UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT

December 2018

UNCTAD Research Paper No. 27
UNCTAD/SER.RP/2018/9/Rev.1

Fernando Cantú
Division on Globalization and Development Strategies, UNCTAD
fernando.cantu@un.org

Estimation of a coincident indicator for international trade and global economic activity

Abstract

World economic aggregates are compiled infrequently and released after considerable lags. There are, however, many potentially relevant series released in a timely manner and at a higher frequency that could provide significant information about the evolution of global aggregates. The challenge is then to extract the relevant information from this multitude of indicators and combine it to track the real-time evolution of the target variables. We develop a methodology based on dynamic factor models adapted to accommodate for variables with heterogeneous frequencies, ragged ends and missing data. We apply this nowcast methodologies to three variables of interest: global trade in goods, global trade in services and world GDP in real terms. In addition to monitoring these variables in real time, this method can also be used to obtain short-term forecasts based on the most up-to-date values of the underlying indicators.

Keywords: Coincident indicators, nowcasting methodologies, short-term forecasts, international trade, economic activity

UNITED NATIONS

Contents

Acknowledgements

This paper was prepared by Fernando Cantú (UNCTAD). The author would like to thank colleagues from UNCTAD's Development Statistics and Information Branch for their helpful comments and suggestions.

1 Introduction

On a global scale, there is a constant stream of economic information released by official and private sources. National statistical authorities publish national accounts, balance of payments, government finances, monetary statistics, data from the banking sector and a multitude of other socio-economic indicators. International actors also compile and publish diverse statistical series. To this, one can add the data produced by commercial statistical providers, with products such as public opinion polls or surveys of business confidence. Lately, there has been an important addition to this list: private actors that collect statistical information as part of their activities and that bundle them as standalone products. This includes, to name just a few examples, stock exchanges, financial actors, port operators, retailers, and social media and social networking companies.

All these series provide a piece of information either on overall economy activity or on specific sectors. Although this means that we can keep track of the state of the economy in real time, it also introduces the challenge of sorting out the information that is relevant from that which is not. This problematic is exacerbated by a long list of statistical complications: missing data, measurement errors, undercoverage, low signal-to-noise ratio, heterogeneous frequencies, different starting and ending dates, asynchronous update schedule, constant revisions, and others.

Assessing the real-time evaluation of macroeconomic variables based on a series of timely, high-frequency indicators (a process that has come to be called “nowcasting”) is not new. The classical literature on coincident and leading indicators (see, for example, Conference Board (2001) and OECD (2012)) is well established and it has been applied in many areas. Bridge models, linking high frequency variables to a target variable of lower frequency, are standard tools in statistical analysis (see Baffigi et al. (2004), Barhoumi et al. (2008) and Rossiter (2010) for recent examples of their application to forecast economic activity). Progress in this area accelerated with the development of more sophisticated techniques of data selection and processing, particularly after the application of dynamic factor models by Stock and Watson (2002a,b), Forni et al. (2005) and Giannone et al. (2008).

Dynamic factor models introduce the assumption that the observed indicators can be divided into two components: one that can be attributed to one or a few (unobserved) factors that are common to all series, and another that is specific or idiosyncratic to each component. The factor model serves to establish a mapping between the common factors and the indicators. It is a dynamic model, so the factors are assumed to change through time according to an autoregressive process. This model can be summarized through a statespace representation, in which the mapping between indicators and factors becomes the measurement equation, and the dynamics of the unobserved factors become the transition equation. The likelihood of this model can then be calculated via the Kalman filter and the maximum likelihood estimators (MLE) obtained via standard optimization techniques.

