




# A dynamic factor model for the Mexican economy: are common trends useful when predicting economic activity?

Francisco Corona<sup>1</sup>  · Graciela González-Farías<sup>2</sup> · Pedro Orraca<sup>3</sup>

Received: 26 January 2017 / Revised: 6 August 2017 / Accepted: 30 August 2017 /  
Published online: 15 September 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** In this paper we propose to use the common trends of the Mexican economy in order to predict economic activity one and two steps ahead. We exploit the cointegration properties of the macroeconomic time series, such that, when the series are  $I(1)$  and cointegrated, there is a factor representation, where the common factors are the common trends of the macroeconomic variables. Thus, we estimate a large non-stationary dynamic factor model using principal components (PC) as suggested by Bai (J Econom 122(1):137–183, 2004), where the estimated common factors are used in a factor-augmented vector autoregressive model to forecast the Global Index of Economic Activity. Additionally, we estimate the common trends through partial least squares. The results indicate that the common trends are useful to predict Mexican economic activity, and reduce the forecast error with respect to benchmark models, mainly when estimated using PC.

**Keywords** Dynamic factor models · Common trends · Factor-augmented vector autoregressive model · Partial least squares · Forecast error

---

✉ Francisco Corona  
fcorona@est-econ.uc3m.es

Graciela González-Farías  
farias@cimat.mx

Pedro Orraca  
porraca@colef.mx

<sup>1</sup> Departamento de Estadística, Universidad Carlos III de Madrid, Calle Madrid, 126, 28903 Getafe, Madrid, Spain

<sup>2</sup> Centro de Investigación en Matemáticas, A.C., Jalisco s-n, Mineral de Valenciana, 36240 Guanajuato, Guanajuato, Mexico

<sup>3</sup> Departamento de Estudios Económicos, El Colegio de la Frontera Norte, Carretera Escénica Tijuana-Ensenada Km 18.5, 22560 Tijuana, Baja California, Mexico

**JEL Classification** C38 · C53 · E00

## 1 Introduction

In recent years, due to the availability of data on a vast number of correlated macroeconomic and financial variables collected regularly by statistical agencies, there has been an increasing interest in modeling large systems of economic time series. Therefore, econometricians have to deal with datasets consisting of hundreds of series, thus making the use of large dimensional dynamic factor models (DFMs) more attractive than the usual vector autoregressive (VAR) models, which usually limit the number of variables; see Boivin and Ng (2006). DFMs were introduced by Geweke (1977) and Sargent and Sims (1977) with the aim of representing the dynamics of large systems of time series through a small number of hidden common factors, and are mainly used for one of the following two objectives: first, forecasting macroeconomic variables and second, estimating the underlying factor in order to carry out policy-making (e.g., the business cycle; lagging, coincident, and leading indicators; or impulse-response functions, among other aspects). Another interesting application is to use the common factors as instrumental variables or exogenous regressors in panel data analysis. See Bai and Ng (2008), Stock and Watson (2011), and Breitung and Choi (2013) for a review of existing literature and applications of DFMs.

Note that macroeconomic time series are generally non-stationary and frequently cointegrated; see, for example, Kunst and Neusser (1997). On the other hand, a cointegrated system can be expressed in terms of a factor representation; see Stock and Watson (1988), Vahid and Engle (1993), Gonzalo and Granger (1995). Furthermore, these authors show that the common factor representation implies that the variables of the system are cointegrated if the common factors are  $I(1)$  and the individual effects are  $I(0)$ . However, in practice, the cointegration results are not frequently used to predict macroeconomic variables, given that we differentiate each element of the factor representation individually.

Empirically, it is interesting to determine the number of common factors of the economy, given that we can summarize a large number of macroeconomic variables in few common trends.<sup>1</sup> Furthermore, by estimating the common factors we can observe the orthogonal dynamics of the economy, which is very important for macroeconomic policy. These common factors, for example, may be related to specific groups of variables, and it is interesting to analyze mechanisms of transmission to other groups of variables. Additionally, we can estimate the confidence intervals for the loading weights, common factors, and common components, which allow us to carry out statistical inference in order to better understand economic phenomena.

