Short-Run Forecasting of Argentine GDP Growth*

Maximo Camacho
Universidad de Murcia
mcamacho@um.es

Marcos Dal Bianco
BBVA Research, Argentina
marcos.dalbianco@bbva.com

Jaime Martínez-Martín
BBVA Research, Spain
j.martinez.martin@bbva.com

Abstract

This paper proposes a small-scale dynamic factor model to monitor the Argentine GDP in real time using economic data at mixed frequencies (monthly and quarterly). Our model not only produces a coincident index of the Argentine business cycle which is in striking accordance with the professional consensus and with the history of Argentine business cycle, but it also generates accurate short-run forecasts of Argentine GDP growth. By means of a simulated real-time empirical evaluation, we show that our model produces reliable backcasts, nowcasts and forecasts well before the official data is released.

Keywords: Real-time forecasting, Argentine GDP, business cycles, state-space models, mixed frequencies.

JEL Classification: C22, C53, E27, E32, E37.

*M. Camacho would like to thank CICYT (ECO2010-19830) for their financial support. All errors are our own responsibility. The views expressed in this paper are those of the authors and do not necessarily represent those of BBVA Research. Preliminar version, comments are welcomed. Corresponding authors: marcos.dalbianco@bbva.com and j.martinez.martin@bbva.com.

1 Introduction

The Gross Domestic Product (GDP) is the most important measure of the aggregate state of economic activity in any market economy. As such, it should be the most relevant business cycle indicator for policymakers and economic agents who must continuously take relevant real-time economic decisions. However, the quarterly GDP figures are typically published with important time lags, constituting a big problem for economic agents who need updated information in order to make a proper assessment of current and future macroeconomic conditions. Needless is to say that this problem is particularly acute in emerging countries, which usually face volatile business cycles and large publication delays in the relevant economic indicators.[1]

In Argentina, for instance, the GDP data of a given quarter is published about 10 weeks after the end of the corresponding quarter; clearly too late for being a useful indicator for real-time decisions. Economic agents (investors, policymakers, consumers) are hence forced to rely on other economic series which are available through the quarter to track the evolution of current GDP. However, those series are related to partial aspects of economic activity, are usually more volatile than GDP and often yield contradictory insights about how GDP is evolving through a given quarter. Moreover, it is difficult to use monthly series to forecast quarterly GDP since the models need to handle data with different frequencies. Hence, having an econometric model that can combine monthly and quarterly economic series to obtain a real-time measure of economic activity as an updating assessment tool for tracking quarterly GDP is of greatest interest. As a consequence, it comes as no surprise that research economists devote an increasing attention to develop econometric techniques for dealing with these shortcomings.

Under such a setting, a very useful small-scale factor model for building a coincident index of business cycle mixing monthly and quarterly series was initially developed by Mariano and Murasawa (2003) for the US economy, and later on refined for specific forecasting purposes by Camacho and Perez-Quiros (2010) for the Eurozone. Therefore, we follow both works using Argentine data in order to produce backcasts, nowcasts and short-run forecasts estimations of Argentine GDP growth. Thus, our model uses partial information from current economic situation mixing just a few monthly and quarterly indicators to obtain an accurate assessment of current and future Argentine GDP growth.

Our main results can be summarized as follows. First, we find a high performance of the coincident indicator as a business cycle indicator since it is in striking accord with the professional consensus of the history of the Argentine business cycle. Second, the percentage of the variance of

[1] For instance, see Lane (2003) and Aguiar and Gopinath (2007).

GDP growth that is explained by the model is 89%, indicating the high potential ability of the indicators used in the model to explain Argentine growth. Third, our pseudo real-time analysis shows that dynamic factor models clearly outperform univariate forecasts, especially when forecasting the next unavailable figure of GDP growth. This encourages real-time forecasters to back-check the bulk of monthly real and survey data which are published in the respective quarter before the next GDP release. Against this background, our model is able to produce accurate forecasts and therefore we strongly consider that our model is a valid tool to be used for short-term analysis.

This paper proceeds as follows. Section 2 briefly reviews the related literature. Section 3 outlines the model, shows how to mix frequencies, describes its dynamic properties along with the state space representation, and states the estimation properties. Section 4 contains data description and highlights the main empirical results, both in- and out-of-sample. Finally, Section 5 concludes and proposes several future lines of research.

2 Brief review of relevant related literature

The modern literature on business cycle estimation starts with the monthly coincident index of Stock and Watson (1989, 1991). They estimate a coincident index of economic activity as the unobservable factor in a dynamic factor model for four coincident indicators: industrial production, real disposable income, hours of work and sales, aiming to provide a formal probabilistic basis for Burns and Mitchell (1946) coincident and leading indicators. However, the dymamic factor model advocated by these authors exhibits three important drawbacks when it is used to monitor the economic activity in real-time. First, their method requires balanced panels, which precluded them from using data with mixed frequency or indicators with different publication delays Therefore, their model ignores the information contained in quarterly indicators such as real GDP, which is probably the main business cycle indicator. Although the index that they obtain is computed as linear combinations of meaningful economic indicators, the fact that it is not related to a particular variable of interest make it difficult to find an economic interpretation of its level or its reactions to shocks.