The simplicity of the dynamic factor model and its good empirical performance explain its positive reception as a tool to nowcast or forecast economic variables based on many heterogeneous indicators. This solution was later extended and applied to many contexts. For example, Mariano and Murasawa (2003) modified the model to allow indicators of mixed frequencies and Camacho and Perez-Quiros (2010) developed this further by taken into consideration variable reporting lags and the availability of early or “flash” estimates. Schumacher and Breitung (2006) implemented some modifications to improve the forecasting performance of the model. Aruoba et al. (2009) and Aruoba and Diebold (2010) constructed a similar model to incorporate very high frequency indicators and a dynamic factor model that does not require the use of approximations. Matheson (2011) applies the basic methodology but extends the application to simultaneously track economic growth in 32 economies by using country-specific and global indicators. The constant stream of economic indicators is called “big data” by Bok et al. (2017) and in their paper they describe the nowcasting methodology of the Federal Reserve Bank of New York, which closely follows some of the research cited above.
However, dynamic factor models are not the only methodologies used to tackle this problem. For example, Clements and Galvão (2008) use the mixed-frequency model (MIDAS) proposed by Ghysels et al. (2006a,b) to

track output growth in the United States. This method allows to incorporate data sampled at different frequencies in a flexible way. Marcellino and Schumacher (2010) and Ferrara and Marsilli (2014) also employ this methodology to nowcast German and global economic activity, respectively. This paper will not explore this approach since, as it will be explained later, recent developments in dynamic factor models also allow to efficiently incorporate variables of mixed frequencies, in addition to overcoming other statistical complications of empirical data.

The main objective of this paper is the development of a nowcasting methodology for world trade. The methods will also be used to nowcast global economic activity to demonstrate how they can be applied to other target variables. The standard dynamic factor model will be adapted to accommodate the characteristics of trade variables, and it will incorporate the information available in an extensive list of indicators. The nowcasts will be published in future releases of UNCTAD's Handbook of Statistics.[1]

The rest of the paper is organized as follows. The next section will introduce the concept of nowcasting as specifically applied to global trade variables. After that, Section 3 will describe the dynamic factor model and the data transformations required. Section 4 will then present the application of this methodology to our variables of interest and the results obtained. A final section will conclude and introduce some possible areas of future work.

2 Nowcasting global trade variables

Most of the literature described above employs a nowcasting methodology to assess the evolution of global economic activity. Only a handful of recent articles have applied this approach to international trade. Guichard and Rusticelli (2011) develop a dynamic factor model for world trade relying on a set of monthly indicators. Golinelli and Parigi (2014) uses an augmented bridge model based on theoretical-level relationships to jointly assess world trade and economic activity. Finally, Barhoumi and Ferrara (2015) construct two leading indicators for global trade: one relying on the traditional methodology of the Conference Board cited above, and another based on dynamic factor models with single-frequency indicators.

Assessing the evolution of world trade is crucial for a comprehensive evaluation and forecasting of the economy. Many countries rely on international trade as an important component of their economy and demand shocks or episodes of price volatility can bring about severe periods of instability. It is therefore essential to identify shifting trends and sudden changes of direction in these variables immediately or as soon as possible after they occur. International trade is also a variable that can affect the entire national economy, creating imbalances or influencing the effectiveness of economic policy. National authorities are therefore interested to closely monitor this variable so that they can adapt their policies in a timely manner.

Figures on world trade are reliably published by international actors. However, they are only available after a considerable lag and with a low frequency (annual or quarterly). Some providers release variables with a higher frequency (monthly), but at the cost of lower coverage and greater variability. At the same time, there are many variables from a multitude of sources that can potentially provide information on international trade. The challenge of any nowcasting exercise is to extract the relevant information from a heterogeneous set of variables affected, to a greater or lesser degree, with different statistical problems, and organize it into a coherent statistical model.

Two approaches could be followed to monitor international trade. One is to target total figures of global trade directly. The other is to monitor trade variables for the main trading countries or the most important sectors, and then aggregate them into an overall estimate of trade. We prefer the former approach. This because trade variables are influenced by global developments that cross national and regional boundaries: globalization and the internationalization of production, growing importance of global value chains, emergence of large multinationals and their reliance on intra-firm industrial processes spread all over the world, generalized

[1] For the latest release of this annual publication, see UNCTAD (2017).

influence of exchange rate dynamics, rise of protectionism, extended effect of new technologies on production, etc. All this may lead to the existence of business cycles that are specific to global trade, but different from those at the country level. In fact, Burgert and Dées (2008) find evidence that this approach leads to improved forecasts in comparison to the other ("bottom-up") approach.