The stochastic common trends represent the long-run behavior of the variables. Previous studies related to this paper are Duy and Thoma (1998), who evaluate the

---

<sup>1</sup> The common factors are equivalent to the common trends when we assume a static representation in the factor model. Barigozzi et al. (2016) develop the econometric theory for non-stationary DFMs with large datasets under different assumptions of the dynamic of the common factors.

improvement of the cointegration relationship in VAR models; Eickmeier and Ziegler (2008), who examine the use of the factor-augmented vector autoregressive (FAVAR) in order to predict macroeconomic and financial variables, focusing on the US economy. They conclude that on average, factor forecasts are slightly better than other models. Banerjee et al. (2014) incorporate the error correction term in FAVAR models, showing that their approach generally offers higher forecasting accuracy relative to the FAVAR. Also, Eickmeier et al. (2014) prove different specifications of FAVAR models, using time-varying coefficients (tv) and stochastic volatility errors. The main conclusion is that the FAVAR models, regardless of the specification, outperform univariate models on in-sample forecasts. When they evaluate the out-sample performance, there is no improvement with respect to the benchmark model. More recently, Hindrayanto et al. (2016) study the forecasting performance of a four factor model approach for large datasets. They conclude that a collapsed DFM is the most accurate model. Finally, Wilms and Croux (2016) propose a sparse cointegration method for a large set of variables by setting some coefficients in the cointegration relationship at exactly zero. They conclude that their method leads to significant forecast accuracy with respect to the method proposed by Johansen (1988, 1991). Other samples of reference are Stock and Watson (2002a, b), Marcellino et al. (2003), Peña and Poncela (2004), Reijer (2005), Schumacher (2007), Giannone et al. (2008), Eickmeier et al. (2014), Lahiri et al. (2015), and Panopoulou and Vrontos (2015), who show that DFMs are useful in order to reduce the forecast errors with respect to the traditional models, as autoregressive approaches and linear regressions with macroeconomic diffusion indexes.<sup>2</sup> It is important to comment that, in a similar way as we have proposed in this study but in a stationary framework, Bräuning and Koopman (2014) propose to use collapsed dynamic factor analysis in order to predict target variables of the economy, exploiting the state-space representation between the target variables and the common factors. They conclude that the forecast accuracy is improved with respect to benchmark models.

However, most empirical applications have been carried out for the US or Euro economies. Therefore, we propose a FAVAR model to predict Mexican economic activity using concepts of cointegration such that we forecast  $I(1)$  time series where the estimated common factors are the common trends of the economy. We readily exploit the long-run relationship between the common factors and the specific variable that we wish to predict.<sup>3</sup>

We contribute to the literature in two ways: (1) empirically in determining the number of common trends of the Mexican economy and (2) the use of the common trends in order to predict economic activity. This is an advantage with respect to

---

<sup>2</sup> Certainly, macroeconomic diffusion indexes are related to DFMs. However, in this study we make a distinction between multivariate predictive approaches and the classic diffusion index forecasts given by Stock and Watson (2002b). Furthermore, we focus on a DFM from a static point of view. Once the common factors are estimated and their non-stationarity disentangled, we use the factors in the proposed FAVAR representation.

<sup>3</sup> A related study is Caruso (2015), who examines the flow of conjunctural data relevant to assess the state of Mexico's economy. Specifically, the author exploits the information embedded in macroeconomic news from both Mexico and the US, in a model constructed to nowcast Mexican real Gross Domestic Product (GDP) based on DFMs and Kalman filters.

other alternatives, given that the FAVAR with cointegrated variables is very straightforward to implement computationally. Furthermore, we also direct our attention to the predictive capacity of macroeconomic variables and their common trends. In this way, we use the traditional method of principal components (PC) and additionally, partial least squares (PLS).