To face these drawbacks, Mariano and Murasawa (2003) proposed a coincident index of business cycle with the distinctive characteristic of blending indicators published both at monthly and quarterly frequency. They apply maximum likelihood factor analysis to the four monthly indicators, but since their methodology is able to handle mixing frequencies, they can also include real GDP as an additional fifth coincident indicator. Their coincident index accurately captures the NBER business cycle reference dates and shows very high statistical correlation with the Stock and Watson

(1991) coincident index. Moreover, their index has an economic interpretation as the common factor component in a (latent) monthly real GDP. A drawback of Mariano and Murasawa (2003) is that they do not explore the forecasting properties of their model.[2]

The forecasting analysis of these models is tackled later on by Camacho and Perez-Quiros (2010), who succesfully modified Mariano and Murasawa's model to compute short-term forecasts of the Eurozone GDP growth in real time. Their small-scale dynamic factor model is able to forecast the eurozone GDP growth at least as well as (and usually better than) professional forecasters. Further developments of their work are Camacho and Domenech (2012) for Spain, Camacho and Garcia-Serrador (2013) for the Euro area, and Camacho and Martinez-Martin (2012) for the USA.

This recent literature on short-run GDP growth forecasting is almost exclusively focused on developed economies. The related literature is very scarce for emerging countries in general and for Argentina in particular. To the best of our knowledge, there are only three similar attempts to our's in the literature. The first is Simone (2001), who constructs coincident and leading indicators of economic activity in Argentina. Although he proposes a useful contribution, he only uses quarterly data and he does not reach a reliable leading indicator for Argentine GDP. Second, there are two recent works by D'Amato, et al. (2011a, 2011b), who employ two techniques to produce predictions of current GDP growth within the quarter – which is called "nowcasting" – and one-quarter ahead forecast of GDP growth.[3] Third, Liu et al. (2011) estimate a large-scale factor model based on monthly data for nowcasting and forecasting. However, the related RMSE show traces of weak forecasting capacity.

3 The econometric model

3.1 Mixing frequencies

We use data at two frequencies, monthly and quarterly. To mix them, we consider all series as being of monthly frequency and treat quarterly data as monthly series. In this case, the monthly series are observed in the last month of the quarter and exhibit missing observations in the first two months of each quarter.

In particular, let G_t be the level of a quarterly flow variable that can be decomposed as the sum of three (usually unobserved) monthly values G_t^*. To avoid using a non-linear state-space model,

[2] Using these time series, Aruoba and Diebold (2010) examine the real-time performance of the common factor as a business cycle indicator.

[3] They do not present "backcasting" results, which are the estimation on a given quarter of the previous quarter rate of growth *before* they are published by the statistical agency, i.e. within the ten weeks of delay. Hence, we cannot compare our backcasting results with theirs.

we follow Mariano and Murasawa (2003) and approximate the arithmetic mean with the geometric mean.[4] Hence, the level of the variable can be written as

$$G_t = 3(G_t^* G_{t-1}^* G_{t-2}^*)^{1/3}. \tag{1}$$

Taking logs in both sides of this expression and computing the three-period differences for all t, we obtain

$$\triangle_3 \ln G_t = \frac{1}{3}(\triangle_3 \ln G_t^* + \triangle_3 \ln G_{t-1}^* + \triangle_3 \ln G_{t-2}^*). \tag{2}$$

Denoting the quarter-on-quarter growth rate $\triangle_3 \ln G_t = g_t$, the monthly-on-monthly growth rate $\triangle \ln G_t^* = g_t^*$, and after a little algebra, we obtain

$$g_t = \frac{1}{3}g_t^* + \frac{2}{3}g_{t-1}^* + g_{t-2}^* + \frac{2}{3}g_{t-3}^* + \frac{1}{3}g_{t-4}^*. \tag{3}$$

Accordingly, we express the quarter-on-quarter growth rate (g_t) as a weighted average of the monthly-on-monthly past growth rates (g_{t-i}^* , $i = 0, ..., 4$) of the monthly series.

3.2 Dynamic properties

We use the common assumption in factor modelling literature that the time series used in the model are the sum of two orthogonal components: a common component, x_t, which represents the overall business cycle conditions, and an idiosyncratic component, which refers to the particular dynamics of the series. The underlying business cycle conditions are assumed to evolve with $AR(p1)$ dynamics:

$$x_t = d_1^x x_{t-1} + ... + d_{p1}^x x_{t-p1} + \varepsilon_t^x, \tag{4}$$

where $\varepsilon_t^x = iN(0, \sigma_x^2)$.