In this paper, we will rely on the dynamic factor models to nowcast global (aggregate) trade variables. We will adapt the existing solutions to the specific characteristics of trade variables and the indicators that can potentially be used to track them. However, it must be noted that this is not a macroeconometric model that selects explanatory variables based on their causal linkages with international trade or by following a structural model of the world economy. Instead, this is a statistical exercise where indicators are selected because of their correlation with the target variable and their availability in a reliable and timely manner. The following section will describe the methodology and the procedures followed to overcome the different data issues.

3 The nowcasting model

The goal of the exercise is to nowcast a target variable observed with low frequency (quarterly or annual) or after a considerable lag based on a set of indicators that are available more frequently or after a shorter publication lag. However these indicators are not necessarily available as a rectangular, balanced dataset. Instead, they are affected with different statistical issues: mixed frequencies, different start and end dates (i.e., ragged ends), asynchronous timing of data publication, and missing data. The methodology employed to calculate the nowcast should take into account these features of the data. This paper takes the methodologies described in Mariano and Murasawa (2003) and Camacho and Perez-Quiros (2010) as starting points, and extends them to incorporate additional types of variables relevant for nowcasting trade variables.

3.1 Mixed frequencies

The approach to overcome this issue is to express all variables in terms of the highest frequency available in the dataset (in this case, monthly). This requires transforming the variables, through approximations if necessary, so that they are all expressed in terms of monthly growth rates.

3.1.1 Annual variables

Let X_t^A be a variable observable once per year. This variable can be written as an aggregation of the last 12 observations of its corresponding (unobservable) monthly time series, in the following way.

$$X_t^A = X_t^m + X_{t-1}^m + \cdots + X_{t-11}^m$$

Note that the t index refers to time in months and that the target variable is available only once every 12 months, with the rest of the observations missing. It is possible to represent this sum as 12 times the average of the last 12 monthly observations.

$$X_t^A = 12\left(\frac{X_t^m + X_{t-1}^m + \cdots + X_{t-11}^m}{12}\right)$$

In order to facilitate the construction of the model, the arithmetic mean will be approximated by the geometric mean.[2]

[2] This approximation is not expected to have a significant effect in the estimation, especially in cases where the rate of growth of the variables are small, leading to a very small difference between the two means. However, in future extensions of this work, it would be interesting to consider exact filtering procedures, such as the one proposed by Aruoba et al. (2009).

$$X_t^A \approx 12 \left(X_t^m X_{t-1}^m \cdots X_{t-11}^m \right)^{1/12}$$
$$\ln X_t^A \approx \ln 12 + \frac{1}{12} \left(\ln X_t^m + \ln X_{t-1}^m + \cdots + \ln X_{t-11}^m \right)$$

The annual growth rate of the target variable can therefore be approximated by

$$x_t^A = \ln X_t^A - \ln X_{t-12}^A \approx \frac{1}{12} \left(\ln X_t^m + \ln X_{t-1}^m + \cdots + \ln X_{t-11}^m \right) - \frac{1}{12} \left(\ln X_{t-12}^m + \ln X_{t-13}^m + \cdots + \ln X_{t-23}^m \right)$$

Rearranging the elements above and letting

$$x_t^m = \ln X_t^m - \ln X_{t-1}^m$$

represent the monthly growth rates, we obtain the following approximation of the annual growth rate of a variable in terms of its (unobservable) monthly growth rates.

$$x_t^A \approx \frac{1}{12} x_t^m + \frac{2}{12} x_{t-1}^m + \cdots + \frac{11}{12} x_{t-10}^m + \frac{12}{12} x_{t-11}^m + \frac{11}{12} x_{t-12}^m + \cdots + \frac{2}{12} x_{t-21}^m + \frac{1}{12} x_{t-22}^m \quad (1)$$

