco Gomes

Nowcasting Brazilian GDP
A Performance Assessment of Dynamic Factor Models

Rio de Janeiro
2018

Guilherme Branco Gomes

Nowcasting Brazilian GDP A Performance Assessment of Dynamic Factor Models

Dissertação para obtenção do grau de mestre apresentada à Escola de Pós-Grauação em Economia

Orientador: João Victor Issler

Rio de Janeiro
2018

Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Gomes, Guilherme Branco.
Nowcasting Brazilian GDP : a performance assessment of dynamic factor models / Guilherme Branco Gomes. – 2018.
42 f.

Dissertação (mestrado) - Fundação Getulio Vargas, Escola de Pós-Graduação em Economia.
Orientador: João Victor Issler.
Inclui bibliografia.

1.Produto interno bruto. 2. Previsão econômica. 3. Indicadores econômicos. I. Issler, João Victor. II. Fundação Getulio Vargas. Escola de Pós-Graduação em Economia. III. Título.

CDD – 339.32

GUILHERME BRANCO GOMES

"NOWCASTING BRAZILIAN GDP: A PERFORMANCE ASSESSMENT OF DYNAMIC FACTOR MODELS".

Dissertação apresentado(a) ao Curso de Mestrado em Economia do(a) Escola de Pós-Graduação em Economia para obtenção do grau de Mestre(a) em Economia.

Data da defesa: 19/03/2018

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

João Victor Issler
Orientador(a)

Felipe S. I
Felipe Saraiva Iachan

Wagner P. Gaglianone
Wagner Piazza Gaglianone

Resumo

Esse trabalho compara previsões para o PIB brasileiro utilizando modelos de fatores dinâmicos. Nossa abordagem leva em consideração frequências mistas e lida com dados incompletos na base (*missing data*). Nós implementamos três modelos: o primeiro é baseado na metodologia de componentes principais; o segundo emprega uma estimação por dois estágio com variáveis trimestrais; o último é similar ao anterior mas utiliza series mensais. Um exercício em tempo real, fora da amostra, é proposto para comparar o desempenho desses modelos. Uma base de dados é criada para cada dia dentro de 27 trimestres - do quarto trimestre de 2010 até o segundo de 2017. Para períodos recentes, os *nowcasts* estimados para ambos os procedimentos de dois estágios se mostram melhores do que a média de previsão da pesquisa Focus, um boletim organizado pelo Banco Central do Brasil. Nós também mostramos evidências que a média das previsões do PIB dessa pesquisa pode ser viesada.

Palavras-chave: PIB, Modelo de Fatores Dinâmicos, Nowcasting

Abstract

This work compares dynamic factor model's forecasts for Brazilian GDP. Our approach takes into account mixed frequencies and can handle missing data. We implement three models: the first is based on the Principal Components Analysis methodology; the second employs a two-step estimation method with quarterly inputs; the last is similar to the former but uses monthly series. A real-time out-of-sample exercise is proposed to assess the performance of these models. A dataset is created for each day within 27 quarters - from the fourth quarter of 2010 up to the second quarter of 2017. For recent periods, the nowcasts estimated by both two-step procedures perform better than the average predictions of *Focus Survey*, a bulletin organized by the Brazilian Central Bank. We also show evidence that the average of GDP forecasts from this survey may be biased.

Keywords: GDP, Dynamic Factor Model, Nowcasting

Contents

1 Introduction

A real-time forecast attempts to measure the current state of a stochastic process when the information about its realization is not yet available. Several economic agents are interested in the current state of economic activity. Market participants, for example, wish to know the realization of the economic scenario as soon as possible, to correctly price assets. The government also has these concerns, either to perform and evaluate public policies according to the current state of the economy or to control the federal budget predicting the present tax revenues.

An economic variable that is fundamental to measure a country's level of activity is the GDP. In Brazil, GDP is measured by *Instituto Brasileiro de Geografia e Estatística* (IBGE) and published, on average, nine weeks after the end of the reference period. Consequently, using only the GDP to guide economic actions would be like driving a car looking solely at the rearview mirror. Although past values of a time series have relevant information about its present state, it is necessary to project it into a more informative set of variables to obtain a better approximation.

In this context, efforts have been made to monitor economic series in real time. For instance, since 1999 the Brazilian Central Bank (BCB) promotes a survey - the so-called *Focus Survey* - to gather the expectations of market participants about a set of macroeconomic series, such as the GDP and the Brazilian Consumer Price Index (CPI). This survey does not provide information about how respondents form their expectations: some may be using reproducible mathematical models and others judgment and not reproducible forecasts. Issler and Lima (2009) and Gaglianone et al. (2017) show that means of surveys' predictions may be considered to be formed by a combination of models and propose a method to evaluate and correct possible biases.

Another monitoring procedure is to construct indicators that match the series of interest. The BCB also works on this front, producing the *Índice de Atividade Econômica do Banco Central* (IBC-Br) and the monthly GDP - monthly frequency indicators that co-move with quarterly GDP when aggregated.[1] These indicators tend to have a good forecasting performance. They are regarded as a monthly preview of GDP but suffer from the same problem, since their delay is equally high. On average, the indicators for the quarter's last month are released two weeks

[1]Issler and Notini (2016) propose an estimation of Brazilian monthly GDP that outperforms those of the BCB. They use a state-space structure where the monthly GDP is considered a latent variable.

before GDP. They are built following a series aggregation methodology that seeks to reproduce, from the bottom-up, IBGE's approach. It means that the same set of information is necessary to develop them.

Because of this relative delay in data disclosure, it is interesting to develop an econometric technique capable of measuring economic variables in real time, even when the set of information needed for the calculation of coincident indicators is incomplete.

The literature that formalizes and treats this class of problems started approximately ten years ago with Evans (2005) and Giannone et al. (2008). The word nowcasting has been coined to name the problem of predicting the present, the recent past, or the near future. This term was already used in meteorology, and Giannone et al. (2008) incorporated it into economics for the first time. Although the literature about nowcasting is new, it emerged from another econometric literature branch on Dynamic Factor Models (DFMs). Some reasons for using these models in real-time forecasting are: they can handle a large number of variables in a joint structure, in a parsimoniously way; they do not discard information; they have good asymptotic properties.

Barhoumi et al. (2008) compare the performance of some DFM estimation techniques to forecast the GDP for several European economies. Concerning emerging markets, Dahlhaus et al. (2017) apply a DFM framework to forecast the GDP of Brazil, Russia, India, China, and Mexico. Their analysis is comprehensive and does not delve deep into any of these economies. For Brazil, they use 36 variables but only estimates 6 predictions for each GDP realization. Specifically for Brazil, Bragoli et al. (2015) employ only one technique of DFM estimation and exploit a reduced dataset with 13 series.

In this scenario, the present work aims to compare the performance of different estimation DFM techniques to predict Brazilian aggregated GDP.[2] Our main contribution to the general literature on DFM is to assess the performance of these models for a developing economy such as Brazil. First, we collected a large dataset with more than fifty macroeconomic series and recreated the information available for these series to more than 2430 days in the past. Then, we reproduced three DFMs. One of them is the classical Stock and Watson (2002a) DFM, which uses a Principal Component Analysis (PCA) and thus demands a balanced panel. The two others, labeled *nowcasting*, employ a Kalman Filter (KF) and therefore can handle

[2]Modeling each component of GDP isolated is possible. However, our approach proposes to model only the aggregated component.

missing data at the edge of the dataset, taking into account the particular way that information is released. An additional goal of this work is to investigate whether the *Focus Survey* is biased for some prediction horizons according to the methodology proposed by Issler and Lima (2009), Gaglianone and Issler (2015) and Gaglianone et al. (2017).

This paper proceeds as follows. In section 2, we develop a literature review on Dynamic Factors Models (DFM) and real-time forecasting, as well as their assumptions and estimation methods. We also discuss the main features of nowcasting models. Section 3 presents our empirical implementation for the Brazilian economy. The study on *Focus Survey*'s bias is presented in section 4. In section 5 we present the comparison of the results of the three models we implemented, together with the outcome of a benchmark univariate model and the Focus Survey mean. Finally, section 6 concludes.

2 Models Review

In this section, we review the literature on Dynamic Factor Model (DFM) and nowcasting models. We chose to divide it into three interdependent parts. In the first part, we present the general structure of DFM, along with its central assumptions. The subsection 2.2 is dedicated to delineate and discuss key estimation methods. The final part of the section discusses the critical concepts of nowcasting.

2.1 Dynamic Factor Model Framework

The DFMs have received particular attention for their ability to model, in a joint fashion, a peculiar macroeconomics data structure. This structure is a panel data with n time series and T observations, where n is large in the sense that there are many contemporaneous series and T is small because some of these series are short in length. These models allow condensing numerous explanatory variables in a parsimonious way and using all the information they contain.[3] DFM estimators have good asymptotic properties for both n and T, which led Stock and Watson (2011) to state that DFMs have a "blessing of dimensionality" in contrast with models where large dimensions are treated as a curse.

[3]It differs from "shrinking" techniques like the LASSO and ridge equation, which select the best series to explain a particular variable and discard the others. For more details see Tibshirani (1996).

Interesting surveys about this theme are found in Stock and Watson (2006), Stock and Watson (2011), and Bai et al. (2008). The first two describe some applications of DFMs for large-scale forecasting, comparing it with other methods. The third presents a more technical survey.

Factor analysis was developed initially by psychologists to understand the principal characteristics of human behavior. Spearman (1904) is the pioneer proposing a measure of the factors in terms of correlations in test scores. Lawley and Maxwell (1962) present a model formulation for the data generation process dependence of factors, called the factor model. The premise of factor model is that a large number n of variables can be explained by a significantly smaller number q of factors. The first appearance of the DFM in the econometric literature was in Geweke (1977) and Sargent et al. (1977), a version of the factor model where the variables of interest have a temporal dimension.

We present the DFM hypotheses below, following Forni et al. (2009). Denote by $\boldsymbol{x}_n^T = (x_{it})_{i=1,..,n;t=1,...,T}$ a panel data with dimensions $n \times T$.