For the sake of simplicitly, let us assume that we consider one quarterly indicator and one monthly indicator.[5] Starting from the quarterly indicator, let us assume that the evolution of its underlying monthly growth rates depends linearly on x_t and on the idiosyncratic dynamics, u_t^g, which evolve as an $AR(p2)$:

$$g_t^* = \beta_g x_t + u_t^g, \tag{5}$$

$$u_t^g = d_1^g u_{t-1}^g + ... + d_{p2}^g u_{t-p2}^g + \varepsilon_t^g, \tag{6}$$

[4] Aruoba et al. (2009) extended this analysis to include high frequency data using an exact algorithm, as opposed to the approximate algorithm of Mariano and Murasawa (2003). However, Aruoba et al. (2009) face the cost of assuming deterministic trends in the series.

[5] Extending the model to k_1 quarterly indicators and k_2 monthly indicators is straightforward.

where $\varepsilon_t^g = iN(0, \sigma_g^2)$. In addition, the evolution of the monthly indicator depends linearly on x_t and on the idiosyncratic component, whose dtnamics can be expressed in terms of autoregressive processes of $p3$ orders:

$$z_t = \beta_z x_t + u_t^z \tag{7}$$

$$u_t^z = d_1^z u_{t-1}^z + ... + d_{p2}^z u_{t-p2}^z + \varepsilon_t^z, \tag{8}$$

where $\varepsilon_t^z = iN(0, \sigma_z^2)$. Finally, all the shocks e_t, ε_t^y, and ε_t^y, are assumed to be mutually uncorrelated in cross-section and time-series dimensions.

Using the assumptions described below, this model can be easily stated in state-space representation and estimated – as further developed in the following section – by using the Kalman filter.

3.3 State-space representation

Let I_a be the identiy matrix of order a and let $0_{a\times b}$ be a $(a \times b)$ matrix of zeroes. For clarity, let us assume that $p1 = p2 = p3 = 1$. In addition, let us start by assuming that all variables are always observed at a monthly frequency. In this simplified version, the *measurement equation*

$$Y_t = Hh_t + E_t, \tag{9}$$

with $E_t \sim i.i.d.N\,(0, R)$, can be stated by defining

$$Y_t = (g_t, z_t)', \tag{10}$$

$$H = \begin{pmatrix} \frac{\beta_g}{3} & \frac{2\beta_g}{3} & \beta_g & \frac{\beta_g}{3} & \frac{2\beta_g}{3} & \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} & 0 \\ \beta_z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{11}$$

$$h_t = \left(x_t, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, u_t^g, u_{t-1}^g, , u_{t-2}^g, u_{t-3}^g, u_{t-4}^g, u_t^z\right)', \tag{12}$$

$$E_t = (0, 0)', \tag{13}$$

$$R = 0_{2\times 2}. \tag{14}$$

In the same way, the *transition equation*

$$h_t = Fh_{t-1} + \xi_t, \tag{15}$$

with $\xi_t \sim i.i.d.N\left(0, Q\right)$ can be stated by defining

$$F = \begin{pmatrix} F_1 & 0_{5\times5} & 0_{5\times1} \\ 0_{5\times5} & F_2 & 0_{5\times1} \\ 0_{1\times5} & 0_{1\times5} & d_1^z \end{pmatrix}, \tag{16}$$

$$F_1 = \begin{pmatrix} d_1^x & 0_{1\times4} \\ I_4 & 0_{4\times1} \end{pmatrix}, \tag{17}$$

$$F_2 = \begin{pmatrix} d_1^g & 0_{1\times4} \\ I_4 & 0_{4\times1} \end{pmatrix}, \tag{18}$$

$$\xi_t = \left(\varepsilon_t^x, 0, 0, 0, 0, \varepsilon_t^g, 0, 0, 0, 0, \varepsilon_t^z\right)', \tag{19}$$

$$Q = diag\left(\sigma_x^2, 0, 0, 0, 0, \sigma_g^2, 0, 0, 0, 0, \sigma_z^2\right)'. \tag{20}$$

The identifying assumption implies that the variance of the common factor, σ_x^2, is normalized to a value of one.[6]

3.4 Estimation

The estimation of the model would be by standard maximum likelihood by using the Kalman filter if all series were observable at the monthly frequency, as we assume so far. However, this assumption is quite restrictive since we are using time series of different length and we are mixing monthly data with quarterly data.

Mariano and Murasawa (2003) develop a framework to easily handle this issue. Following these authors, the unobserved cells can be treated as missing observations and maximum likelihood estimation of a linear Gaussian state-space model with missing observations can be applied straightforwardly after a subtle transformation of the system matrices. The missing observations can be replaced with random draws ϑ_t, whose distribution cannot depend on the parameter space that characterizes the Kalman filter.[7] Thus, the likelihood function of the observed data and that of the data whose missings are replaced by the random draws are equivalent up to scale. In particular, we assume that the random draws come from $N(0, \sigma_\vartheta^2)$. In addition, the measurement equation must be transformed conveniently in order to allow the Kalman filter to skip the missing observations when updating.

Let Y_{it} be the i-th element of the vector Y_t and R_{ii} be its variance. Let H_i be the i-th row of the matrix H which has ς columns and let $0_{1\varsigma}$ be a row vector of ς zeroes. The measurement

[6] This is a very standard assumption in factor models.

