Toward a more reliable picture of the economic activity: An application to Argentina

Maximo Camacho
Universidad de Murcia
mcamacho@um.es
Shipping address: Departamento de metodos cuantitativos para la economia y la empresa, Facultad de Economia y Empresa, Universidad de Murcia, 30100, Murcia, Spain

Marcos Dal Bianco (**Corresponding author)**
BBVA Research
marcos.dalbianco@bbva.com
Shipping address: BBVA Frances, Peron 380, 1°. C1003ABC, Buenos Aires, Argentina
Tel: +54 9 11 36818361 (mobile)
Fax: +54 11 4346 4416

Jaime Martinez-Martin
BBVA Research
AQR-IREA (U Barcelona)
jaime.martinezmartin@ub.edu

Abstract

We advocate a dynamic factor model to provide alternative measures of output data using indirect information from economic indicators. Notably, the method is valid regardless of the length and the frequency of the indicators used in the analysis. We apply the method to show evidence of a significant gap between estimated and official measures of Argentine GDP since 2007.

JEL Classification: C53, E01, E27, E32, E37.
Keywords: Business Cycles, Output Growth, Dynamic Factor Models.

1. Introduction

Questioning the reliability of official economic data has a long history (Rawski, 1993) that continues nowadays (Michalski and Stoltz, 2013), especially in emerging-market economies, whose structures of production and expenditure change fast. As Koech and Wang (2012) pointed out, the Great Recession renewed the skepticism about the official figures of output in some of these economies, which might understate the effect of the economic slowdown. The issue has attracted considerable interest not only in academic literature, but also in the mass media (Bradsher, 2012).

To get a more reliable picture of these economic aggregates, the literature typically follows three stages. The first stage implies collecting an alternative set of reliable indicators of economic activity, which serve as a proxy. In the second stage, a statistical model captures the in-sample statistical relationships across the indicator and the aggregate over a reliable sample period. The final step consists on computing out-of-sample forecasts of the economic aggregates for the questionable period and checking whether they diverge substantially from the actual data. The projections can be either direct forecasts, i.e., based on the National Accounts methodology used in the in-sample period (Coremberg, 2014) or model-based forecasts (Fernald, Malkin and Spiegel, 2013).

The focus of this paper is on model-based forecasts. Recently, Fernald et al. (2013) use uniequational projections of Chinese GDP growth on the first principal component of eight economic indicators and lagged growth. However, the implementation of this approach is limited since it requires indicators of the same frequency as GDP, which must also be available for the same sample.

To overcome this problem, we advocate that a dynamic factor framework that allows for missing observations is an extremely useful framework to compute the out-of-sample forecasts, while handling both mixed frequencies and data availability problems. After the last trustworthy observation, missing values are added to the suspicious aggregate. The variables observed at a lower frequency can also be viewed as being periodically missing. Finally, unavailable data are also treated as missing. Now, the maximum likelihood estimation of state-space models with missing observations can be applied straightforwardly and carry on with the Kalman filter that accounts for missing data. Typically, the method consists on replacing the missing figures with random draws, which are skipped from the Kalman recursions when updating. Therefore, the projections are a direct output of the filter. Importantly, the projections are obtained in real time.

To illustrate our proposal, we focus on Argentina. In February, 2013, the International Monetary Fund took the unprecedented step of censuring a member, encouraging the country to improve its efforts to meet the IMF standards for inflation and GDP data. Using the model proposed by Camacho, Dal Bianco, and Martinez-Martin (2014), we truncate the sample at December 2006 and then take the point estimates of our specification to predict from 2007 onwards. We show that the recent data are inconsistent with the relationships observed in the past. In particular, we find that the gap with our estimates is about 14.4%.

2. Dynamic factor model

Let $X_{t,}$ be a vector of N economic indicators that may include N_1 quarterly indicators, $X^q_{t,1}$, and N_2 monthly indicators, $X^m_{t,2}$. Let us assume that all the indicators are observed for every month, including $X^m_{t,1}$, the monthly series underlying $X^q_{t,1}$, although this assumption will be relaxed soon. Also, let us assume that the logs of $X_{t,}$ contain a unit root.