Assumption 1: $\boldsymbol{x}_n^T$ is a finite realization of a real-valued stochastic process

$$\boldsymbol{X} = \{x_{it}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}, \quad x_{it} \in L_2(\Omega, \mathscr{F}, P)\}$$

indexed by $\mathbb{N} \times \mathbb{Z}$, where the n-dimensional vector processes

$$\boldsymbol{x}_{nt} = (x_{1t}, ..., x_{nt})', \quad t \in \mathbb{Z}, \quad n \in \mathbb{N}$$

are stationary, with zero mean and finite second-order moments:

$$\Gamma_k^x = E[\boldsymbol{x_{nt}x'_{n,t-k}}], \quad k \in \mathbb{Z}$$

We assume that each variable x_{it} is a sum of two orthogonal and nonobservable components, the *common component* χ_{it} and the *idiosyncratic component* ξ_{it}. The common components are a function of q common shocks $\boldsymbol{u}_t = (u_{1t}, ..., u_{qt})'$, where q is smaller than n. Define $\boldsymbol{\chi_{nt}} = (\chi_{1t}, ..., \chi_{nt})'$ and $\boldsymbol{\xi_{nt}} = (\xi_{1t}, ..., \xi_{nt})'$, then

$$\boldsymbol{x}_{nt} = \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} \tag{1}$$
$$\boldsymbol{\chi}_{nt} = B_n(L)\boldsymbol{u}_t \tag{2}$$

where the following conditions hold:

Assumption 2: $\boldsymbol{u}_t$ is an orthonormal white noise vector of dimension q and $B_n(L)$ is a summable $n \times q$ matrix of lag polynomials. Further, there exists an integer $r \geq q$, λ_n a $n \times r$ matrix, and $N(L)$ a summable $r \times q$ matrix of lag polynomials such that

$$B_n(L) = \lambda_n N(L) \tag{3}$$

Define $\boldsymbol{f}_t$ as an $r \times 1$ vector such that

$$\boldsymbol{f_t} = N(L)\boldsymbol{u}_t \tag{4}$$

The Equation (1) can be written in a *static form*, that is

$$\boldsymbol{x}_{nt} = \lambda_n \boldsymbol{f}_t + \boldsymbol{\xi}_{nt} \tag{5}$$

Forni et al. (2009) labels the vector $\boldsymbol{f}_t$ *static factors.* At the same time, common shocks $\boldsymbol{u}_t$ are called *dynamic factors.*

Assumption 3: For all n, the vector $\boldsymbol{\xi}_{nt}$ is stationary. Moreover, the dynamic factors $\boldsymbol{u}_t$ are orthogonal to each idiosyncratic shock $\xi_{i\tau}, \quad i \in \mathbb{N}, \quad t \in \mathbb{Z}, \quad \tau \in \mathbb{Z}$.

Following Stock and Watson (2011), if one imposes the hypothesis that idiosyncratic terms are mutually independent for all lags and leads, i.e. $\mathbb{E}[\xi_{it}\xi'_{js}] = 0$ for $s \in \mathbb{Z}$ and $i \neq j \in \mathbb{N}$ given $t \in \mathbb{Z}$, then it is an *exact DFM.* In some studies, like in Banbura et al. (2011) and Banbura et al. (2013), an exact DFM is not imposed. Instead, an autoregressive structure is proposed for the idiosyncratic terms.

If $\boldsymbol{u}_t$ belongs to the subspace spanned by present and past values of $\boldsymbol{f}_t$, it is said that the representation in (4) is *fundamental* and $\boldsymbol{u}_t$ is fundamental to $\boldsymbol{f}_t$. For this condition to be valid, implying that it is possible to obtain the dynamic factors from the static factors, the following hypothesis is sufficient:[4]

[4]Forni et al. (2000), Forni and Lippi (2001) and Forni et al. (2009) further discuss this repre-

Assumption 4: (Fundamentalness) There exist an $r \times q$ matrix denoted by R, and a quantity p, $p \in \mathbb{N}$, of $r \times r$ matrices, such that $(I - A_1 L - ... - A_p L^p)N(L) = R$.

With this assumption, the Equation (4) can be rewritten as a VAR(p)

$$\boldsymbol{f}_t - A_1 \boldsymbol{f}_{t-1} - ... - A_p \boldsymbol{f}_{t-p} = R\boldsymbol{u}_t \tag{6}$$

Considering the generation process of the time series $\boldsymbol{x}_{nt}$, questions about identification and estimation follow.

2.2 Identification and Estimation

Identification problems are related to the structure of the generating process. Assuming that the ideal model to describe the data is the DFM, then it is necessary to find the number of dynamic factors q, static factors r, and the degree p of the VAR (which updates the factors). We also need to verify if the model is exact and if the variance of the idiosyncratic components changes. Since our objective is forecasting, we use as identification technique the model that produces the best out-of-sample results. For more details about alternative identification techniques see Giannone et al. (2004), Forni et al. (2009) and Stock and Watson (2011).

Stock and Watson (2011) present a chronology of estimation methods. We discuss some of them throughout this section. It begins with Geweke (1977), who implemented an algorithm for estimating a one-factor dynamic model in the frequency domain. Sargent et al. (1977) presented an application of the former with more factors and developed an economic intuition to them. Later, Engle and Watson (1981) and Watson and Engle (1983) introduced the temporal domain via the Kalman Filter, but the implementation was unfeasible due to tremendous computational efforts demanded at that time which induced the use of models with small datasets. Other methodologies were used to overcome the numerical problem, for instance, those non-parametric of the PCA family proposed by Stock and Watson (2002a). More recently, a mixed strategy has been proposed using the PCA together with the Kalman filter as in Doz et al. (2011).

Stock and Watson (2011) divides the estimation procedures in three generations. The first generation of DFM estimates the maximum likelihood via a Kalman Filter. Note that Equations (5) and (6) can be rewritten in the *companion form* as a

sentation and generalize the conditions under which it is valid.

VAR(1).

$$\boldsymbol{x}_{nt} = \underbrace{\begin{bmatrix} \lambda_n & 0_{n\times r} & ... & 0_{n\times r} \end{bmatrix}}_{\Lambda} \underbrace{\begin{pmatrix} \boldsymbol{f}_t \\ ... \\ \boldsymbol{f}_{t-p+1} \end{pmatrix}}_{\boldsymbol{F}_t} + \boldsymbol{\xi}_{nt} \tag{7}$$

$$\begin{pmatrix} \boldsymbol{f}_t \\ ... \\ \boldsymbol{f}_{t-p+1} \end{pmatrix} = \underbrace{\begin{bmatrix} A_1 & A_2 & ... & A_{p-1} & A_p \\ I_r & 0_r & ... & 0_r & 0_r \\ ... & ... & ... & ... & ... \\ 0_r & 0_r & ... & I_r & 0_r \end{bmatrix}}_{A} \begin{pmatrix} \boldsymbol{f}_{t-1} \\ ... \\ \boldsymbol{f}_{t-p} \end{pmatrix} + \underbrace{\begin{bmatrix} I_r \\ 0_r \\ ... \\ 0_r \end{bmatrix}}_{G} R\boldsymbol{u}_t \tag{8}$$

which is equivalent to a state-space representation,

Measurement equation

$$\boldsymbol{x}_{nt} = \Lambda \boldsymbol{F}_t + \boldsymbol{\xi}_{nt} \tag{9}$$

Transition equation

$$\boldsymbol{F}_t = A\boldsymbol{F}_{t-1} + GR\boldsymbol{u}_t \tag{10}$$

Equations (9) and (10) represent the *static form* of the DFM. To complete the identification, a parametric form is proposed for the shocks $\boldsymbol{u}_t$, typically a normal distribution.

The factors can be estimated via KF (Kalman Filter)[5] if the parameters of these equations are known; otherwise, the EM (Expected Maximization) algorithm can be implemented. Because EM is recursive, there were numerical impediments to estimate it when n is large. Hence, empirical studies reduced the number of parameters using fewer series and introducing restrictions from economic theory. Some examples of the first generation are found in Engle and Watson (1981),Watson and Engle (1983), Stock and Watson (1989) and Quah and Sargent (1993).

In the second generation, non-parametric estimation methods are used, where the most important equation is Equation (9), and no structure is imposed on the transition Equation (10). Moreover, no functional form is imposed on the distribution of

[5]For more details see Durbin and Koopman (2012).

errors and idiosyncratic terms. It explains why they are called non-parametric methods. Stock and Watson (2002a) (2002b) and Bernanke et al. (2005), for example, use this estimation technique.

In this methodology, the orthogonality between common and idiosyncratic components is explored. With a large number of variables, the cross-sectional mean equals the common factor, for idiosyncratic factors vanish as a consequence of the law of large numbers. Results can be obtained under the very general conditions presented in the *approximate factor model* of Chamberlain and Rothschild (1983) and Chamberlain (1983) and deeply explored in Stock and Watson (2002a), Stock and Watson (2002b) and Forni et al. (2009). A simplified form of these conditions is given below.

Assumption 5:

1. $n^{-1}\lambda_n'\lambda_n \to D_\Lambda$ as $n \to \infty$, and D_Λ has full rank equals to q;
2. $maxeval(E[\boldsymbol{\xi}_{nt}\boldsymbol{\xi}_{nt}'])\leq c<\infty$ for all n, where *maxeval* denotes the maximum eigenvalue.

Condition (1) is to ensure that the factors are pervasive (i.e., they affect all or almost all series) and that factor loadings matrix are heterogeneous (Λ columns are not similar). Condition (2) guarantees that the weak law of large numbers holds for the cross-section and the idiosyncratic terms fade as n increases.