[7] Note that replacements by constants would also be valid.

equation can be replaced by the following expressions

$$Y_{it}^{*} = \begin{cases} Y_{it} & \text{if } Y_{it} \text{ observable} \\ \vartheta_t & \text{otherwise} \end{cases}, \tag{21}$$

$$H_{it}^{*} = \begin{cases} H_{i} & \text{if } Y_{it} \text{ observable} \\ 0_{1\varsigma} & \text{otherwise} \end{cases}, \tag{22}$$

$$E_{it}^{*} = \begin{cases} 0 & \text{if } Y_{it} \text{ observable} \\ \vartheta_t & \text{otherwise} \end{cases}, \tag{23}$$

$$R_{iit}^{*} = \begin{cases} 0 & \text{if } Y_{it} \text{ observable} \\ \sigma_{\vartheta}^{2} & \text{otherwise} \end{cases}. \tag{24}$$

According to this transformation, the time-varying state space model can be treated as having no missing observations so the Kalman filter can be directly applied to Y_t^*, H_t^*, E_t^*, and R_t^*.

The estimation of the model's parameters can be developed by maximizing the log-likelihood of $\{Y_t^*\}_{t=1}^{t=T}$ numerically with respect to the unknown parameters in matrices. Let $\widehat{\zeta}_{t|\tau}$ be the estimate of ζ_t based on information up to period τ. Let $P_{t|\tau}$ be its covariance matrix. The prediction equations are

$$\widehat{\xi}_{t|t-1} = F\widehat{\xi}_{t-1|t-1}, \tag{25}$$

$$P_{t|t-1} = FP_{t-1|t-1}F' + Q. \tag{26}$$

Hence, the predicted value of Y_t with information up to $t-1$, denoted $\widehat{Y}_{t|t-1}$ is

$$\widehat{Y}_{t|t-1} = H^{*}\widehat{\xi}_{t|t-1}, \tag{27}$$

and the prediction error is

$$\eta_{t|t-1} = Y_t^{*} - \widehat{Y}_{t|t-1} = Y_t^{*} - H^{*}\widehat{\xi}_{t|t-1}, \tag{28}$$

with covariance matrix:

$$\zeta_{t|t-1} = H^{*}P_{t|t-1}H^{*} + R_t^{*}. \tag{29}$$

The way missing observations are treated implies that the filter, through its implicit signal extraction process, will put no weight on missing observations in the computation of the factors.

In each iteration, the log-likelihood can be computed as

$$\log L_{t|t-1} = -\frac{1}{2}\ln\left(2\pi\left|\zeta_{t|t-1}\right|\right) - \frac{1}{2}\eta'_{t|t-1}\left(\zeta_{t|t-1}\right)^{-1}\eta_{t|t-1}. \tag{30}$$

It is worth noting that the transformed filter to handle missing observations has no impact on the model estimation. In that sense, the missing observations simply add a constant to the likelihood function of the Kalman filter process. Hence, the parameters that maximize the likelihood are achieved as if all the variables were observed.

The updating equations are:

$$\widehat{\xi}_{t|t} = \widehat{\xi}_{t|t-1} + P_{t|t-1} H_t^{+\prime} \left(\zeta_{t|t-1}\right)^{-1} \eta_{t|t-1}, \tag{31}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} H_t^{+\prime} \left(\zeta_{t|t-1}\right)^{-1} H_t^{+} P_{t|t-1}. \tag{32}$$

Therefore, missing observations are from the updating recursion.

Remarkably, computing short-term forecasts in real time from this model is straightforward. The reason is that one can regard the future values of the time series as a set of missing observations at the end of the sample periods. After the last observation, missing values are added to the data set and the Kalman accounts for the missing data which are replaced by forecasts. In particular, the k-period-ahead forecasts are

$$\widehat{Y}_{t+k|t} = H^{*} \widehat{\xi}_{t+k|t}, \tag{33}$$

where $\widehat{\xi}_{t+k|t} = F^k \widehat{\xi}_{t|t}$.

4 Empirical Results

4.1 Preliminary analysis of the data

The data employed in this paper, which were collected on September 15, 2012, spans the period from January 1993 to August 2012. Regarding the relatively wide potential set of indicators that could be used in the analysis, we only choose those that verify the following four properties. First, they must exhibit high statistical correlation with the GDP growth rate, which is the target series to be estimated and predicted. Second, for a given quarter they should refer to data of this quarter, which implies that they must be published before the GDP figure becomes available in the respective quarter. Third, they must be relevant in the model from both theoretical and empirical (statistical) points of view. Finally, they must be available in at least one third of the sample.

We started the analysis with the Argentinean version of the set of coincident economic indicators used in Aruoba and Diebold (2010): real quarterly GDP, monthly industrial production, quarterly employment, monthly real personal income, and real trade sales, which exhibit a strong (statistical) link with the GDP cycle. However, income was discarded as its loading factor was not statistically significant in favor of a Synthetic Indicator of Construction Activity, which exhibits higher correlation and a statistically significant loading factor. The potential enlargements of the data set were sequentially tested by including additional indicators such as Consumer Confidence index. However, the loading factors were not statistically significant and they were not included in the model.