If the sample mean of the within-quarter activity can be well approximated by the geometric mean, Mariano and Murasawa (2003) show that the quarter-over-quarter growth rate of quarterly indicators computed at each month of the sample, $Y^q_{t,1}$, can be expressed as the averaged sum of previous month-over-month growth rates

$$Y^q_{t,1} = \frac{1}{3}Y^m_{t,1} + \frac{2}{3}Y^m_{t-1,1} + Y^m_{t-2,1} + \frac{2}{3}Y^m_{t-3,1} + \frac{1}{3}Y^m_{t-4,1}. \tag{1}$$

We assume that the month-over-month growth rates admit a factor decomposition, i.e., they can be expressed as the sum of a common factor and idiosyncratic components

$$Y^m_{t,i} = \beta_i f_t + U_{t,i}, \tag{2}$$

where i=1, 2.

To complete the dynamic specification of the model, we assume that $U_{t,i}$ and f_t follow autoregressive processes of orders p1, p2 and p3, respectively

$$U_{t,1} = A_{1,1}U_{t-1,1} + ... + A_{p1,1}U_{t-p1,1} + \varepsilon_{t,1}, \tag{3}$$

$$U_{t,2} = A_{1,2}U_{t-1,2} + ... + A_{p1,2}U_{t-p2,2} + \varepsilon_{t,2}, \tag{4}$$

$$f_t = b_1 f_{t-1} + ... + b_{p3} f_{t-p3} + \varepsilon_{t,f}, \tag{5}$$

where $A_{j,i}$ are diagonal matrices; $\varepsilon_{t,1}$ and $\varepsilon_{t,2}$ are multivariate Gaussian processes with zero mean and diagonal covariance matrices Ω_1 and Ω_2; and $\varepsilon_{t,f}$ is an univariate Gaussian process with zero mean and variance σ^2_f. All of these errors are uncorrelated at all leads and lags.

It is convenient to cast these expressions into state-space form. Let $0_{a\times b}$ and $1_{a\times b}$ be the $(a\times b)$ matrices of zeroes and ones and let I_a the $(a\times a)$ identity matrix. Let v be the vector $\left(\frac{1}{3},\frac{2}{3},1,\frac{2}{3},\frac{1}{3}\right)'$. Without loss of generality, let us assume that p1=p2=p3=1. The *measurement equation* is $Y_t = H\beta_t + E_t$, where $Y_t = \left(Y^q_{t,1}{}', Y^m_{t,2}{}'\right)'$, $E_t \sim i.i.d.N(0,R)$, $R = 0$, $\beta_t = \left(f_t, f_{t-1},..., f_{t-4}, U'_{t,1}, U'_{t-1,1},...,U'_{t-4,1}, U'_{t,2}\right)'$ and

$$H = \begin{pmatrix} \frac{\beta_1}{3} & \frac{2\beta_1}{3} & \beta_1 & \frac{2\beta_1}{3} & \frac{\beta_1}{3} & v'\otimes I_{N_1} & 0_{N_1\times N_2} \\ \beta_2 & 0_{N_2\times 1} & 0_{N_2\times 1} & 0_{N_2\times 1} & 0_{N_2\times 1} & 0_{N_2\times 5N_1} & I_{N_2} \end{pmatrix}, \qquad (6)$$

with β_1 and β_2 being the $(N_1 \times 1)$ and $(N_2 \times 1)$ loading vectors of $Y_{t,1}^m$ and $Y_{t,2}^m$.

Let us now specify the *transition equation*, $Y_t = F\beta_t + V_t$. If $\bar{b}_1 = (b_1, 0_{1\times 3})$

$$B = \begin{pmatrix} \bar{b}_1 & 0_{1\times 4} \\ I_4 & 0 \end{pmatrix}, \qquad (7)$$

the matrix *F* is

$$F = \begin{pmatrix} B & 0_{5\times N_1} & 0_{5\times N_1} & 0_{5\times N_1} & 0_{5\times N_1} & 0_{5\times N_1} & 0_{5\times N_2} \\ 0_{N_1\times 5} & A_{1,1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_2} \\ 0_{N_1\times 5} & I_{N_1} & 0_{5\times 5} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_2} \\ 0_{N_1\times 5} & 0_{N_1\times N_1} & I_{N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_2} \\ 0_{N_1\times 5} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & I_{N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_2} \\ 0_{N_1\times 5} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_1} & I_{N_1} & 0_{N_1\times N_1} & 0_{N_1\times N_2} \\ 0_{N_2\times 5} & 0_{N_2\times N_1} & 0_{N_2\times N_1} & 0_{N_2\times N_1} & 0_{N_2\times N_1} & 0_{N_2\times N_1} & A_{1,2} \end{pmatrix}. \qquad (8)$$