The rest of this paper is structured as follows. In Sect. 2, we summarize the historical behavior of the Mexican economy. In Sect. 3, we briefly explain the DFM and the FAVAR representation. In Sect. 4, we describe the estimation methods used. In Sect. 5, we describe the data, present the cointegration results, their descriptive predictive features, the empirical strategy to obtain small forecast errors, and the results obtained. Finally, we conclude in Sect. 6.

## 2 The Mexican economy

Mexico is one of the world's leading emerging countries. According to data from the International Monetary Fund (IMF), it is the 11th largest economy and the largest emerging country outside the BRICs, i.e., Brazil, Russia, India, and China (Rodríguez-Pose and Villarreal 2015). In 2015, Mexico's population ascended to 119.9 million and its GDP per capita stood at 9005.0 US dollars (World Bank 2016b).

After a prolonged period of sustained economic growth, the country began decelerating in the late 1970s. Following a sovereign default, Mexico encountered a severe economic crisis from 1982 to 1985. Subsequently, the Mexican government put forward a series of market reforms that culminated with the implementation of the North American Free Trade Agreement in 1994. These reforms included the privatization of previously government-operated firms, fiscal reforms, and opening the country to trade and foreign direct investment (FDI), among other actions (Kehoe and Ruhl 2010). Mexico's economic transformation was successful in reducing inflation, maintaining fiscal discipline, reducing its external debt burden, and increasing trade as a share of GDP (Hanson 2010). Nonetheless, this was not accompanied by high levels of economic growth, as between 1982 and 2015 the country's GDP grew at an average annual rate of 2.3% (World Bank 2016a).

Moreover, in recent decades the Mexican economy has regressed on different fronts. In 1995, Mexico was responsible for 7.0% of trade among economies classified by the IMF as emerging and developing, behind China at 12.7%. By 2008, China's share was still the largest and totaled 22.3%, while Mexico dropped to third at 5.4%, behind Russia. Additionally, in 1995, China was the largest recipient of FDI to emerging and developing economies, accounting for 33.4%, while Mexico was second, receiving 8.5%. By 2008, Mexico's position had weakened considerably, falling to seventh at 3.2% (Kehoe and Ruhl 2010). Factors, such as rigidities in the labor market, an inefficient financial sector, and a lack of contract enforcement, have limited Mexico's capacity to benefit from its reforms and therefore hindered the country's economic growth (Kehoe and Ruhl 2010).

As an economy open to world markets, Mexico hosts many modern firms, notably in the automobile, aerospace, and foods and beverages sectors, which

employ highly skilled, well-educated, and well-remunerated workers. Nonetheless, this represents only a small part of the economy and is concentrated in a few regions of the country. Another segment of the economy is characterized by high levels of informality, low-skilled work, low productivity levels, and out-of-date technologies. Unregistered firms in the informal sector employ close to half of Mexico's labor force, with workers that frequently lack access to a good education, reliable health care, and affordable financial services, conditions that strongly hamper their accumulation of human capital. Lastly, a third segment of the economy is made up of companies that, for long periods, have been protected from competition, particularly dominant firms in the energy and telecommunications sectors (Dougherty 2015).

During the global economic crisis of 2008–2009, Mexico's GDP suffered a significant contraction, which was accentuated by a reduction in the amount of remittances the country receives from migrant workers based in the US and the outbreak of influenza A(H1N1) throughout the country. Nevertheless, due to an upgraded macroeconomic policy framework and careful regulation of the financial system, Mexico did not suffer the type of financial and fiscal crises experienced in other countries (Schwellnus 2011).

In the years following the global economic crisis, the Mexican economy was characterized by a persistent trend of increasing debt-to-GDP, which grew from 29.0% in 2007 to 50.5% in 2016, and a significant decline in the volume of oil production (World Bank 2016a). Output volatility in Mexico remained high, which can have high costs for individuals and for long-term growth. Furthermore, in Mexico, temporary disturbances in output are usually accompanied by temporary reductions in consumption, since a substantial share of the population is credit-constrained and the social safety net is generally weak. This is costly for individuals who prefer a smooth path of consumption and are averse to periods of unemployment or poverty (Schwellnus 2011).