The five indicators used in the empirical analysis and their respective publication delay are summarized in Table 1. All the variables are seasonally adjusted and are (weekly) stationary or transformed to be stationary.[8] Accordingly, the quarterly indicators enter in the model in quarterly growth rates while the monthly indicators enter in monthly growth rates. Before estimating the model, the variables are standardized to have zero mean and variance equal to one. Therefore, the final forecasts are computed by multiplying the initial forecasts of the model by sample standard deviation, and then adding the sample mean.

[Insert Table 1 about here]

4.2 In sample analysis

We show in this section the results obtained by the model outlined in Section 3. In Table 2 we present the estimated values for the factor loadings which reflect the degree to which variations in each observed variable are correlated with the latent factor. As observed, all variables show statiscally significant loading factors. Notably, Table 2 also shows that the percentage of variance of the actual Argentine GDP growth that is explained by the factor is very high, reaching about 90%.

[Insert Table 2 about here]

The empirical reliablity of the inferred factor as an Argetinean business cycle indicator is examined in Figure 1. Together with this series, the figure plots the corresponding growth rates of the Monthly Estimator of Economic Activity (EMAE because of its acronym in Spanish), which is a widely acepted proxy of Argentine GDP. According to this graph the evolution of the inferred factor is in striking accord with that of EMAE.

[Insert Figure 1 about here]

With the aim of deeply checking the accuracy of the common factor as a real-time business cycle indicator, we assume that the indicator is subject to regime switches.[9] For this purpose,

[8] In particular, if the logs of a variable appears as non-stationary according to Ng and Perron (2001) unit root tests, then the data are used in growth rates. To save space, the results are not presented but they are available from the authors upon request.

[9] Camacho, Pérez-Quirós and Poncela (2012) show that although the fully Markov-switching dynamic factor model is generally preferred to the shortcut of computing inferences from the common factor obtained from a linear factor model, its marginal gains rapidly diminish as the quality of the indicators used in the analysis increases. This is precisely our case.

we assume that the switching mechanism of the common factor at time t, x_t, is controlled by an unobservable state variable, s_t, which is allowed to follow a first-order Markov chain. Following Hamilton (1989), a simple switching model may be specified as :

$$x_t = c_{s_t} + \sum_{j=1}^{p} \alpha_j x_{t-j} + \varepsilon_t, \tag{34}$$

where $\varepsilon_t \sim iidN(0, \sigma^2)$.[10] The nonlinear behavior of the time series is governed by c_{s_t}, which is allowed to change within each of the two distinct regimes $s_t = 0$ and $s_t = 1$. The Markov-switching assumption implies that the transition probabilities are independent of the information set at $t-1$, x_{t-1}, and of the business cycle states prior to $t-1$. Accordingly, the probabilities of staying in each state are:

$$p(s_t = i/s_{t-1} = j, s_{t-2} = h, ..., x_{t-1}) = p(s_t = i/s_{t-1} = j) = p_{ij}. \tag{35}$$

Taking the maximum likelihood estimates of parameters, reported in Table 3, in the regime represented by $s_t = 0$, the intercept is positive and statistically significant while in the regime represented by $s_t = 1$, it is negative and statistically significant. Hence, we can associate the first regime with expansions and positive values of the indicator and the second regime with recessions and negative values of the indicator. According to the related literature, expansions are more persistent than downturns (estimated p_{00} and p_{11} of about 0.97 and 0.87, respectively). These estimates are in line with the well-known fact that expansions are longer than contractions, on average. Using the transition probabilities, one can derive the expected number of months that the business cycle phases prevail as $(1 - p_{ii})^{-1}$. Conditional on being in $s_t = 0$, the expected duration of a typical Argentine expansion is 33 months, and the expected duration of recession is approximately 8 months.

[Insert Table 3 about here]

Figure 1 displays the smoothed probabilities of being in state $s_t = 1$ that comes from this model as shaded areas. The figure shows that there is a high commonality in switch times of probabilities with Argentine business cycle phases as identified by both the common factor and EMAE. It validates the interpretation of state $s_t = 1$ as recession and the probabilities plotted in this chart as probabilities of being in recession.

[10] Based on Camacho and Perez Quiros (2007), we did not include any lags in the factor. We checked that the resulting model is dynamically complete in the sense that the errors are white noise.

4.3 Simulated real-time analysis

Among many others, Stark and Croushore (2002) suggest that the analysis of in-sample forecasting performance of competitive models is questionable since the results can be deceptively lower when using real time vintages. This happens because the in-sample analysis misses three aspects of real time forecasting: (i) the recursive estimation of the model parameters; (ii) the real time data flow, i.e. the fact that data are released at different point in time; and (iii) the real time data revisions.