Finally, the Gaussian errors V_t are serially uncorrelated with zero mean and covariance matrix $Q = diag\left(\left(\sigma_f^2, 0_{1\times 4}, \bar{\Omega}_1', 0_{1\times 4N_1}, \bar{\Omega}_2'\right)'\right)$, where $\bar{\Omega}_i$ is the vector that contains the elements of the main diagonal of Ω_i, where *i*=1,2.

If there were no missing data, the Kalman filter could be used to estimate model's parameters by maximizing the likelihood function. However, this assumption is quite unrealistic in empirical applications since the analysts have to deal with mixing quarterly and monthly frequencies and with time series which are published with different time lags. In our context, additional projections of some unreliable data are also required, which can be considered as missing data from certain date as well.

Mariano and Murasawa (2003) show that the system of equations remains valid with missing data after a subtle transformation. They propose replacing the missing observations with random draws r_t, whose distribution cannot depend on the parameter space that characterizes the Kalman filter.[1] Then, the measurement equation is 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 let R_{ii} be its variance. Let H_{it} be the *i*-th row of the matrix H_t, which has *z* columns. The measurement equation can be replaced by the following expressions

[1] We assume that $r_t \sim N\left(0, \sigma_r^2\right)$ but replacements by constants would also be valid.

$$Y_{it}^{+} = \begin{cases} Y_{it} & \text{if } Y_{it} \text{ is observable} \\ r_t & \text{otherwise} \end{cases}, \tag{9}$$

$$H_{it}^{+} = \begin{cases} H_i & \text{if } Y_{it} \text{ is observable} \\ 0_{1\times z} & \text{otherwise} \end{cases}, \tag{10}$$

$$E_{it}^{+} = \begin{cases} E_{it} & \text{if } Y_{it} \text{ is observable} \\ r_t & \text{otherwise} \end{cases}, \tag{11}$$

$$R_{iit}^{+} = \begin{cases} R_{ii} & \text{if } Y_{it} \text{ is observable} \\ \sigma_r^2 & \text{otherwise} \end{cases}, \tag{12}$$

This substitution leads to a time-varying state space model with no missing observations so the Kalman filter can be directly applied to Y_{it}^{+}, H_{it}^{+}, E_{it}^{+}, R_{iit}^{+}.

3. Empirical example

We use the indicators employed in Camacho et al. (2014), which were useful to account for the real activity in Argentina. The data are real quarterly GDP, quarterly employment, monthly industrial production, monthly real trade sales, and the monthly indicator of construction activity. The data were collected on December 1, 2013 and span the period from January 1993 to October 2013.

All the variables were seasonally adjusted, they were stationary or transformed to be stationary, and they were standardized to have zero mean and unit variance before estimating the model. For each indicator, the effective sample, source, publication delay and data transformation are summarized in Table 1. According to the method described in Section 2, only data of GDP until the fourth quarter of 2006 were used to obtain the maximum likelihood estimates. Table 2 shows that all the loading factors are positive and statistically significant, which provide evidence that these indicators were procyclical.

Figure 1 shows the main result of the procedure. The top line refers to the Monthly Estimator of Economic Activity (EMAE, for its Spanish acronym), which is a monthly proxy of quarterly GDP published by the official statistical office (INDEC). The middle line plots the estimate of $X_{t,GDP}^{m}$, which is a reproducible GDP series that captures the in-sample statistical relationships across the economic indicators and GDP over the sample period 1993 to 2006 and that projects GDP since then. The large discrepancies between the official figures and our estimates that appeared since 2006 raised suspicions about uneven levels of the official estimates.