3.1.2 Quarterly variables

Let X_t^Q be a variable observable once per quarter. Parallel to the treatment of annual variables described above, this variable can be approximated by the geometric mean of the last three monthly observations.

$$X_t^Q \approx 3 \left(X_t^m X_{t-1}^m X_{t-2}^m \right)^{1/3}$$

Taking logarithms and rearranging the terms, we obtain the following approximation of x_t^Q, the quarterly growth rate of the variable, in terms of lags of x_t^m, the corresponding (unobservable) monthly growth rate.

$$x_t^Q \approx \frac{1}{3} x_t^m + \frac{2}{3} x_{t-1}^m + \frac{3}{3} x_{t-2}^m + \frac{2}{3} x_{t-3}^m + \frac{1}{3} x_{t-4}^m \quad (2)$$

Quarterly variables are observed only once every three months and the rest of the series is treated as missing.

3.1.3 Monthly variables

Monthly time series expressed as monthly growth rates can be incorporated directly into the model. We will denote them in this paper simply as

$$x_t^m = \ln X_t^m - \ln X_{t-1}^m \quad (3)$$

However, some sources publish their data not as changes with respect to the previous month, but relative to the same month of the previous year. This requires a special transformation before they are incorporated into the model.

Let x_t^{m12} be the rate of change of X_t with respect to the same month of the previous year. Then the difference in this rate over the preceding month can be expressed in terms of monthly growth rates.

$$x_t^{m12} - x_{t-1}^{m12} = \left(\ln X_t^m - \ln X_{t-12}^m\right) - \left(\ln X_{t-1}^m - \ln X_{t-13}^m\right) \\ = x_t^m - x_{t-12}^m \tag{4}$$

3.2 Ragged ends and missing data

The mixed-frequency database that will be used to estimate the model may be affected by the presence of missing data from three sources. First, the series may have some missing information directly from the source and no official imputation is available. Second, some variables are only observed once per year or quarter. Once they are transformed, by following the procedures described above, into functions of monthly growth rates, the data will still be available only when a data point is published and the rest of the series will be treated as missing. And third, because each series has its own starting and ending dates. Although it could be possible to restrict the database to the time window where all the variables are available, this would mean throwing away valuable information that could be used to estimate the model. Moreover, one of the benefits of the nowcasting methodology is that it uses the most recent information available to estimate the target variable, even if this means that only some of the variables are available in the most recent months.

In order to overcome the problem of estimating the model with missing data, we follow Mariano and Murasawa (2003) and substitute all variables by artificial (random normal) observations independent of the model parameters. The authors show that this only adds a constant to the likelihood and does not impact the estimation of the parameters. Let x_t be any of the variables included in the model. This will be substituted by

$$x_t^* = \begin{cases} x_t & \text{if } x_t \text{ is observable} \\ \lambda_t & \text{otherwise} \end{cases} \tag{5}$$

where $\lambda_t \sim \mathcal{N}\left(0, \sigma_\lambda^2\right)$. In this way, for the purpose of estimating the parameters of the model, we will have a database of monthly variables with no empty cells.

3.3 Building the model

With this conformable database, we will structure the model as a dynamic factor model in state-space representation. We therefore assume that the target variables and the different indicators in the model share a common (time-varying) factor ft, in addition to their own idiosyncratic component. We can then apply the Kalman filter and obtain maximum likelihood estimates of the parameters. In this way, we can obtain predictions of the target variable based on the most recent information provided by a list of mixed-frequency indicators.

3.3.1 Annual target variable

To build the state-space model, assume that we have one target variable (y_t^A) available once per year. This will be included in the model by following (1). We also have one indicator of each of the following types: one quarterly variable (x_t^Q) as in (2), one monthly variable measured as month-on-month growth rates (x_t^m) as in (3), one monthly variables measured as year-on-year growth rates (x_t^{m12}) as in (4), and one forecast of the target variable ($\hat{y}_t^A$). If there is more than one variable of any type available, it would suffice to add the corresponding rows to the Kalman matrices.