However, although developing real time data sets is conceptually simple, producing real time vintages is, as in our case, unfeasible since the historical records of many time series are not available. In the context of dynamic factor models, an interesting alternative to the real time forecasting analysis is the pseudo real time forecasting exercise suggested by, among others, Giannone, Reichlin and Small (2008). Their proposal consists of taking into account the recursive estimate of the models and the real time data flow (and hence the publication lags) but, due to data availability constraints, does not consider data revisions.

The proposal is based on trying to mimic as closely as possible the real time analysis that would have been performed by a potential user of dynamic factor models when forecasting, at each period of time, on the basis of different vintages of data sets. The experiment considers that the releases of each vintage contain missing data at the end of the sample reflecting the lags in the calendar of data releases that has been summarized in Table 1. This allows us to reproduce every 15 days the typical end of the sample unbalanced panel faced by the forecaster due to the lack of synchronization of the data releases. Accordingly, the experiment is labeled as "pseudo" because the vintages are not obtained in pure real time but from the latest available data set.

The forecast performance analysis was conducted to simulate real-time forecasting. The first data vintage of this experiment refers to April 1, 2002 and, although it was collected from the information of the latest available data set, it preserved the data release calendar that a forecaster would have faced on that day. Using this data vintage, we computed nine-month blocks of forecasts. Among them, some refer to the last quarter's GDP growth before its official release (backcasts), others refer to current quarter GDP growth (nowcasts), while others refer to the next quarter's GDP growth (forecasts). Against this background, the data vintages were recursively updated on the first day and fifteenth day of each month. All parameters, factors, and so forth were then re-estimated, and nine-month blocks of backcasts, nowcasts and forecasts were then computed. The final pseudo real time nine-month block of forecasts was made on January 15, 2013, leading to 246 different blocks of forecasts (from first quarter 2002 to 2012).

The predictive accuracy of our model is examined in Table 4. The table shows the mean-squared forecast errors (MSE), which are the average of the deviations of the predictions from the final releases of GDP available in the data set. Results for backcasts, nowcasts and forecasts appear in the second, third and fourth columns of the table, respectively. In addition to the factor model described in Section 2 (labeled as MICA), two benchmark models are included in the forecast evaluation. The former is an autoregressive model of order two (AR) which is estimated in real-time producing iterative forecasts, and the latter is a random walk (RW) model whose forecasts are equal to the average latest available real-time observations.

The MSE leads to a ranking of the competing models according to their forecasting performance. However, it is advisable to test whether the forecasts made with the dynamic factor model are significantly superior to the others models' forecasts. To analyze whether empirical loss differences between two or more competing models are statistically significant, the last three rows of the table shows the pairwise test introduced by Diebold and Mariano (DM, 1995).

The immediate conclusion obtained when comparing the forecasts is that the dynamic factor model unequivocally outperforms the alternative forecasting models, although the magnitude of these gains depend on the forecast horizon. In the backcasting exercise, the differences between the MSE results using the factor model and the benchmark models are noticeable. The relative MSE of the dynamic factor model versus RW and AR are 0.368 to 0.350 and, according to the p-values of the DM test, the differences are statistically significant (p-values of 0.001 in both cases). The relative gains diminishes with the forecasting horizon, reducing to 0.742 and 0.709 in nowcasting and to 0.880 and 0.866 in forecsating. This result is quite intuitive because the backcasts and nowcasts are computed immediately before the end of the quarter, which allow the model to use the latest available information of the respective quarter from the early available indicators. Notably, although the gains diminish, they are still statistically significant, according to the p-values of the DM test reported in the bottom panel of this table.

[Insert Table 4 about here]

5 Conclusions

In this paper, we propose a small-scale factor model with mixed frequencies to produce accurate backcasts, nowcasts and short-run forecasts of Argentine GDP growth. Our model is succesful, not only in computing a coincident indicator, which is in striking accord with the actual history of Argentine business cycle, but also in explaining a very high percentage of the variance of actual GDP growth. Moreover, our pseudo real-time analysis shows that our dynamic factor model clearly outperforms univariate forecasts, especially when forecasting the next unavailable figure of GDP growth. This encourages real time forecasters to back-check the bulk of monthly real and survey data which are published in the respective quarter before the next GDP release. Therefore, we strongly consider that it is a valid tool to be used to monitor the business cycle and to compute short-term forecasts of Argentine GDP.

References

[1] Aguiar, Mark, and Gita Gopinath (2007), "Emerging Market Business Cycles: The Cycle Is the Trend", Journal of Political Economy, Vol. 114, No. 1, pp. 69-102.

[2] Aruoba, S. Boragan, Francis X. Diebold, and Chiara Scotti (2009), "Real-Time Measurement of Business Conditions", Journal of Business and Economic Statistics, Vol. 27, No. 4, October, pp. 417-427.

[3] Aruoba, S. Boragan, and Francis X. Diebold (2010), "Real-Time Macroeconomic Monitoring: Real Activity, Inflation, and Interactions", American Economic Review Papers& Proceedings, Vol. 100, No. 2, pp. 20-24

[4] Burns and Mitchel (1946), Measuring Business Cycless, NBER, New York.