To check for the robustness of our results, Figure 1 also plots (bottom line) the monthly evolution the General Indicator of Activity (IGA, for its Spanish acronym), which is published by the private economic research company Orlando Ferreres. According to the figure, the level and the dynamics of our estimated time series are in striking accord with those of IGA, which reinforces the method proposed in this paper.

Figure 2 plots the official values of quarterly GDP from 1993 to 2012 and three alternative projections from 2008 to 2012, which follow three different approaches. The first projection is the direct forecast done by Coremberg (2014), in which the GDP is estimated using the same methodology that had been employed by the official statistical agency up to 2006. The second approach is the model-based forecast described in this paper, i.e., the figure plots the estimates of $X^{q}_{t,GDP}$. The third approach is a statistical approach that consists on projecting GDP by using the annual growth rates of IGA.

The figure reveals that the different approaches yield similar information regarding the reliability of the official GDP, but our method can obtain a reliable picture of economic activity much sooner than the alternative ones. After 2007, the plots show an important gap between the official figures and all alternatives, which accumulates into a substantial difference increasing every year since then. In particular, the official real GDP shows an accumulated positive gap of 12.2% with respect to ARKLEMS GDP. According to our results, the direct approach followed by ARKLEMS could even underestimate this gap. Our model-based projections and IGA-based forecasts find that the accumulated positive gap could be about 14.4%.

4. Conclusion

Since the Great Recession, there has been a great deal of speculation that output slowed more than official figures indicate in some emerging markets economies. To examine this potential mismatch, we propose a simple and very flexible method, based on dynamic factor models, that accounts for unbalanced panels of economic indicators that could be available at different frequencies. We apply the method to Argentina, which became the first country to be censured by the International Monetary Fund for not providing accurate data on inflation and economic growth. From 2007 to 2012, our results indicate that official real GDP shows an accumulated positive gap of about 14%.

Acknowledgements

We are grateful to Daniel Aromí and to seminar participants at the University of CEMA and the University of Buenos Aires for their very useful comments. All remaining errors are our own responsibility. Maximo Camacho thanks CICYT for its support through grants ECO2010-19830 and ECO2013-45698. The views expressed in this paper are those of the authors and do not represent those of BBVA.

References

Bradsher, K. 2012. China Data Mask Depth of Slowdown, Executives Say. New York Times, June 22.

Camacho, M., Dal Bianco, M., and Martinez-Martin, J. 2014. Short-run forecasting of Argentine GDP growth. *Emerging Markets Finance and Trade*, forthcoming.

Coremberg, A. 2014. Measuring Argentina's GDP growth. Myths and facts. *World Economics* 15: 1-31.

Fernald, J., Malkin, I., and Spiegel, M. 2013. On the reliability of Chinese output figures. *San Francisco Fed Economic Letter* 8, March.

Koech, J., and Wang, J. 2012. China's slowdown may be worse than the official data suggest. *Dallas Fed Economic Letter* 7: August.

Mariano, R., and Murasawa, Y. 2003. A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics* 18: 427-443.

Michalski, T., and G. Stoltz. 2013. Do countries falsify economic data strategically? Some evidence that they might. *Review of Economics and Statistics* 95: 591-616.

Rawski, T. 1993. How fast has Chinese industry grown? WB Policy Research Paper No. 1194.

Table 1: Variables included in the model

Series	Sample	Source	Publication delay	Data transformation
Real Gross Domestic Product (GDP, SAAR, Mill. 1993 ARS)	1993.1 2013.6	INDEC	2.5 to 3 months	QGR
Industrial Production Index (IPI, SA, 1993=100)	1993.01 2013.10	FIEL	25 days	MGR
All Employees: Total Urban Population (EMPL, SA, Thous)	1993.1 2013.6	INDEC	1.5 to 2 months	QGR
Real Retail Sales: Total Supermarket Sales (SALES, SA, IPC deflated, constant. ARS)	1997.06 2013.10	INDEC/ FIEL	25-30 days	MGR
Synthetic Indicator of Construction Activity (ISAC, SA)	1993.01 2013.10	INDEC	30 days	MGR

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

Table 2: Loading factors

GDP	IP	EMPL	ISAC	SALES
0.26	0.43	0.24	0.34	0.12
(0.06)	(0.10)	(0.12)	(0.08)	(0.08)

Notes. The loading factors (standard errors are in brackets) measure the correlation between the common factor and each of the indicators.