Finally, during the last decade, Mexico has implemented a more ambitious and wide-ranging innovation policy aimed at getting closer to the technological frontier and generating higher levels of economic growth (Rodriguez-Pose and Villarreal 2015). By 2016, the expansion of economic activity was highly dependent on private consumption as low levels of investment and export demand were scarcely contributing to growth. In the medium term, economic and financial stability and an increase in external competitiveness, derived from the depreciation of the country's currency, are expected to boost private investment, exports, and hence economic growth (World Bank 2016b).

In this context, it is very important to formulate an econometric approach in order to successfully predict Mexican economic activity, which takes into account a large number of macroeconomic and financial variables associated with the behavior of the economy. Hence, we use a large non-stationary DFM in which the variables are summarized in the common trends of the Mexican economy. These common trends are used in a FAVAR representation, evaluating the out-sample forecast error one and two steps ahead during a recent phase of the economy.

### 3 The models

In this section we introduce notation and describe the non-stationary DFM and the FAVAR model in order to use the common trends to predict a specific macroeconomic variable.

#### 3.1 Non-stationary dynamic factor model

Suppose that  $N$  large economic time series  $Y_t = (y_{1t}, \dots, y_{Nt})'$ , observed from  $t = 1, \dots, T$ , are  $I(1)$  and any two series  $y_{it}$  and  $y_{jt}$  are cointegrated, such that the idiosyncratic component  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$  is stationary. In this case, and following Stock and Watson (1988), there is a common factor representation as follows:

$$Y_t = PF_t + \varepsilon_t, \quad (1)$$

where  $P = (p'_1, \dots, p'_N)'$  is the  $N \times r$  ( $r < N$ ) matrix of factor loadings, where  $p_i = (p_{i1}, \dots, p_{ir})$  and  $F_t = (F_{1t}, \dots, F_{rt})'$  is the  $r \times 1$  vector of common factors, or in this case, the common trends.

Given that any two series of  $Y_t$  are cointegrated in the sense of Engle and Granger (1987), the dynamic of the common factors and the idiosyncratic components are the following:

$$F_t = F_{t-1} + \eta_t \quad (2)$$

$$\varepsilon_t = \Gamma(L)\varepsilon_{t-1} + a_t \quad (3)$$

where the factor disturbances,  $\eta_t = (\eta_{1t}, \dots, \eta_{rt})'$ , are  $r \times 1$  vectors, distributed independently from the idiosyncratic noises for all leads and lags. Furthermore,  $\eta_t$  and  $a_t$  are white noises with positive definite covariance matrices  $\Sigma_\eta$  and  $\Sigma_a$ , respectively. Additionally,  $\Gamma(L)$  collected the  $N \times N$  matrices containing the autoregressive parameters of the idiosyncratic components where  $L$  is the lag operator such that  $L^k \varepsilon_t = \varepsilon_{t-k}$ . These autoregressive matrices satisfy the usual stationarity assumptions.

The DFM in Eqs. (1)–(3) is not identified because for any  $r \times r$  non-singular matrix  $H$ , the system can be expressed in terms of a new loading matrix and a new set of common factors as follows:

$$Y_t = P^* F_t^* + \varepsilon_t \quad (4)$$

$$F_t^* = F_{t-1}^* + \eta_t^*, \quad (5)$$

where  $P^* = PH$ ,  $F_t^* = H^{-1}F_t$ , and  $\eta_t^* = H^{-1}\eta_t$ . The DFM in Eqs. (4) and (5) is observationally equivalent to that in Eqs. (1) and (2). A normalization is necessary to solve this identification problem and uniquely define the factors. In the context of PC factor extraction, it is common to impose the restriction  $P'P/N = I_r$  and  $F'F$  being diagonal, where  $F = (F_1, \dots, F_T)'$  is a  $T \times r$  matrix of common factors. Alternatively, we can set the restrictions  $F'F/T^2 = I_r$  and  $P'P$  being diagonal; see Bai (2004) for identification issues in the non-stationary case.