In this generation of models, the matrix Λ and the factors in the static form $\boldsymbol{F}_t$ are estimated by PCA. Stock and Watson (2011) show that these estimators can be derived with OLS. Starting from Equation (9) and considering a quadratic loss function, the minimization problem can be stated as

$$\min_{\boldsymbol{F}_1,\ldots,\boldsymbol{F}_T,\Lambda} \frac{1}{nT}\sum_{t=1}^{T}(\boldsymbol{x}_{nt}-\Lambda\boldsymbol{F}_t)'(\boldsymbol{x}_{nt}-\Lambda\boldsymbol{F}_t) \tag{11}$$

if $\boldsymbol{F}_t$, $t=1,...,T$, and Λ are parameters, this problem admits infinite solutions since we have one degree of freedom in the determination of the estimators. To observe this, denote a scalar $\alpha \neq 0$ and consider the parameters Λ^* and $\boldsymbol{F}_t^*$ $t = 1,...,T$

$$\Lambda^* = \alpha\Lambda$$
$$\boldsymbol{F}_t^* = \frac{1}{\alpha}\boldsymbol{F}_t$$

Note that

$$\Lambda^*\boldsymbol{F}_t^* = (\alpha\Lambda)\Big(\frac{1}{\alpha}\boldsymbol{F}_t\Big) = \Lambda\boldsymbol{F}_t$$

We impose an identification constraint on the model in which the matrix Λ is orthogonal, to avoid the case of infinite solutions. This way, the optimization problem can be rewritten as

$$\min_{\boldsymbol{F}_1,\dots,\boldsymbol{F}_T,\Lambda} \frac{1}{nT}\sum_{t=1}^{T}(\boldsymbol{x}_{nt} - \Lambda\boldsymbol{F}_t)'(\boldsymbol{x}_{nt} - \Lambda\boldsymbol{F}_t) \quad s.t. \quad N^{-1}\Lambda'\Lambda = I_r \tag{12}$$

Assuming that we know Λ, the estimator of $\boldsymbol{F}_t$ is the OLS applied to the cross-section

$$\hat{\boldsymbol{F}}_t = (\Lambda'\Lambda)^{-1}\Lambda'\boldsymbol{x}_{nt} \tag{13}$$

Denote $\hat{\Sigma}_X = T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{nt}\boldsymbol{x}_{nt}'$. Solving the Equation (12) for Λ using the estimator $\hat{\boldsymbol{F}}_t$ we have

$$\begin{aligned}
&\min_{\Lambda} T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{nt}'[I_n - \Lambda(\Lambda'\Lambda)^{-1}\Lambda]\boldsymbol{x}_{nt} \quad s.t. \quad N^{-1}\Lambda'\Lambda = I_r\\
\Rightarrow &\max_{\Lambda} tr\{(\Lambda'\Lambda)^{-\frac{1}{2}}\Lambda'(T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{nt}\boldsymbol{x}_{nt}')\Lambda(\Lambda'\Lambda)^{-\frac{1}{2}}\} \quad s.t. \quad N^{-1}\Lambda'\Lambda = I_r\\
\Rightarrow &\max_{\Lambda} tr\{(\Lambda'\Lambda)^{-\frac{1}{2}}\Lambda'(\hat{\Sigma}_X)\Lambda(\Lambda'\Lambda)^{-\frac{1}{2}}\} \quad s.t. \quad N^{-1}\Lambda'\Lambda = I_r\\
\Rightarrow &\max_{\Lambda} tr\{\Lambda'\hat{\Sigma}_X\Lambda\}
\end{aligned}$$

We can see, in this last equation, that the columns of matrix Λ are the eigenvectors associated to the r larger eigenvalues of matrix $\hat{\Sigma}_X$, which in turn is the estimation of variance-covariance matrix of $\boldsymbol{x}_{nt}$. Having an estimate for matrix Λ, it is possible to obtain $\hat{\boldsymbol{F}}_t$ through Equation (13) using OLS.

The variance-covariance matrix and the PCA vary with the series' unit of measurement. To overcome this issue and not to assign different weights to the series due to the measurement difference, the panel series $\boldsymbol{x}_n^T$ are standardized (meaning their mean is set to zero and their variance, to one).

The third generation of DFM estimation techniques is characterized by hybrid methods using PCA and KF. They allow the use of many variables, overcoming a limitation of the first generation, and also estimate the way in which the factors are updated, which was not explored in the second. This generation of models begins with the two-step estimation implemented in Giannone et al. (2008) and Giannone et al. (2004), and proposed by Doz et al. (2011).

On the first step, the parameters of Equation (9) and factors are estimated following the PCA methodology. Let $\boldsymbol{\Phi}^T = (\hat{\boldsymbol{F}}_t)_{t=2,\dots,T}$ and $\boldsymbol{\Phi}^{T-1} = (\hat{\boldsymbol{F}}_t)_{t=1,\dots,T-1}$. The estimators of Equation (10) are obtained by OLS:

$$\hat{A} = (\boldsymbol{\Phi}^{T-1}\boldsymbol{\Phi}^{T-1\prime})^{-1}\boldsymbol{\Phi}^{T-1}\boldsymbol{\Phi}^{T\prime} \tag{14}$$

The estimates of the first r rows of matrix A are the values of the first r rows of matrix $\hat{A}$. Let $\hat{\hat{A}}$ be the estimate of the matrix A and define

$$\widehat{RR'} = (T-1)^{-1}\sum_{t=2}^{T}(\hat{\boldsymbol{F}}_t - \hat{\hat{A}}\hat{\boldsymbol{F}}_{t-1})(\hat{\boldsymbol{F}}_t - \hat{\hat{A}}\hat{\boldsymbol{F}}_{t-1})' \tag{15}$$

Denote by $\widehat{\widehat{RR'}}$ the estimator of matrix RR'. Its coefficients are the first r rows and r columns of matrix $\widehat{RR'}$.

On the second step, the factors $\boldsymbol{f}_t$ are reestimated applying the Kalman smoother. The estimators found in the previous step are used as parameters in the following state-space representation

Measurement Equation

$$\boldsymbol{x}_{nt} = \hat{\Lambda}\boldsymbol{F}_t + \boldsymbol{e}_t \tag{16}$$

where $\hat{\Lambda} = \begin{bmatrix} \hat{\lambda}_n & 0_{n\times r} & \dots & 0_{n\times r} \end{bmatrix}$ and $\hat{\lambda}_n$ is the PCA estimator of matrix λ_n, i.e. the r eigenvectors associated with the r biggest eigenvalues of $\hat{\Sigma}_X$.

Transition Equation

$$\boldsymbol{F}_t = \hat{\hat{A}}\boldsymbol{F}_{t-1} + GR\boldsymbol{u}_t \tag{17}$$

and

$$\boldsymbol{u}_t \sim i.i.d.N(0, \widehat{\widehat{RR'}}) \tag{18}$$

Let $\hat{\hat{\boldsymbol{f}}}_t$ be the static factors estimated by the Kalman smoother applied on the above representation. The consistency properties of this estimator are discussed in Doz et al. (2011). In the next section, we explain how this two-step estimation method is essential for nowcasting.

2.3 Nowcasting

Banbura et al. (2013) define nowcasting as the prediction of the present, the near future and the recent past. It can also be considered the real-time monitoring of a group of variables, or even high-frequency predictions, where models are reestimated daily. These types of models use the DFM theoretical framework but powerfully explore the relationship between the data structure and the information set available at each instant of time.

Data are not available contemporaneously for all the economic series. Consider two monthly series: the CPI and Industrial Production (IP). In Brazil, the CPI is reported, on average, 11 days after the end of the reference period, while the IP after 30 days. An individual who collects information between these two dates has an unbalanced panel since at the end the information is incomplete. This feature of the real-time data structure is called *jagged edge*.[6]

We present insights that emerge when working with such feature. Let v be the vintage or the date when data are collected. The information set available in v is

$$\Omega_v = \{x_{it|v}, \quad i = 1, ..., n, \quad t = 1, ..., T_{iv}\} \tag{19}$$

where i identifies the n variables and $t = 1, ..., T_{iv}$ the time from the first to the last available observation, which depend on both the series i and the vintage v. Let y_t be the economic time series that one wants to predict, constrained by the

[6] Croushore (2011) presents a survey of real-time econometrics and the data structure issues concerning this approach.

information set Ω_v. More precisely, one is interested in the realization of $y_K \notin \Omega_v$. The conditional expectation is used to project y_K over Ω_v

$$\hat{y}_{K|v} = \mathbb{E}[y_K|\Omega_v] \tag{20}$$

When news is released, the information set is updated to another vintage w. Then a new expectation can be calculated.

$$\hat{y}_{K|w} = \mathbb{E}[y_K|\Omega_w] \tag{21}$$

If $\Omega_v \subset \Omega_w$, we presume that $\hat{y}_{K|w}$ is better than than $\hat{y}_{K|v}$, in the sense that it is closer to y_K.

This real-time estimation with an incomplete information set is the *nowcast* itself. An advantage of the two-step estimation proposed by Giannone et al. (2008) is that it is able to deal with data for any vintage, without discarding information.

In the first estimation step, we truncate the original panel to construct a new one with complete information, eliminating the jagged edge problem. In the second step, a Kalman smoother is applied to the original database and not only to the truncated panel since this filter is capable of dealing with missing information by imposing a very high variance to them.

We assume that y_t does not depend on the dynamics of any specific variable x_{it}, but rather on their joint dynamics. When we have dozens of series, a projection of y_t on all of them is unfeasible and over-parameterized. A parsimonious approach that takes into account all available information is to project y_t on the static factors extracted from these series via DFM.

Let y_t be the variable of interest. We suppose that this variable depends linearly on the common factors of $\boldsymbol{x}_{in}$ labeled $\boldsymbol{f}_t$ and an idiosyncratic component ϵ_t.

$$y_t = \alpha + \boldsymbol{\beta}'\boldsymbol{f}_t + \epsilon_t \tag{22}$$

Equation (22) is commonly called *bridge equation.* Stock and Watson (2011) state that factors can be treated as data, as long as we have a good approximation to them.[7] Let $\hat{\hat{\boldsymbol{f}}}_{t|v}$ be the static factors estimated using the information available in Ω_v, so

[7]In theory, there is a mismeasured error in $\boldsymbol{f}_t$ estimation which is not observable. For more details see Pagan (1984), Murphy and Topel (2002) and Hausman (2001). Nevertheless, our objective is prediction ou-of-sample and then this issue becomes less relevant.

$$y_t = \alpha + \boldsymbol{\beta}' \hat{\hat{\boldsymbol{f}}}_{t|v} + \epsilon_t \tag{23}$$

The parameters α and β are estimated via OLS using complete observations for y_t. The prediction for y_t to date K is obtained by updating the estimates of the factors $\hat{\hat{\boldsymbol{f}}}_t$ via Kalman smoother and applying the following linear transformation

$$\mathbb{E}[y_K|\Omega_v] = \hat{\alpha} + \hat{\boldsymbol{\beta}}' \hat{\hat{\boldsymbol{f}}}_{K|v} \tag{24}$$

where $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are estimates for the parameters α and $\boldsymbol{\beta}$. We obtain real-time forecasts for y_t every time the information set Ω_v is updated.