Figure 1. EMAE, Estimated Monthly GDP, and IGA (levels, 1993=100)

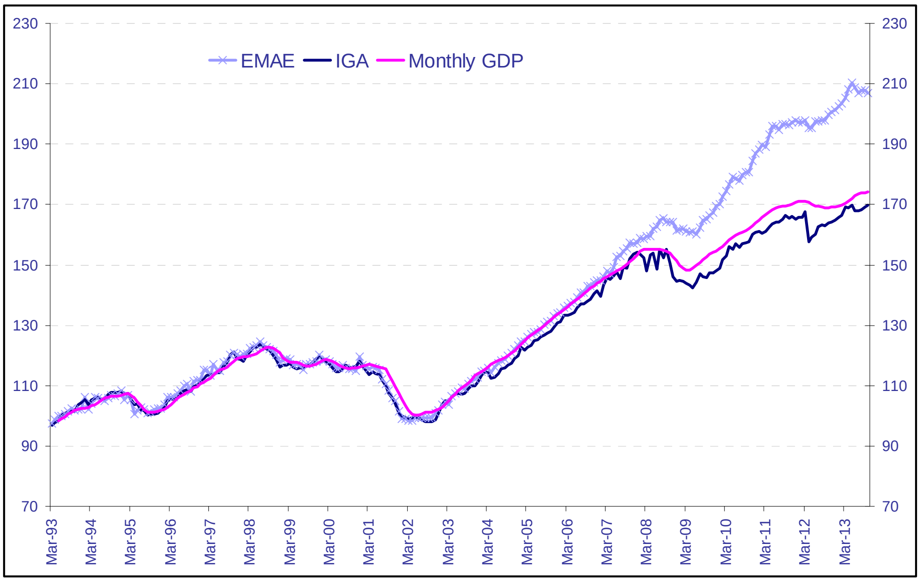
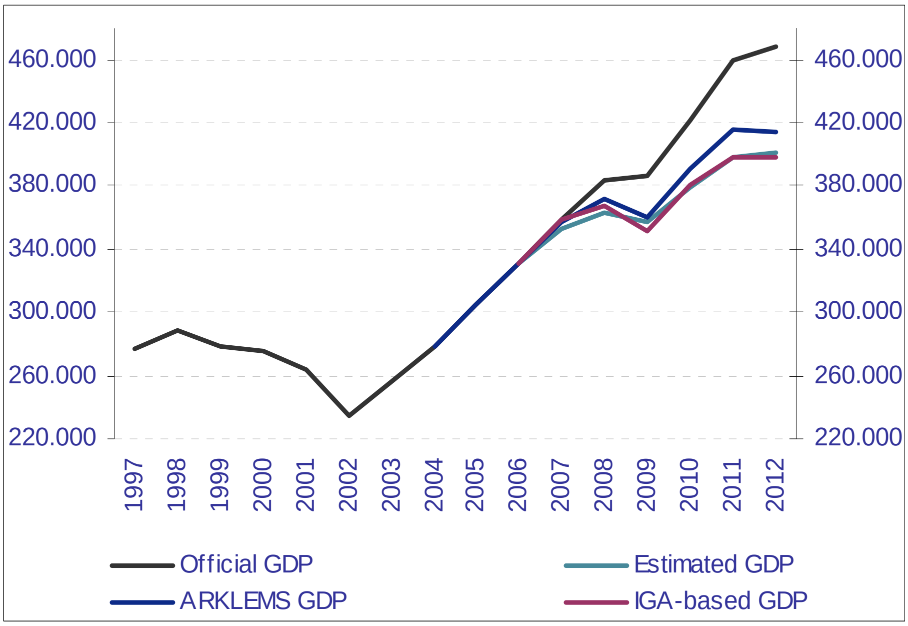


Figure 2. Argentine GDP (annual data, levels, 1993 prices)



Notes. Argentine GDP: Official, ARKLEMS, IGA-growth-based, and Estimated GDP