In Eqs. (1)–(3), if  $\Gamma = 0$  and  $\Sigma_a$  is diagonal, then the DFM is known as *strict*; see Bai and Ng (2008). On the other hand, when  $\Gamma \neq 0$  and  $\Sigma_a$  is diagonal, there is serial correlation in the idiosyncratic noises and in this case the DFM is called *exact*; see Stock and Watson (2011). Chamberlain and Rothschild (1983) introduce the term *approximate model* when the idiosyncratic term does not need to have a diagonal covariance matrix. Forni et al. (2000) and Stock and Watson (2002a) generalize the approximate model, allowing for weak serial and cross-correlation. Furthermore, Bai and Ng (2002) and Bai (2003) also allow for heteroskedasticity in both the time and cross-sectional dimensions.

### 3.2 The factor-augmented autoregressive model

Once the non-stationary DFM is specified, we can use the common trends to predict a specific macroeconomic variable, denoted as  $x_t$ ; see Stock and Watson (2005), Banerjee et al. (2013), among many others. It is important to comment that if forecasting is the objective, although  $(x_t, F_t)'$  are integrated or cointegrated, the FAVAR specification is quite convenient. Note that the optimal conditional expectation does not require stationarity or stability in the system. The theoretical justification is given by Lütkepohl (2006). More recently, Barigozzi et al. (2016) discuss several situations where a VAR estimated in levels is equivalent to a cointegrated VAR. Basically, when estimating a VAR in levels we can consistently estimate the parameters of a cointegrated VAR. Therefore, the FAVAR is provided as follows:

$$\begin{pmatrix} x_t \\ F_t \end{pmatrix} = \Phi(L) \begin{pmatrix} x_{t-1} \\ F_{t-1} \end{pmatrix} + v_t, \tag{6}$$

where

$$\Phi(L) = \begin{pmatrix} \Phi_{11}^{(k)} & \Phi_{12}^{(k)} & \Phi_{13}^{(k)} & \cdots & \Phi_{1r+1}^{(k)} \\ 0 & \Phi_{22}^{(k)} & 0 & \cdots & 0 \\ 0 & 0 & \Phi_{33}^{(k)} & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & \Phi_{r+1r+1}^{(k)} \end{pmatrix} L^{k+1},$$

for  $k = 1, \dots, p$ , where  $v_t \sim N(0, \Sigma_v)$ . Intuitively, common economic trends that summarize its long-run behavior are used to predict a specific macroeconomic variable. For a similar representation but in the stationary context, see Bräuning and Koopman (2014). Note that expression (6) is the prediction equation; for example, assuming  $p = 1$ , for any  $h$  step ahead, then

$$x_{t+h} = \Phi_{11}x_{t-(h+1)} + \Phi_{12}F_{1,t-(h+1)} + \cdots + \Phi_{1r+1}F_{r+1,t-(h+1)} + v_{t+h}. \tag{7}$$

Additionally, we can introduce deterministic components and seasonal and exogenous variables in Eq. (6).

## 4 Estimation

The estimation of the parameters of Eq. (6) is obtained using restricted ordinary least squares (OLS). Furthermore, it is necessary to define the number of common factors and their factor estimates. In this section we describe some traditional methods to determine the number of non-stationary common factors. Additionally, we describe PC factor extraction and comment on an alternative method to estimate the common factors, known in literature as PLS.