One question that arises in empirics is how to estimate Equation (23) if the dependent and independent variables are in different frequencies. A first procedure proposed by Giannone et al. (2008) is to aggregate all variables to the lowest frequency so that the estimated factors are in this frequency. Another methodology, following Bańbura and Rünstler (2011), is to estimate the factors with the variables at their original frequency and only then proceed with the aggregation.[8]

Nowcasting models are based on the structural form of the DFM with some additional specifications about the data structure. Banbura et al. (2013) present a relevant survey on this theme. In order to trace the rise of nowcasting from DFM literature we start with Stock and Watson (2002a) (2002b), who inaugurated the second generation of DFM using PCA estimation. We saw that it allows the introduction of many variables, taking advantage of asymptotic properties of DFMs, overcoming those of the first generation. Mariano and Murasawa (2003) proposed a method with mixed frequencies (monthly and quarterly) using the same estimation found in Stock and Watson (1989), i.e., MLE and KF. Next, Giannone et al. (2008) merged the previous works and inaugurated the third generation of DFMs, using the mixed frequency approximation of Mariano and Murasawa (2003), the estimation of PCA in the first step, and KF in the second.

[8]We use the filters presented in the appendix (including the one proposed by Mariano and Murasawa (2003)) to transform monthly in quarterly series. Other techniques that incorporate mixed frequencies into time-series analysis are presented in Ghysels et al. (2004) and Aruoba et al. (2009).

3 Empirical Results

The database used to perform these exercises is presented in the appendix and was built with monthly series[9]. They were chosen to represent the following categories of the Brazilian economy: general indexes, prices, labor, industry, sales, energy, international, monetary and finance, surveys and others. In order to compose this database, we selected the most significant number of economic series that are available at the Central Bank of Brazil (BCB) online databases and other data sources [10]. All series were collected on November 23, 2017, and we used a time window that starts on January 1, 2002.

Since real-time information is not available for all selected series, we follow Giannone et al. (2008) and create a proxy for each past information set vintage (exactly 2432 days). This is a proxy for the database collected if we were in the days inside the window defined above. It is built considering a stylized outreach calendar that is adapted for every month, as presented in the last column of the appendix table. To illustrate, we consider that the IP is released with a delay of 30 days every month.

Generally, the seasonal adjustment is re-estimated whenever new data is released, changing the past values of the entire series. Therefore we collected the original series in levels and without seasonal adjustment. In the models we studied, the series must be stationary. For this reason, the transformations presented in the appendix table's fifth column are applied as in Giannone et al. (2008).

3.1 PCA Estimation

Mariano and Murasawa (2003) propose a coincident indicator for the US economy based on the model in Stock and Watson (1989). Here we perform a similar exercise. However, we employ the PCA estimation proposed in Stock and Watson (2002a) (2002b) which allow us to extract the factors from a large information set. The model follows the Equation (5):

$$\boldsymbol{x}_{nt} = \lambda_n \boldsymbol{f}_t + \boldsymbol{\xi}_{nt}$$

where $\boldsymbol{x}_{nt}$ are monthly variables. We use a complete panel to estimate the static factors of Equation (9) via PCA following Stock and Watson (2002a) (2002b). As presented in the literature review, all the variables of the month must have

[9]The only exception is the GDP which is quarterly.

[10]The sources are listed in the appendix table.

been disclosed to implement this estimation, for this method does not deal with an unbalanced panel.

We are interested in nowcasting GDP, which is a quarterly time series. Nevertheless, only monthly variables are used to extract the factors. Then, we suppose that the monthly factors $\boldsymbol{f}_t$ are aggregated to represent quarterly quantities $\boldsymbol{f}_t^Q$ following the Mariano and Murasawa (2003) filter:

$$\boldsymbol{f}_t^Q = 1\boldsymbol{f}_t + 2\boldsymbol{f}_{t-1} + 3\boldsymbol{f}_{t-2} + 2\boldsymbol{f}_{t-3} + 1\boldsymbol{f}_{t-4} \tag{25}$$

and can linearly explain the quarterly variable y_t that represents GDP, as in Equation (22)

$$y_t = \alpha + \boldsymbol{\beta}'\boldsymbol{f}_t^Q + \epsilon_t \tag{26}$$

We transform the monthly factors in quarterly quantities applying the transformation (25). After that, we estimate the coefficients of Equation (26) by OLS. Finally, we obtain the one-step-ahead forecast applying the last estimated factor in the same equation. Since all the monthly variables are released before the GDP, the last factor represents precisely the quarter of interest.

Let's take a look at the forecasting model for the second quarter of 2017[11]. We initially used the first 5 principal components to calculate 5 quarterly factors that subsequently were regressed on the available GDP information. Table 1 presents the estimation results

The only significant regressor is *Factor 1*, so we perform an exercise to re-estimate the model excluding the non-significant regressors. The results are reported in the Table 2. Comparatively, the loss of explanatory power is irrelevant, but the second model is more parsimonious.

Notice that even with the reduction in the number of regressors the model presents a good in-sample fit (this means that the model fits well with the sample used to estimate it). The Coefficient of Determination (R^2) is greater than 90%. Still, this work aims to perform nowcasts, so it is necessary to evaluate the out-of-sample results. In other words, test the model on a sample not used for estimation.

Given the factor of the 2017's second quarter and using the coefficient presented

[11]On November 23, 2017, the day we collect data, every monthly variable had been disclosed.

Table 1: Bridge equation with 5 Factors

	Dependent variable:
	GDP
Factor 1	0.219***
	(0.018)
Factor 2	0.005
	(0.013)
Factor 3	−0.011
	(0.013)
Factor 4	0.031
	(0.020)
Factor 5	−0.020
	(0.019)
Constant	−0.062
	(0.105)
Observations	52
R^2	0.935
Adjusted R^2	0.928
Residual Std. Error	0.751 (df = 46)
F Statistic	131.863*** (df = 5; 46)
Note:	*p<0.1; **p<0.05; ***p<0.01

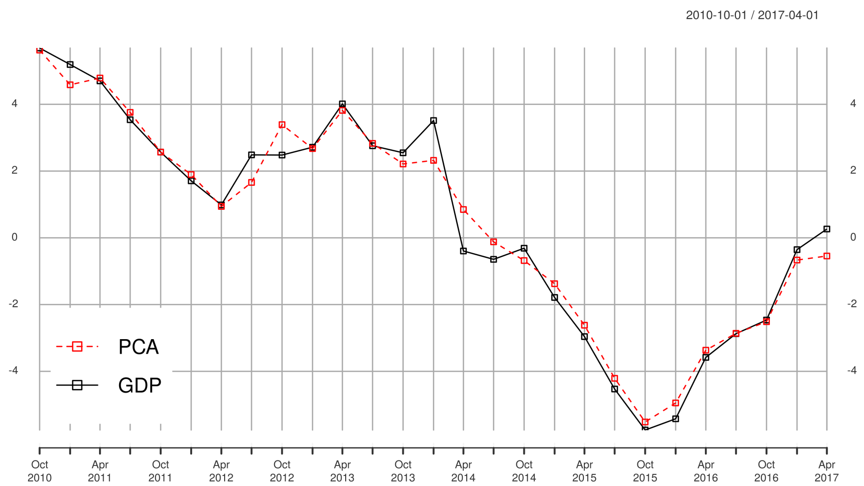
Table 2: Bridge equation with one Factor

	Dependent variable:
	GDP
Factor 1	0.223***
	(0.009)
Constant	−0.075
	(0.104)
Observations	52
R^2	0.929
Adjusted R^2	0.927
Residual Std. Error	0.753 (df = 50)
F Statistic	651.261*** (df = 1; 50)
Note:	*p<0.1; **p<0.05; ***p<0.01

in Table 2, the out-of-sample forecast of GDP annual rates of change (YoY) in the second quarter of 2017 was -0.36% when, in fact, the observed value was 0.26%.

We performed an out-of-sample evaluation exercise employing panels built with the stylized calendar. The model is reproduced for a period of 27 quarters, from the fourth quarter of 2010 to the second quarter of 2017. Figure 1 presents out-of-sample estimates transformed into (YoY) so that we could better visualize the results. The prediction horizon in this model is of 9 days, i.e., the predictions are made nine days before the release of GDP.

Figure 1: Out-of-sample PCA estimates



We calculated the Root Mean Squared Forecast Error (RMSFE) as a single measure of prediction quality, according to the equation:

$$RMSFE = \sqrt{\sum_{t=1}^{T} \frac{(\hat{y}_t - y_t)^2}{T}} \tag{27}$$

where y_t is the observed YoY GDP and $\hat{y}_t$ is the prediction of y_t. We found an $RMSFE$ equals to 0.51%.

As discussed in the literature review, the jagged edge problem does not allow real-time prediction using all variables, unless the panel is already complete. Thus, only one forecast is produced for each quarter, 142 days after it begins. Giannone et al. (2008) overcome this difficulty by proposing the two-step estimation method, which allows to re-estimate the forecast whenever new data is added to the information set.

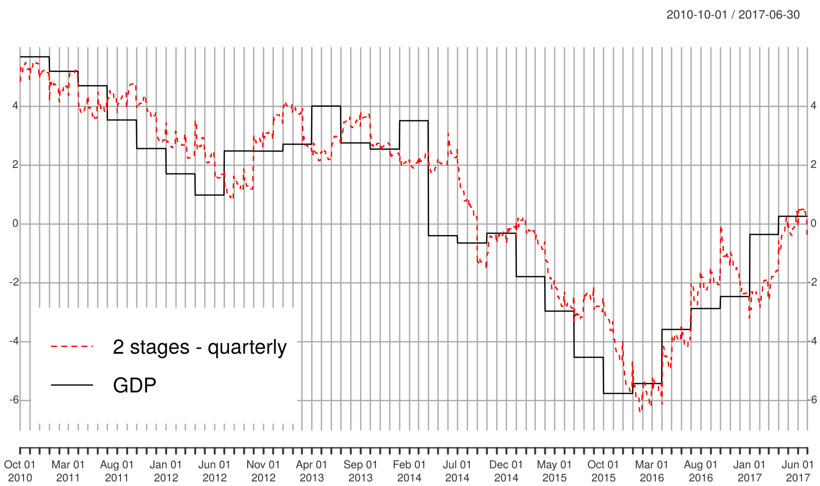
3.2 Two-step Estimation

We estimate the model as in Giannone et al. (2008) considering that the input variables are aggregated to a quarterly frequency. Then, the factors represent quarterly units. Regarding the specification of the model, following Giannone et al. (2008), we use the one that has the best out-of-sample performance for our database: 2 dynamic factors, 2 static factors, and a VAR of degree 1 for the loadings ($q = 2$, $r = 2$ and $p = 1$). In this case, the Equation (6) is represented as follows.

$$\begin{bmatrix} f_{1,t} \\ f_{2,t} \end{bmatrix} - \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} f_{1,t-1} \\ f_{2,t-1} \end{bmatrix} = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} u_t$$

In the model estimated via PCA, only one forecast is computed for each quarter, but in the two-step estimation, a forecast is obtained for each day whenever new information is disclosed. Figure 2 presents the YoY GDP, as well as the forecasts within each quarter. This graph is equivalent to the one shown in the Figure 1, but with more than one GDP estimate for of each quarter. This is exactly the nowcasting: predictions are calculated for the current quarter.