[5] Camacho, Maximo, and Gabriel Pérez-Quirós (2007). Jump-and-rest effect of U.S. business cycles. *Studies in Nonlinear Dynamics and Econometrics* 11(4): article 3.

[6] Camacho, Maximo, and Rafael Doménech (2012), "MICA-BBVA: A factor model of economic and financial indicators for short-term GDP forecasting", SERIES: Journal of the Spanish Economic Association, forthcoming.

[7] Camacho, Maximo, and Jaime Martinez-Martin (2012), "Real-time forecasting US GDP from small-scale factor models", BBVA Working Paper 12/10, June.

[8] Camacho, Maximo, and Gabriel Perez-Quiros (2010), "Introducing the Euro-Sting: short term indicator of euro area growth", Journal of Applied Econometrics, Vol. 25, No. 4, June/July, pp. 663-694.

[9] Camacho, Maximo, and Agustin Garcia-Serrador (2013), "The Euro-Sting revisited: the usefulness of financial indicators to obtain euro area GDP forecasts", Journal of Forecasting, forthcoming.

[10] Camacho, Maximo, Pérez-Quirós, Gabriel and Pilar Poncela (2012), "Markov-switching dynamic factor models in real time", CEPR Discussion Papers 8866.

[11] D'Amato, Laura, Lorena Garegnani, and Emilio Blanco (2011), "Nowcasting economic activity in Argentina with many predictors", Annals of the XLVI Annual Meeting of the Argentine Association of Political Economy, November.

[12] D'Amato, Laura, Lorena Garegnani, and Emilio Blanco (2011), "Uso de flujos de informacion de alta frecuencia para pronosticos de corto plazo de la actividad economica en Argentina", Investigaciones Economicas, Vol. 64, Oct.-Dic, pp. 7-33.

[13] Diebold F., and Mariano, R. 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13: 253-263.

[14] Giannone, D., Lucrezia Reichlin, and D. Small (2008), "Nowcasting: The real-time informational content of macroeconomic data", Journal of Monetary Economics 55: 665-676.

[15] Hamilton, James (1989), "A new approach to the economic analysis of nonstationary time series and the business cycles" , Econometrica 57: 357-384.

[16] Jorrat, Juan Mario (2005), "Construccion de Indices Compuestos Mensuales Coincidente y Lider en Argentina", Chapter 4 in Mariana Marchionjni (Ed.), Progresos en Econometria, Ed. Temas and AAEP, pp. 43-100.

[17] Lane, Philip R. (2003), "Business Cycles and Macroeconomic Policy in Emerging Market Economies", International Finance, Vol. 6, No. 1, pp. 89-108.

[18] Liu, Phillip, Troy Matheson and Rafael Romeu (2011), "Real-Time Forecasts of Economic Activity for Latin American Economies", IMF Working Paper N^o 11/98.

[19] Mariano, Roberto S., and Yasumoto Murasawa (2003), "A new coincident index of business cycles based on monthly and quarterly series", Journal of Applied Econometrics, Vol. 18, pp. 427–443.

[20] Ng, Serena, and Pierre Perron (2001), "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," Econometrica, Vol. 69, No. 6, 1519-54.

[21] Simone, Alejandro (2001), "In Search of Coincident and Leading Indicators of Economic Activity in Argentina", IMF Working Paper WP/03/30, Marcn.

[22] Stark, Tom, and Dean Croushore (2002), "Forecasting with Real-Time Data Set for Macroeconomists", Journal of Macroeconomics 24: 507-531.

[23] Stock, James H., and Mark Watson (1989), "New Indexes of Coincident and Leading Economic Indicators", in Oliver Blanchard and Stanley Fischer (Eds.), NBER Macroeconomics Annual 1989 , Vol. 4, pp. 351-409.

[24] Stock, James H., and Mark Watson (1991), "A Probability Model of the Coincident Economic Indicators", in K. Lahiri and G. Moore (Eds.), Leading Economic Indicators: New Approaches and Forecasting Records, Cambridge University Press, UK, pp. 63-89.

Table 1: Final variables included in the model

	Series	Sample	Source	Publication delay	Data transform.
1	Real Gross Domestic Product (GDP, SAAR, Mil.1993 Pesos)	1993.1 2012.3	INDEC	2.5 to 3 months	QGR
2	Industrial Production Index (IPI) (SA, 1993=100)	1993.01 2012.07	FIEL	25 days	MGR
3	All Employees: Total Urban Population (Empl, SA, Thous)	1993.1 2012.2	INDEC	1.5 to 2 months	QGR
4	Real Retail Sales:Total Supermarket Sales (Sales, SA, IPC deflated, constant. ARS)	1997.06 2012.06	INDEC/ Census	25-30 days	MGR
5	Synthetic Indicator of Construction Activity (ISAC, SA)	1993.01 2012.07	UTDT	30 days	MGR

Notes: SA means seasonally adjusted. MGR and QGR mean monthly growth rates, quarterly growth rates and levels, respectively. INDEC: National Institute of Statistics and Census; FIEL: Latin American Foundation of Economic Investigations.