### 4.1 Determining the number of factors

It was previously assumed that  $r$  is known, however in practice it is frequently necessary to estimate it. In this subsection we describe two procedures to determine  $r$  in the context of large DFMs, namely the Onatski (2010) procedure and the ratio of eigenvalues proposed by Ahn and Horenstein (2013). There are other alternatives to determine  $r$ , the Bai and Ng (2002) information criteria, the respective correction proposed by Alessi et al. (2010), and the procedure carried out by Kapetanios (2010), among many other approaches. Moreover, in the context of non-stationary DFMs, Ergemen and Rodriguez-Caballero (2016) use the procedure given by Hallin and Liska (2007) to determine the number of regional and global factors allowing fractional differencing in  $Y_t$ . However, and following Corona et al. (2017a), they show that the finite sample performance of the proposed procedures in this paper exhibits a good performance when we use data in first differences or levels and the common factors are  $I(1)$ .<sup>4</sup>

#### 4.1.1 Onatski (2010) procedure

The Onatski (2010) procedure is based on the behavior of two adjacent eigenvalues of  $\hat{\Sigma}_Y = T^{-1}Y'Y$ , for  $j = 1, \dots, r_{\max}$ . Intuitively, it is reasonable that when  $N$  and  $T$  tend to infinity, the difference between  $\hat{\lambda}_j - \hat{\lambda}_{j+1}$  tends to zero while  $\hat{\lambda}_r - \hat{\lambda}_{r+1}$  diverges to infinity, where  $\hat{\lambda}_i$  is the  $i$ -largest eigenvalue of  $\hat{\Sigma}_Y$ . Onatski (2010) points out the necessity of determining a “sharp” threshold that separates convergent and divergent eigenvalues, denoted as  $\delta$ . The author gives the empirical procedure to determine the threshold and in practice, this approach is more robust in presence of non-stationarity and when the factors are weak or when the proportion of variance attributed to the idiosyncratic components is larger than the variance of the common component. We denote this estimator as  $\hat{r}_{ED} = \max\{j \leq r_{\max} : \hat{\lambda}_j - \hat{\lambda}_{j+1} \geq \delta\}$ .

#### 4.1.2 The Ahn and Horenstein (2013) ratios of eigenvalues

Ahn and Horenstein (2013) develop the consistency of the estimation of  $r$  when they use the ratio of two adjacent eigenvalues of  $\hat{\Sigma}_Y$  under the traditional assumptions of

<sup>4</sup> Alternatively, Bai (2004) proposes an information criteria for data in levels, modifying the PC( $k$ ) information criteria of Bai and Ng (2002) by multiplying the penalty function by  $\alpha_T = T/[4 \log \log(T)]$ .



PC factor extraction. The two criteria provided by the authors are based on maximizing with respect to  $j = 1, \dots, r_{\max}$  the following ratios:

$$ER(j) = \frac{\hat{\lambda}_j}{\hat{\lambda}_{j+1}}, \tag{8}$$

$$GR(j) = \frac{\ln(1 + \hat{\lambda}_j^*)}{\ln(1 + \hat{\lambda}_{j+1}^*)}, \tag{9}$$

where  $\hat{\lambda}_0 = \frac{1}{m} \sum_{i=1}^m \hat{\lambda}_j / \ln(m)$  and  $\hat{\lambda}_j^* = \hat{\lambda}_j / \sum_{i=k+1}^m \hat{\lambda}_i$ . The value of  $\hat{\lambda}_0$  has been chosen following the definition of Ahn and Horenstein (2013), according to which  $\hat{\lambda}_0 \rightarrow 0$  and  $m\hat{\lambda}_0 \rightarrow \infty$  as  $m \rightarrow \infty$ .

### 4.2 Principal components factor extraction

The most popular method to extract static common factors is based on PC given that it does not require assumptions of the error distribution, and the estimation of the common component is consistent, among many other properties; see Bai (2003). This procedure separates the common component from the idiosyncratic noises by considering a cross-sectional averaging of the variables within  $Y_t$  such that when  $N$  and  $T$  tend simultaneously to infinity, the weighted averages of the idiosyncratic noises converge to zero, with only the linear combinations of the factors remaining. Therefore, this method requires that the cumulative effects of the common component increase proportionally with  $N$ , while the eigenvalues associated with the idiosyncratic components remain bounded.