Figure 2: Out-of-sample two-step estimates
Quarterly Factors

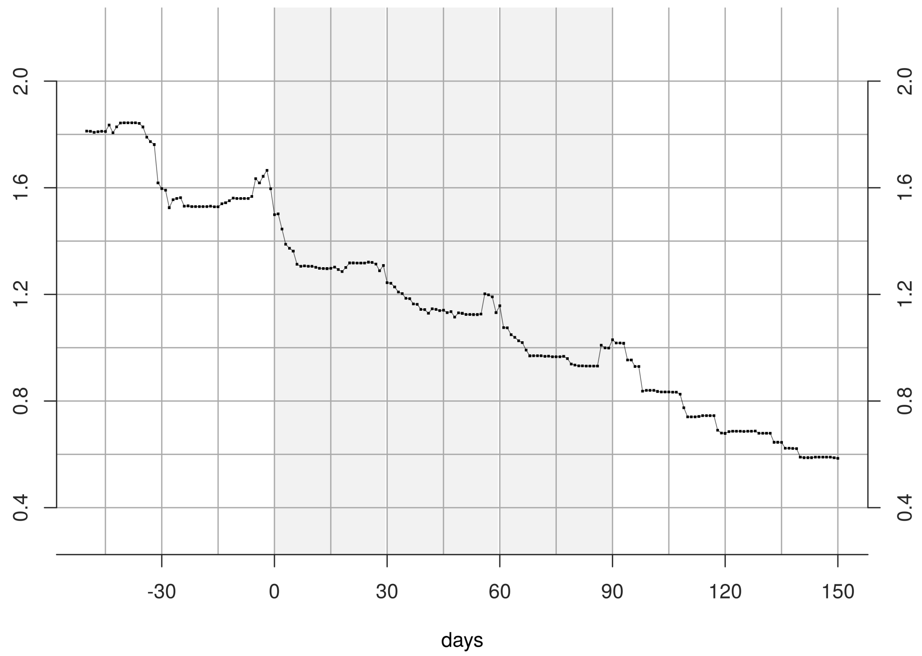


Not all quarterly information is available at the reference quarter. In fact, as shown in the last column of the appendix table, much of the information about the quarter is released after the quarter is over. The RMSFE is calculated for each vintage v following a variation of Equation (27)

$$RMSFE_v = \sqrt{\sum_{t=1}^{T} \frac{(\hat{y}_{t|v} - y_t)^2}{T}} \tag{28}$$

where $\hat{y}_{t|v}$ is the y_t prediction for vintage v. We construct Figure 3 aggregating these measures. The highlighted areas indicate the months of the reference quarter. RMSFE percentages are shown in the vertical axis, and the horizontal axis displays the time in days, being 1 the first day of the reference quarter.

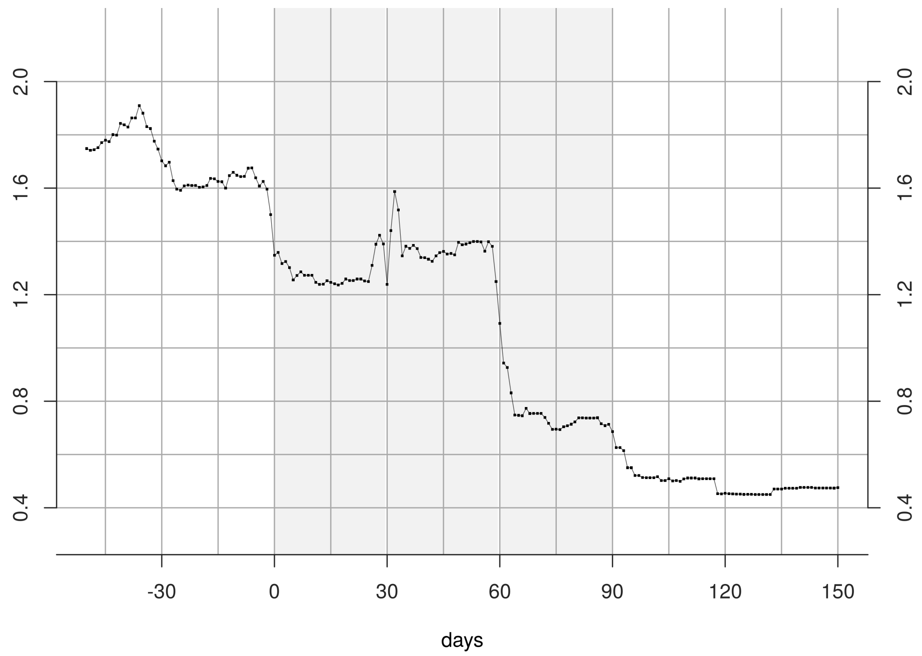
Figure 3: RMSFE two-step estimates
Quarterly Factors



As expected, the RMSFE generally drops with time and the expansion of the information set. One significant result in the nowcasting literature is that the quality of new information is higher the smaller the RMSFE. However, the quarterly quantities are updated only when the data of the last month of the quarter is disclosed. When new data is missing, the information available for the first two months is not used. In spite of that, the data of these two months can be used to improve nowcasts, as proposed by Bańbura and Rünstler (2011). They calculate the monthly factors from these data. Next, they use Equation (25) to obtain the quarterly factor which is in accordance to Equation (26).

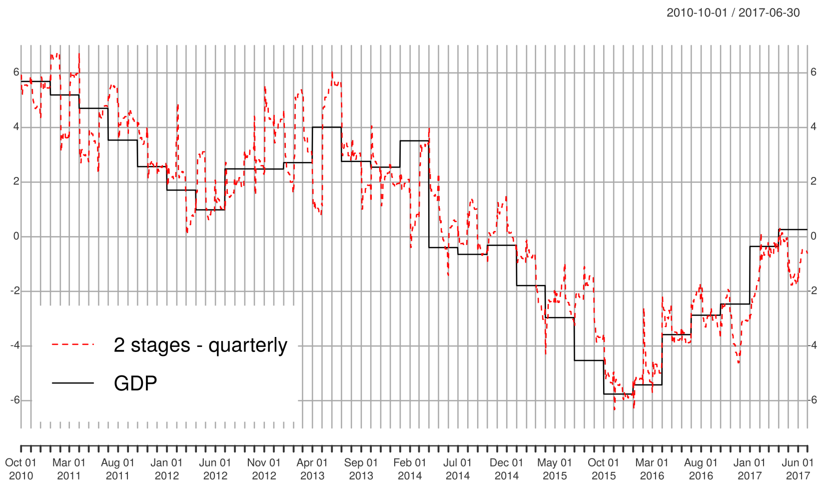
We reproduce Bańbura and Rünstler (2011) procedure and also calculate the RMSFE for each forecast horizon as shown in Figure 4. We can see that RMSFE decreases significantly within the quarter, more specifically at the beginning of the third month of the quarter, when the industrial production of the first month is disclosed. The GDP of the previous quarter is released a few days later, which helps to explain this result.

Figure 4: RMSFE two-step estimates
Monthly Factors



This sharp fall in RMSFE can also be observed in the Figure 5, which is equivalent to Figure 2 but shows forecasts of the monthly model. These forecasts are more volatile than those of the previous model. Predictions are affected because revisions within the quarter are more significant.

Figure 5: Out-of-sample two-step estimates
Monthly Factors



4 Professional Forecasts Combination

The Brazilian Central Bank (BCB) coordinates the *Focus Survey* of professional forecasters in order to unveil the expectations about some macroeconomic variables. This survey is conducted every workday, and the individuals are free to answer it. Nonetheless, there are incentives to do so. The first is the top five ranking: the five better forecasters of some variables have their names published, and it helps to promote their businesses. Another incentive is that the respondents can be invited to participate in BCB quarterly meetings and express their opinions about economic conjuncture. The database has approximately 250 registered individuals, but only 100 regularly update their forecasts. The BCB does not disclose information about each individual. Instead, it reveals some descriptive statistics of all forecasts in the survey, such as average, standard deviation, and so on.[12]

Regarding the GDP forecast, it is possible to find historical data since the third quarter of 2001 in the BCB online database. To the empirics, we select 62 quarters, from the third quarter of 2001 up to the fourth of 2016.

[12]For more details about the Focus Survey see Marques et al. (2013) and Carvalho and Minella (2012).

4.1 Theoretical Framework

This section follows Issler and Lima (2009), Gaglianone and Issler (2015), and Gaglianone et al. (2017) to produce optimal forecasts with a combination of individual forecasts. Here, we present and reproduce the *extended Bias-Corrected Average Forecast* (eBCAF). Denote $f_{i,t}^h$, the forecast informed to the *Focus Survey* made for period t, in the horizon h, by individual i:

$$f_{i,t}^h = k_i^h + \beta_i^h \mathbb{E}_{t-h}[y_t] + \varepsilon_{i,t}^h \tag{29}$$

where y_t is a weakly stationary univariate process (that, in this case, is the monthly difference of seasonally adjusted GDP), h is the forecast horizon, k_i^h and β_i^h are the level and slope biases for each individual respectively, $\varepsilon_{i,t}^h$ is the idiosyncratic specification error for an individual i at horizon h with zero mean. The optimal forecast with MSE loss function is $\mathbb{E}_{t-h}[y_t]$, the conditional expectation given the information set available at $t-h$. An individual j model with optimal forecast has no bias, then $k_j^h = 0$ and $\beta_j^h = 1$.