Following the transformations described in the previous section, we obtain the following measurement equation of the model.

$$\begin{pmatrix} y_t^A \\ x_t^Q \\ x_t^m \\ x_t^{m12} - x_{t-1}^{m12} \\ \hat{y}_t^A \end{pmatrix} = \begin{pmatrix} \beta_1\left(\frac{1}{12}f_t + \cdots + \frac{12}{12}f_{t-11} + \cdots + \frac{1}{12}f_{t-22}\right) \\ \beta_2\left(\frac{1}{3}f_t + \frac{2}{3}f_{t-1} + \frac{3}{3}f_{t-2} + \frac{2}{3}f_{t-3} + \frac{1}{3}f_{t-4}\right) \\ \beta_3 f_t \\ \beta_4\left(f_t - f_{t-12}\right) \\ \beta_1\left(\frac{1}{12}f_t + \cdots + \frac{12}{12}f_{t-11} + \cdots + \frac{1}{12}f_{t-22}\right) \end{pmatrix} +$$

$$\begin{pmatrix} \frac{1}{12}\epsilon_{1,t} + \cdots + \frac{12}{12}\epsilon_{1,t-11} + \cdots + \frac{1}{12}\epsilon_{1,t-22} \\ \frac{1}{3}\epsilon_{2,t} + \frac{2}{3}\epsilon_{2,t-1} + \frac{3}{3}\epsilon_{2,t-2} + \frac{2}{3}\epsilon_{2,t-3} + \frac{1}{3}\epsilon_{2,t-4} \\ \epsilon_{3,t} \\ \epsilon_{4,t} - \epsilon_{4,t-12} \\ \frac{1}{12}\epsilon_{1,t} + \cdots + \frac{12}{12}\epsilon_{1,t-11} + \cdots + \frac{1}{12}\epsilon_{1,t-22} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ fce_t \end{pmatrix}$$

This can be written more compactly as

$$Gt = Bht + ut \tag{6}$$

with

$$h_t = (f_t, \ldots, f_{t-23}, \epsilon_{1,t}, \ldots, \epsilon_{1,t-23}, \epsilon_{2,t}, \ldots, \epsilon_{2,t-5}, \epsilon_{3,t}, \ldots, \epsilon_{3,t-m_3}, \epsilon_{4,t}, \ldots, \epsilon_{4,t-13}, fce_t)$$

The factor loading matrix is given by

$$B = \begin{bmatrix} B_{11} & B_{12} & \mathbf{0}_{1\times6} & \mathbf{0}_{1\times m_3} & \mathbf{0}_{1\times14} & 0 \\ B_{21} & \mathbf{0}_{1\times24} & B_{22} & \mathbf{0}_{1\times m_3} & \mathbf{0}_{1\times14} & 0 \\ B_{31} & \mathbf{0}_{1\times24} & \mathbf{0}_{1\times6} & B_{32} & \mathbf{0}_{1\times14} & 0 \\ B_{41} & \mathbf{0}_{1\times24} & \mathbf{0}_{1\times6} & \mathbf{0}_{1\times m_3} & B_{42} & 0 \\ B_{11} & B_{12} & \mathbf{0}_{1\times6} & \mathbf{0}_{1\times m_3} & \mathbf{0}_{1\times14} & 1 \end{bmatrix}$$

$$B_{11} = \beta_1\left(\frac{1}{12}, \ldots, \frac{12}{12}, \ldots, \frac{1}{12}, 0\right), \quad B_{12} = \left(\frac{1}{12}, \ldots, \frac{12}{12}, \ldots, \frac{1}{12}, 0\right)$$

$$B_{21} = \beta_2\left(\frac{1}{3}, \frac{2}{3}, \frac{3}{3}, \frac{2}{3}, \frac{1}{3}, 0, \ldots, 0\right), \quad B_{22} = \left(\frac{1}{3}, \frac{2}{3}, \frac{3}{3}, \frac{2}{3}, \frac{1}{3}, 0\right)$$

$$B_{31} = \beta_3\left(1, 0, \ldots, 0\right), \quad B_{32} = \left(1, 0, \ldots, 0\right)$$