The PC estimator of  $F_t$  can be derived as the solution to the following least squares problem:

$$V_r(P, F) = \min_{F_1, \dots, F_T, P} (NT)^{-1} \sum_{t=1}^T (Y_t - PF_t)'(Y_t - PF_t) \tag{10}$$

subject to the restrictions  $F'F/T^2 = I_r$  and  $P'P$  being diagonal, where  $F = (F_1, \dots, F_T)'$  is a  $r \times T$  matrix of common trends. The minimization problem in (10) is equivalent to maximizing  $\text{tr}[F'(YY')F]$  where  $Y = (Y_1, \dots, Y_T)'$  with the dimension  $T \times N$ . This maximization problem is solved by setting  $\hat{F}$  equal to  $T$  times the eigenvectors corresponding to the  $r$  largest eigenvalues of the  $T \times T$  matrix  $YY'$ . The corresponding PC estimator of  $P$  is given by<sup>5</sup>:

<sup>5</sup> Alternatively, if we choose the restrictions  $P'P/N = I_r$  and  $F'F$  being diagonal, the estimator of the matrix of factor loadings,  $\hat{P}$ , is  $\sqrt{N}$  times the eigenvectors corresponding to the  $r$  largest eigenvalues of the  $N \times N$  matrix  $Y'Y$ , with estimated factor matrix  $\hat{F} = Y\hat{P}/N$ . The difference is only computational, these latest restrictions are less costly when  $T > N$ , whereas  $F'F/T^2 = I_r$  with  $P'P$  being diagonal are less costly when  $T < N$ .

$$\hat{P} = \frac{Y' \hat{F}}{T^2}. \quad (11)$$

Bai (2003) deduces that the rate of convergence and the limiting distributions of the estimated factors, factor loadings and common component is the stationary DFM when the cross-section and time dimensions tend towards infinity. Under the assumptions considered in this paper, Bai (2004) shows the consistency of  $\hat{F}_t$ ,  $\hat{P}$ , and the common component. The finite sample performance of  $\hat{F}_t$  is analyzed by Bai (2004) and recently by Corona et al. (2017b). The first author concludes that  $\hat{F}_t$  is a close estimation of  $F_t$  when the idiosyncratic errors are stationary and even if the sample size is moderately small (i.e.,  $N = 100$  and  $T = 40$ ). The second authors consider structure dependence in the idiosyncratic errors and show that when the idiosyncratic errors are  $I(0)$  we can obtain close estimations of  $F_t$  even if  $N$  is small and the variance of  $\varepsilon_t$  is large. When the idiosyncratic errors are  $I(1)$ ,  $\hat{F}_t$  works poorly. In this case, it is convenient to use the estimator given by Bai and Ng (2004).<sup>6</sup> Other alternatives to estimate the common factors without differencing them are given by Barigozzi et al. (2016) and Corona et al. (2017b).

Note that the estimation of  $F_t$  disregards the dynamic of Eqs. (2) and (3). The reason for this is that we estimate the common factors from static representation in the factor model. Corona et al. (2017b) incorporate the dynamic of Eq. (2) in a second step to obtain smooth estimates of the common factors using the Kalman filter. However, in this study we focus on a large DFM from static representation. For a technical discussion of the analogies and differences between static and dynamic representation, see for example Bai and Ng (2007). Thus, our approach is related to Bai (2004) and Bai and Ng (2004), who study non-stationarity in DFMs under a static representation.

### 4.3 Partial least squares

In addition to the estimation of  $F_t$  by PC, we consider the PLS estimation, which takes into account the effect of a dependent variable. In a similar manner to Fuentes et al. (2015), we estimate the common factors in economic time series using PLS, denoted as  $\tilde{F}_t$ . Intuitively, the idea is to find the orthogonal latent variables that maximize the covariance between  $Y$  and  $x = (x_1, \dots, x_T)'$ . The estimation process is iterative, and the first step consists in the eigenvalue decomposition of the following  $T \times T$  matrix:

$$M = Yxx'Y' \quad (12)$$

Then, the first common factor,  $\tilde{F}_{1t}$ , is the first eigenvector associated to the first eigenvalue of  $M$ . The second common factor is estimated from the residual matrix of  $e = Y - B\tilde{F}'_1$ . To obtain the following common factors, the process is repeated  $r - 2$  steps using the  $\tilde{F}_{1t}, \dots, \tilde{F}_{r-1t}$  common factors in each step.