Let N be the total of respondents, then $i = 1, .., N$. The cross-sectional average of Equation (29) can be stated as

$$\sum_{i=1}^{N} f_{i,t}^h = \sum_{i=1}^{N} k_i^h + \sum_{i=1}^{N} \beta_i^h \mathbb{E}_{t-h}[y_t] + \sum_{i=1}^{N} \varepsilon_{i,t}^h \tag{30}$$

Since $\mathbb{E}_{t-h}[y_t]$ is common for all models, it follows that

$$\bar{f}_{.,t}^h = \bar{k}^h + \bar{\beta}^h \mathbb{E}_{t-h}[y_t] + \varepsilon_{.,t}^h \tag{31}$$

According to Issler and Lima (2009), it is possible to decompose y_t in $\mathbb{E}_{t-h}[y_t]$ and η_t^h, an unforecastable component such that $\mathbb{E}_{t-h}[\eta_t^h] = 0$

$$y_t = \mathbb{E}_{t-h}[y_t] - \eta_t^h \tag{32}$$

From the Equations (31) and (32)

$$\bar{f}^{h}_{.,t} = \bar{k}^{h} + \bar{\beta}^{h} y_t + \underbrace{\bar{\beta}^{h}\eta^{h}_{t} + \varepsilon^{h}_{.,t}}_{u^h_t} \tag{33}$$

Gaglianone and Issler (2015) show that, under suitable assumptions, $\mathbb{E}_{t-h}[u^h_t] = 0$. If there is a set of valid instruments z_{t-s}, such that $dim(z_{t-s}) \geq 2$, where $s \geq h$, and it belongs to the information set available in $t-h$, we obtain:

$$\mathbb{E}[u^h_t z_{t-s}] = \mathbb{E}[\mathbb{E}_{t-h}[u^h_t z_{t-s}]] = \mathbb{E}[z_{t-s}\mathbb{E}_{t-h}[u^h_t]] = \mathbb{E}[z_{t-s}0] = 0 \tag{34}$$

The first equality comes from the law of iterated expectations. We can write the following equation for each horizon:

$$\mathbb{E}[(\bar{f}^{h}_{.,t} - \bar{k}^{h} - \bar{\beta}^{h} y_t) z_{t-s}] = 0 \tag{35}$$

Gaglianone et al. (2017) show that it is possible to exploit all possible moments expressed by the equation above by pooling the horizons $h = h_1, h_2, ..., h_H$, when $s \geq h_H$ and the instruments belong to the information set available in $t - h_H$.[13]

$$\mathbb{E}\left[\begin{pmatrix} \bar{f}^{h_1}_{.,t} - \bar{k}^{h_1} - \bar{\beta}^{h_1} y_t \\ \bar{f}^{h_2}_{.,t} - \bar{k}^{h_2} - \bar{\beta}^{h_2} y_t \\ \cdots \\ \bar{f}^{h_H}_{.,t} - \bar{k}^{h_H} - \bar{\beta}^{h_H} y_t \end{pmatrix} \otimes z_{t-s}\right] = 0 \tag{36}$$

Using the GMM estimates Gaglianone et al. (2017) show that when $N \to \infty$ and $T \to \infty$

$$N^{-1} \sum_{i=1}^{N} \frac{f^h_{i,t} - \hat{\bar{k}}^h}{\hat{\bar{\beta}}^h} \quad \xrightarrow{p} \quad \mathbb{E}_{t-h}[y_t] \tag{37}$$

Equation (37) immediately suggests a combination of forecasts that should asymptotically outperform the simple average when there is bias.

[13] Unfortunately, this estimation is not feasible when we need more than 4 instruments, because we have few observations, resulting in a singular system.

4.2 Data and Empirical Results

For each observation of GDP, Focus Survey registers past forecasts for more than 252 business days. We could choose among any of these horizons, but we selected only six days to test the bias. We elected the days when important data on the reference quarter are released. We consider the six days when the Monthly Industrial Production Survey (PIM) and the Monthly Retail Sales Survey (PMC) are published. The implicit hypothesis behind this selection is that agents revise their expectations after observing such disclosures.[14]

So the first horizon is the day when the Industrial Production (IP) of the first month of the reference quarter is announced, the second horizon is the day when Retail Sales of the first month of the quarter is disclosed, and so on.

The format of *Focus Survey* data is YoY but, by construction, this measure generates serial autocorrelation. To overcome this difficulty, we transform the data to the first monthly difference. We explain this conversion below.

Let Y_t be the time series that represents the seasonally adjusted GDP index and $y_t = (\frac{Y_t}{Y_{t-4}} - 1) \times 100$ be the YoY rate. We want to find Y_t^h given y_t^h. As Y_{t-4} is known in t, we have the following expression

$$Y_t^h = \left(\frac{y_t^h}{100} + 1\right) \times Y_{t-4} \tag{38}$$

We find this measure for those six forecast horizons mentioned above and take the first difference, which we presume is stationary

$$\Delta Y_t^h = Y_t^h - Y_{t-1}^h \tag{39}$$

We lose five observations because of these transformations, four due to the YoY conversion and one because of the first difference. Note that depending on the forecast horizon (h), Y_{t-1} is known. Consequently there exists $\hat{h}$ such that for all $h > \hat{h}$ we have $Y_{t-1}^h = Y_{t-1}$. In this exercise, it happens for horizons 5 e 6.

To verify if this transformation is stationary as expected, we propose the augmented Dickey-Fuller and Phillips-Perron tests. The results are shown in the Table 3. In both tests, the null hypothesis of unit root is rejected for all series at 99% of confidence.

[14]These time series are published monthly by IBGE in the *Pesquisa Industrial Mensal* (PIM) survey and *Pesquisa Mensal do Comércio* (PMC). In the period under study, the PIM is disclosed, on average, 30 days after the end of the reference month. The PMC, 45 days. Only in the first and second horizons GDP forecasts are made within the quarter itself.

Table 3: Unit Root tests

	DF test	PP test
GDP	−3.687	−4.984
Fore. Horiz. 1	−3.164	−6.037
Fore. Horiz. 2	−3.335	−6.018
Fore. Horiz. 3	−3.576	−6.165
Fore. Horiz. 4	−3.531	−6.066
Fore. Horiz. 5	−6.010	−4.265
Fore. Horiz. 6	−6.110	−4.270

Note1: Critic value are -2.6 for 1%, -1.95 for 5% and -1.61 for 10%.
Note2: For P.P. test the lag is 10.

We estimate the parameters of Equation (35) by GMM for each horizon. As instruments, we use the first difference of the seasonally adjusted monthly industrial production and the seasonally adjusted monthly retail sales volume variation. The equations' variables are quarterly, and proposed instruments are monthly. To overcome this difficulty, we use the aggregation procedures presented in the appendix.

Table 4 describes the results of six these estimations. Each has a different dependent variable, the cross-sectional average of the *Focus Survey* forecasts registered in the respective horizon.

Table 4: One GMM for each horizon

	Dependent variable:					
	Focus forecast					
	pim1	pmc1	pim2	pmc2	pim3	pmc3
	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.656***	0.896***	0.992***	0.976***	0.523***	0.593***
	(0.132)	(0.113)	(0.101)	(0.108)	(0.048)	(0.045)
Constant	0.357*	−0.038	−0.056	−0.008	−0.373***	−0.571***
	(0.215)	(0.157)	(0.144)	(0.126)	(0.109)	(0.102)
J-test	5.833	2.649	3.065	8.275	3.484	6.711
P-value	0.559	0.915	0.879	0.309	0.837	0.46
Wald test	7.363	1.596	0.346	0.109	120.436	129.713
P-value	0.013	0.225	0.421	0.473	0	0
Observations	57	57	57	57	57	57

Note: *p<0.1; **p<0.05; ***p<0.01
instruments: lag 1, 2, 3 e 4 of PIM and PMC quarterly aggregated

The J Sargan-Hansen tests show evidence that all models are well specified.

Following Gaglianone et al. (2017) we carry out a joint Wald test for the coefficients, where under the null hypothesis the linear coefficient is equal to 0 and the angular coefficient 1. It suggests that there is evidence of bias in the *Focus Survey* for GDP in horizons 1, 5, and 6. For such horizons, a bias correction is proposed following Issler and Lima (2009) and Gaglianone et al. (2017).

Table 5 shows the RMSFE in percentage for all horizons. The mean is computed for 57 quarters, from the fourth quarter of 2002 to the fourth quarter of 2016. Unfortunately, we do not have enough observations to estimate the model and evaluate the out-of-sample performance, at the same time.

Table 5: RMSFE eBCAF and Focus

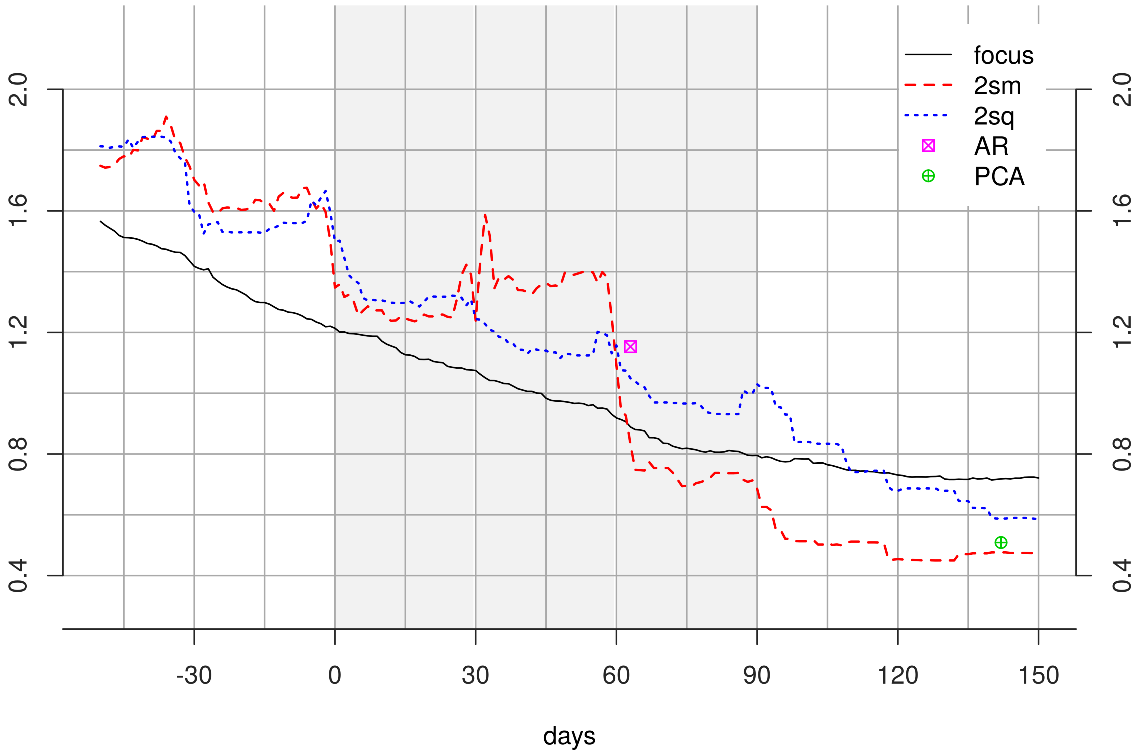
Horizon	eBCAF	focus
1	2.53	1.98
2	—	1.89
3	—	1.80
4	—	1.69
5	1.94	1.46
6	1.68	1.40