Table 2: Loading factors

GDP	IP	Empl	ISAC	Sales	% Var
0.28 (0.05)	0.41 (0.09)	0.28 (0.12)	0.35 (0.07)	0.16 (0.08)	89.9%

Notes. The loading factors (standard errors are in brackets) measure the correlation between the common factor and each of the indicators appearing in columns. See Table 1 for a description of these indicators.

Table 3. Markov-switching estimates

c_0	c_1	σ^2	p_{00}	p_{11}
1.09 (0.15)	-5.39 (0.43)	3.45 (0.32)	0.97 (0.01)	0.87 (0.06)

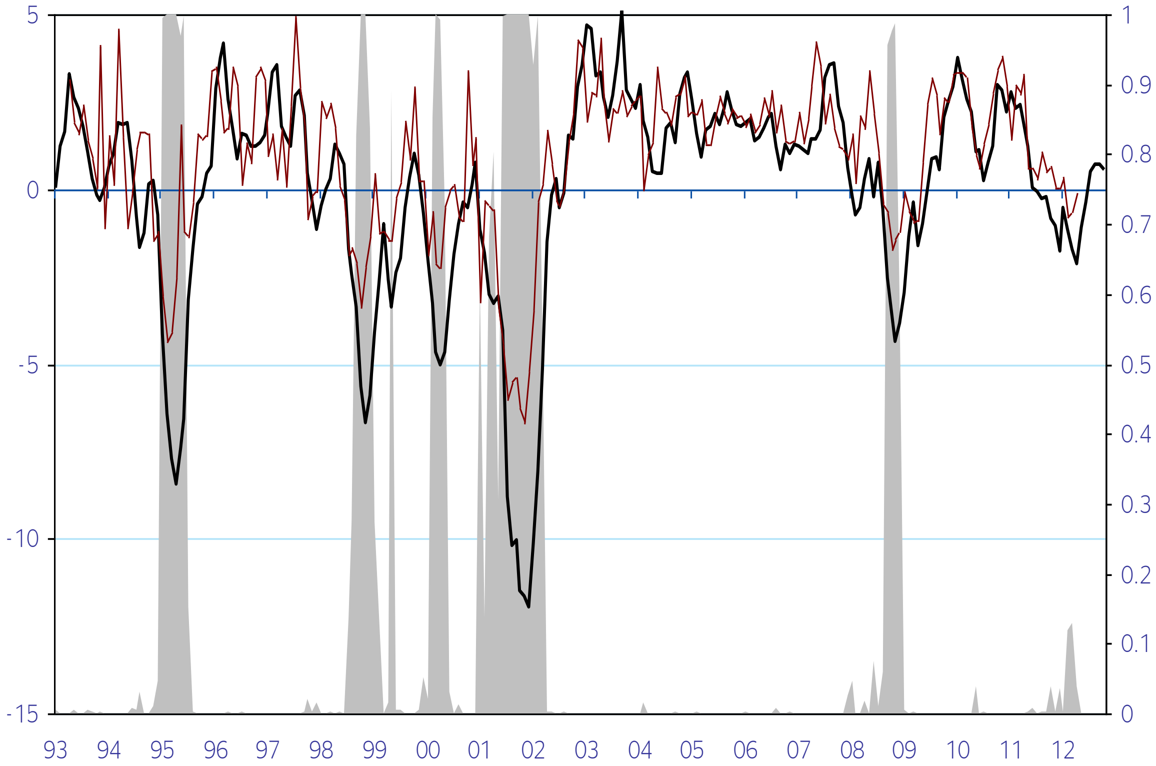
Notes. The estimated model is $x_t = c_{s_t} + \varepsilon_t$, where x_t is the common factor, s_t is an unobservable state variable that governs the business cycle dynamics, $\varepsilon_t \sim iidN(0,\sigma)$, and $p(s_t = i / s_{t-1} = j) = p_{ij}$.

Table 4: Predictive accuracy

	Backcasts	Nowcasts	Forecasts
	Mean Squared Errors		
Our model	1.049	1.552	1.884
RW	2.851	2.090	2.139
Our model/RW	0.368	0.742	0.880
AR	2.999	2.189	2.174
Our model/AR	0.350	0.709	0.866
	Equal predictive accuracy tests		
Our model vs RW	0.001	0.032	0.041
Our model vs AR	0.001	0.015	0.021

Notes. The forecasting sample is 2002.1-2012.1, which implies comparisons over 246 forecasts. The top panel shows the Mean Squared Errors (MSE) of our dynamic factor model, a random walk (RW), an autoregressive model of order two (AR), along with the relative MSEs over that of our model. R and E refer to recessions and expansions periods. The bottom panel shows the *p*-values of the Diebold-Mariano (DM) test of equal predictive accuracy.

Figure 1. Common factor, EMAE (var. % 3M-3M) and recession probabilities



Notes. Black line refers to the common dynamic factor (1993.03-2012.12, left-hand scale). Red line refers to EMAE (1993.06-2012.06, left-hand scale). Shaded areas refer to smoothed probabilities of recession (right-hand scale).