Looking at Table 5, we see that the model correction eBCAF appears to perform worse than the simple average of forecasters. This result may be explained by low power estimates once asymptotic results are not guaranteed with few observations. Another explanation is that the sample contains 2008 global crisis when the projection errors are usually more significant. Despite the evidence of possible bias in the average of *Focus Survey* predictions, we decided not to use the bias correction methodology because it delivers forecasts with larger RMSFE.

5 Models Comparison

In this section, we compare all the models proposed in this work. Figure 6 shows a comparison of RMSFEs. The vertical axis represents the values in percentages and the horizontal axis measures time in days where 1 is the first day of the quarter. The labels *focus*, *2sm*, *2sq*, *PCA*, and *AR* are respectively associated with the Focus Survey average, the two-step monthly and quarterly forecasts, principal components forecasts and an AR model as a benchmark.[15] The AR and PCA forecasts are estimated once a quarter, the former when the GDP of the previous quarter is available and the last when all variables in the dataset have been disclosed.

Figure 6: RMSFE: all models comparison



The decline of Focus Survey's prediction errors is smoother than it is for the two-step models. This may happen because the first is an average of predictions, so it tends to be smoother than other individual forecasts. In addition, forecasts are not updated daily by all survey participants, meaning that predictions are more persistent. Focus Survey seems to behave better than the other models until the end of the second month of the quarter when the two-step model with monthly factors starts performing better on average.

As noticed, the models are estimated for different horizons and, consequently,

[15]The AR benchmark model is a SARIMA (4,1,0)(1,1,0):

$$(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \phi_4 L^4)(1 - L)(1 - L^4)y_t = \varepsilon_t \tag{40}$$

information sets are different. For a fair evaluation, their performance must be compared within the same horizon. We choose then the days 63 and 142 for which the models AR and PCA are respectively estimated. In table 6, we present the RMSFE of each model for these days.

Table 6: RMSFE comparison

model	day 63	day 142
focus	0.91	0.72
2sq	1.07	0.59
2sm	0.93	0.48
PCA	—	0.51
AR	1.15	—

note: In percentage values.

Table 7: Diebold Mariano Test

	day 63	day 142
2sq	0.418	0.320
2sm	0.707	0.022
PCA	—	0.088
AR	0.171	—

note: p-values of two sided test in a quadratic loss function.

We compare the models' performance with the average Focus Survey using the Diebold and Mariano test. The results are presented in Table 7. It is not possible to reject that the three estimated models perform equal to the Focus Survey for the day 63. In contrast, the two-step monthly prediction can be considered better than the Focus Survey for day 142 at 95% confidence level and the PCA forecast, at a 90%. [16]

[16] Another forecast comparison tests are proposed in Giacomini and Rossi (2010) and can be used in unstable environments.

6 Conclusion

In this work, we reproduced three estimation approaches of *exact* Dynamic Factor Models. The first is based on the of Principal Components Analysis, following Stock and Watson (2002a) and Stock and Watson (2002b). The second relies on the groundwork laid by Giannone et al. (2008) who proposed a two-step estimation method and used quarterly inputs. The last employed monthly series, an alternative proposed by Bańbura and Rünstler (2011).

The dataset is composed of 61 monthly series plus the GDP, which is quarterly. The main technique we used to aggregate monthly data was devised by Mariano and Murasawa (2003), but we also employed some other well-known methods. In order to perform a real-time out-of-sample evaluation, we created a proxy information set vintage for each day for 27 quarters, from the fourth quarter of 2010 to the fourth quarter of 2016.

The nowcast estimated by the two-step procedure with monthly variables performs better than the average predictions of Focus Survey for recent periods. This result depends on the selected variables and the structure proposed for the DFM. Also, following the methodology proposed by Issler and Lima (2009) and Gaglianone et al. (2017) we showed evidence that the average of Focus Survey is biased for the GDP in some cases. In future works, we can test for the robustness of these results by changing the dataset, estimation techniques, generator process structure, or even try a combination of models following, for instance, Granger and Ramanathan (1984).

Moreover, the empirical results indicated that the data for the reference quarter improve the precision of the nowcast since the prediction error decreases with time. It suggests that the information flow must be taken into account and several estimates can be obtained as data become available. It means that the economic activity can be monitored every day in real-time. Another suggestion for further work is to use intra-daily data in nowcasting models as in Andreou et al. (2013). Lastly, we demonstrated that these models are feasible and can be implemented by market participants and policymakers to monitor the current state of the economy.

References

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

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, 27(4):417–427.

Bai, J., Ng, S., et al. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2):89–163.

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. *Handbook of economic forecasting*, 2:193–224.

Banbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. *Oxford Handbook on Economic Forecasting.*

Bańbura, M. and Rünstler, G. (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, 27(2):333–346.

Barhoumi, K., Benk, S., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Rünstler, G., Ruth, K., and Nieuwenhuyze, C. V. (2008). Short-term forecasting of GDP using large monthly datasets–a pseudo real-time forecast evaluation exercise.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics*, 120(1):387–422.

Bragoli, D., Metelli, L., and Modugno, M. (2015). The importance of updating: Evidence from a Brazilian nowcasting model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2015/1.

Carvalho, F. A. and Minella, A. (2012). Survey forecasts in Brazil: a prismatic assessment of epidemiology, performance, and determinants. *Journal of International Money and Finance*, 31(6):1371–1391.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models. *Econometrica: Journal of the Econometric Society*, 51(5):1305–1323.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of economic literature*, 49(1):72–100.

Cusinato, R., Minella, A., and Júnior, S. (2010). Hiato do produto e pib no brasil: uma análise de dados em tempo real. *Trabalhos para discussão*, (203).

Dahlhaus, T., Guénette, J.-D., and Vasishtha, G. (2017). Nowcasting BRIC+M in real time. *International Journal of Forecasting*, 33(4):915–935.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. OUP Oxford.

Engle, R. and Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76(376):774–781.

Evans, M. D. (2005). Where are we now? real-time estimates of the macroeconomy. *International Journal of Central Banking*, 1(2).

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, 25(5):1319–1347.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The review of Economics and Statistics*, 82(4):540–554.

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: representation theory. *Econometric theory*, 17(6):1113–1141.

Gaglianone, W. P. and Issler, J. V. (2015). Microfounded forecasting. *Ensaios Economicos*, (766).

Gaglianone, W. P., Issler, J. V., and Matos, S. M. (2017). Applying a microfounded-forecasting approach to predict brazilian inflation. *Empirical Economics*, 53(1):137–163.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

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

Giacomini, R. and Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Giannone, D., Reichlin, L., and Sala, L. (2004). Monetary policy in real time. *NBER macroeconomics annual*, 19:161–200.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Granger, C. W. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204.

Hausman, J. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *Journal of Economic perspectives*, 15(4):57–67.

Issler, J. V. and Lima, L. R. (2009). A panel data approach to economic forecasting: The bias-corrected average forecast. *Journal of Econometrics*, 152(2):153–164.

Issler, J. V. and Notini, H. H. (2016). Estimating Brazilian monthly GDP: a state-space approach. *Revista Brasileira de Economia*, 70(1):41–59.

Lawley, D. and Maxwell, A. (1962). Factor analysis as a statistical method. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 12(3):209–229.

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

Marques, A. B. C. et al. (2013). Central bank of brazil's market expectations system: a tool for monetary policy. *Bank for International Settlements*, 36:304–324.

Murphy, K. M. and Topel, R. H. (2002). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, 20(1):88–97.

Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review*, pages 221–247.

Quah, D. and Sargent, T. J. (1993). A dynamic index model for large cross sections. In *Business cycles, indicators and forecasting*, pages 285–310. University of Chicago Press.

Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, 1:145–168.

Spearman, C. (1904). General Intelligence objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford Handbook on Economic Forecasting*.

Stock, J. H. and Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*, 4:351–394.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

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

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. *Handbook of economic forecasting*, 1:515–554.

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

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3):385–400.

7 Appendix

7.1 Time Aggregation

An important issue that arises when we are working with many time series is that they may be at different frequencies. The major nowcasting models with a single frequency, transforming the data using the methods we describe in this section.

Let Y_t^M be the level of a monthly variable and Y_t^Q its quarterly representation. The quarter is always indexed by its last month. If the variable represents a stock, then

$$Y_t^Q = \begin{cases} Y_t^M & \text{if t is the last month of the quarter} \\ 0 & \text{otherwise} \end{cases}$$

if the variable represents a flow, we have

$$Y_t^Q = \begin{cases} Y_t^M + Y_{t-1}^M + Y_{t-2}^M & \text{if t is the last month of the quarter} \\ 0 & \text{otherwise} \end{cases}$$

Consider now the first difference of the monthly variable y_t^M and its quarterly correspondence y_t^Q. For stocks we have,

$$\begin{aligned} y_t^Q &= Y_t^Q - Y_{t-3}^Q = Y_t^M - Y_{t-3}^M = Y_t^M - Y_{t-1}^M + Y_{t-1}^M - Y_{t-2}^M + Y_{t-2}^M - Y_{t-3}^M = \\ &= y_t^M + y_{t-1}^M + y_{t-2}^M \end{aligned}$$

and flows,

$$\begin{aligned} y_t^Q &= Y_t^Q - Y_{t-3}^Q = Y_t^Q - Y_{t-1}^Q + Y_{t-1}^Q - Y_{t-2}^Q + Y_{t-2}^Q - Y_{t-3}^Q = \\ &= Y_t^M + Y_{t-1}^M + Y_{t-2}^M - (Y_{t-1}^M + Y_{t-2}^M + Y_{t-3}^M) + (Y_{t-1}^M + Y_{t-2}^M + Y_{t-3}^M) - \\ &- (Y_{t-2}^M + Y_{t-3}^M + Y_{t-4}^M) + (Y_{t-2}^M + Y_{t-3}^M + Y_{t-4}^M) - (Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M) = \\ &= y_t^M + 2y_{t-1}^M + 3y_{t-2}^M + 2y_{t-3}^M + y_{t-4}^M \end{aligned}$$

Finally, let x_t^M be the difference of the log of the monthly variable and x_t^Q its quarterly correspondent. For a stock variable,

$$x_t^Q = log(Y_t^Q) - log(Y_{t-3}^Q) = x_t^M + x_{t-1}^M + x_{t-2}^M$$

For flows, we use the approximation of Mariano and Murasawa (2003), which combines rates in a linear way:

$$
\begin{aligned}
x_t^Q &= log(Y_t^Q) - log(Y_{t-3}^Q) = \\
&= log(Y_t^Q) - log(Y_{t-1}^Q) + log(Y_{t-1}^Q) - log(Y_{t-2}^Q) + log(Y_{t-2}^Q) - log(Y_{t-3}^Q) = \\
&= log(Y_t^M + Y_{t-1}^M + Y_{t-2}^M) - log(Y_{t-1}^M + Y_{t-2}^M + Y_{t-3}^M) + log(Y_{t-1}^M + Y_{t-2}^M + Y_{t-3}^M) - \\
&- log(Y_{t-2}^M + Y_{t-3}^M + Y_{t-4}^M) + log(Y_{t-2}^M + Y_{t-3}^M + Y_{t-4}^M) - log(Y_{t-3}^M + Y_{t-4}^M + Y_{t-5}^M) \approx \\
&\approx \frac{1}{3}\Big[log(Y_t^M) + log(Y_{t-1}^M) + log(Y_{t-2}^M) - (log(Y_{t-1}^M) + log(Y_{t-2}^M) + log(Y_{t-3}^M)) + \\
&+ (log(Y_{t-1}^M) + log(Y_{t-2}^M) + log(Y_{t-3}^M)) - (log(Y_{t-2}^M) + log(Y_{t-3}^M) + log(Y_{t-4}^M)) + \\
&+ (log(Y_{t-2}^M) + log(Y_{t-3}^M) + log(Y_{t-4}^M)) - (log(Y_{t-3}^M) + log(Y_{t-4}^M) + log(Y_{t-5}^M))\Big] = \\
&= \frac{1}{3}\Big[x_t^M + 2x_{t-1}^M + 3x_{t-2}^M + 2x_{t-3}^M + x_{t-4}^M\Big]
\end{aligned}
$$

This approximation is obtained considering that the arithmetic mean approximately equals to the geometric average:

$$
\frac{1}{3}Y_t^Q = \frac{1}{3}\Big[Y_t^M + Y_{t-1}^M + Y_{t-2}^M\Big] \approx \sqrt[3]{Y_t^M Y_{t-1}^M Y_{t-2}^M} \tag{41}
$$

7.2 Data set

Here we present the dataset used. All series were collected in a single day (November 23, 2017). When we reconstruct the data available for the past vintages, we do not consider data revisions. We suppose that the data disclosed in the past was equal to the one already obtained. We also consider that they become available in regular periods. These assumptions are equivalent to those in Giannone et al. (2008).

We suppose that ultra-realistic dataset vintages could change some results. Nevertheless, they are not available for Brazil in the time we write this paper, and the construction of it would a separate research work. Cusinato et al. (2010) show that some revisions of Brazilian GDP were significative. Croushore (2011) presents a rich survey about the ways the real data structure can be exploited.

Table 8: Data set

	Category	Name	Source	Original	Transformation	Publication delay in days
1	General Index	GDP	IBGE	Index	diff diff 12	63
2	General Index	Central Bank Economic Activity Index – IBCBR	BCB	Index	diff diff 12	52
3	General Index	GDP monthly - current prices (R$ million)	BCB	Index	diff diff 12	48
4	Prices	Broad National Consumer Price Index (IPCA)	IBGE	Monthly var %	diff 12	11
5	Prices	National Consumer Price Index (INPC)	IBGE	Monthly var %	diff 12	10
6	Prices	General Price Index-Market (IGP-M)	FGV	Monthly var %	diff 12	−2
7	Prices	General Price Index-Domestic Supply (IGP-DI)	FGV	Monthly var %	diff 12	8
8	Prices	Consumer Price Index-Brazil (IPC)	FGV	Monthly var %	diff 12	0
9	Prices	Wholesale Price Index (IPA)	FGV	Monthly var %	diff 12	0
10	Prices	Consumer Price Index-São Paulo (IPC-Fipe)	FIPE	Monthly var %	diff 12	17
11	Labor	Registered Employees Index	MTE	Index	diff diff 12	21
12	Labor	Hours worked in production (June/1994=100)	Fiesp	Index	diff diff 12	33
13	Labor	Hours worked in industry (mean 2006=100)	CNI	Index	diff diff 12	32
14	Industry	Industrial Production, Total	IBGE	Index	diff diff 12	30
15	Industry	Industrial Production, Extractivism	IBGE	Index	diff diff 12	30
16	Industry	Industrial Production, Manufacturing	IBGE	Index	diff diff 12	30
17	Industry	Vehicles production (total)	Anfavea	Units	diff diff 12	6
18	Industry	Passenger cars and light commercial vehicles production	Anfavea	Units	diff diff 12	6
19	Industry	Truck production	Anfavea	Units	diff diff 12	6
20	Industry	Bus production	Anfavea	Units	diff diff 12	6
21	Industry	Installed Capacity Utilization (UCI)	CNI	Percentage	diff diff 12	32
22	Sales	Construction sales	IBGE	Index	diff diff 12	45
23	Sales	Retail sales	IBGE	Index	diff diff 12	45
24	Sales	Retail sales extended	IBGE	Index	diff diff 12	45
25	Sales	Vehicles	IBGE	Index	diff diff 12	45
26	Sales	Sales of factory authorized vehicle outlets - Sales (total)	Fenabrave	Units	diff diff 12	3
27	Sales	Sales of factory authorized vehicle outlets - Passenger cars sales	Fenabrave	Units	diff diff 12	3
28	Sales	Sales of factory authorized vehicle outlets - Light commercial cars sales Units	Fenabrave	Units	diff diff 12	3
29	Sales	Sales of factory authorized vehicle outlets - Trucks sales	Fenabrave	Units	diff diff 12	3
30	Sales	Sales of factory authorized vehicle outlets - Buses sales	Fenabrave	Units	diff diff 12	3
31	Energy	Electric energy consumption - Brazil – commercial	Eletrobras	Gwh	diff diff 12	22
32	Energy	Electric energy consumption - Brazil – residential	Eletrobras	Gwh	diff diff 12	22
33	Energy	Electric energy consumption - Brazil – industrial	Eletrobras	Gwh	diff diff 12	22
34	Energy	Electric energy consumption - Brazil – other	Eletrobras	Gwh	diff diff 12	22
35	Energy	Electric energy consumption - Brazil – total	Eletrobras	Gwh	diff diff 12	22

	Category	Name	Source	Original	Transformation	Publication delay in days
36	International	Exports (Fob) – Total	MDIC/Secex	US$	diff diff 12	8
37	International	Imports (Fob) – Total	MDIC/Secex	US$	diff diff 12	8
38	International	Imports (Fob) - Capital Goods	MDIC/Secex	US$	diff diff 12	8
39	International	Imports (Fob) - Consumption goods	MDIC/Secex	US$	diff diff 12	8
40	International	Current account - monthly – net	MDIC/Secex	US$	diff diff 12	50
41	Public Sector	Net public debt - Balances in c.m.u. (million) - Total - State governments	MF-STN	R$	diff diff 12	31
42	Public Sector	Primary Result of the Central Government - Primary Result of the National Treasury	MF-STN	R$	NONE	31
43	Monetary and Finance	Monetary base - Monetary base (working day balance average)	BCB	R$	diff diff 12	31
44	Monetary and Finance	International reserves - Liquidity concept – Total	BCB	US$	diff	31
45	Monetary and Finance	Money supply - M1 (working day balance average)	BCB	R$	diff diff 12	31
46	Monetary and Finance	Interest rate - Selic accumulated in the month	BCB	Percentage	diff diff 12	0
47	Monetary and Finance	Ibovespa - monthly percent change	BM&F Bovespa	Percentage	NONE	1
48	Others	Corrugated fiberboard producation	ABPO	Tonne	diff diff 12	5
49	Others	Steel production	BCB	Index	diff diff 12	20
50	Others	Toll road flow – Total	ABCR	Index	diff diff 12	10
51	Others	Toll road flow – light vehicles	ABCR	Index	diff diff 12	10
52	Others	Toll road flow – gross vehicles	ABCR	Index	diff diff 12	10
53	Survey	Level of Installed Capacity Utilization (NUCI)	FGV	Percentage	diff diff 12	−1
54	Survey	Industrial Confidence Index (ICI)	FGV	Index	diff diff 12	−1
55	Survey	Industrial Current Situation (ISA)	FGV	Index	diff diff 12	−1
56	Survey	Industrial Expectation Index (IE)	FGV	Index	diff diff 12	−1
57	Survey	Consumer Confidence Index (ICI)	FGV	Index	diff diff 12	−5
58	Survey	Consumer Current Situation (ISA)	FGV	Index	diff diff 12	−5
59	Survey	Consumer Expectation Index (IE)	FGV	Index	diff diff 12	−5
60	Survey	Entrepreneur Confidence Index (ICI)	FGV	Index	diff diff 12	2
61	Survey	Entrepreneur Current Situation (ISA)	FGV	Index	diff diff 12	2
62	Survey	Entrepreneur Expectation Index (IE)	FGV	Index	diff diff 12	2