<sup>6</sup> According to Bai and Ng (2004), if  $\varepsilon_t$  is not stationary, it is necessary to differentiate  $Y_t$  to consistently estimate  $F_t$  using  $\hat{F}_t = \sum_{s=1}^t \hat{f}_s$  where  $\hat{f}_t$  is the PC factor estimated when using data in first differences.

## 5 Empirical analysis

In this section we present the data from the Mexican economy used in this study, describe the methodology to evaluate the accuracy of the common trends in order to predict the Mexican economy, determine the number of common trends, and analyze their dynamic by evaluating the performance of the FAVAR model proposed in Sect. 2 with respect to benchmark models.

### 5.1 Data

Initially, we consider 511 macroeconomic and financial variables obtained from the *Banco de Información Económica* (BIE) of the *Instituto Nacional de Geografía y Estadística* (INEGI), Mexico's national statistical agency. The analysis covers from March 2005 to April 2016, hence,  $T = 133$ . The blocks of variables are considered according to the INEGI division, and additionally, we compare this division with respect to the National Institute's Global Econometric Model, which considers nine blocks with a total of 67 variables.<sup>7</sup>

According to our approach, it is necessary that all variables are integrated of order one, so that, if factors are found, these are the common trends of the observations. In this case, we consider only  $I(1)$  variables according to the Augmented Dickey–Fuller (ADF) test.<sup>8</sup> When needed, the time series have been deseasonalized and corrected by outliers using X-13ARIMA-SEATS developed by the US Census Bureau.<sup>9</sup> Following Stock and Watson (2005), outliers are substituted by the median of the five previous observations.

Finally, according to these non-stationary conditions, we work with the following database of  $N = 211$  variables (number between parentheses)<sup>10</sup>:

- Balance of trade (19)
- Consumer confidence (18)
- Consumption (9)
- Economic activity (13)
- Employment (5)
- Financial (35)
- Industrial (58)
- International (17)
- Investment (8)
- Miscellaneous (18)
- Prices (11).<sup>11</sup>

---

<sup>7</sup> <https://nimodel.niesr.ac.uk>.

<sup>8</sup> See Barigozzi et al. (2016) for a similar approach.

<sup>9</sup> <https://cran.r-project.org/web/packages/seasonal.pdf>.

<sup>10</sup> See Annex with the description of each block of variables included.

<sup>11</sup> It is well known that the prices can be  $I(2)$ . In our sample period seven prices are  $I(2)$  and they are differenced.

Furthermore, we define  $x_t$  as the Global Index of Economic Activity (IGAE, *Indicador Global de la Actividad Económica*).<sup>12</sup>

## 5.2 Estimating the common trends

We apply the criteria described in this article to detect the number of factors. We use an  $r_{\max} = 11$ . The results indicate that  $\hat{r}_{ER} = \hat{r}_{GR} = 2$  and  $\hat{r}_{ED} = 5$ .<sup>13</sup> Given that Onatski (2010) is more robust in the presence of non-stationarity, see Corona et al. (2017a), we work with this number of factors.<sup>14</sup>

Figure 1 plots the results of the ratios and differences of eigenvalues with the respective threshold. Note that, intuitively, it is congruent that the ratio of eigenvalues of Ahn and Horenstein (2013) determines one and two common factors, given that the first and second differences of eigenvalues are large, while the others are practically zero. However, the estimation of the “sharp” threshold is one of the main contributions from Onatski (2010), which as we have mentioned, consistently separates convergent and divergent eigenvalues. Furthermore, the five common factors explain 79.19% of the total variability. Specifically, the first common factor explains 46.14%, the second 19.59%, the third 6.2%, the fourth 4.75%, and the fifth 2.51%.

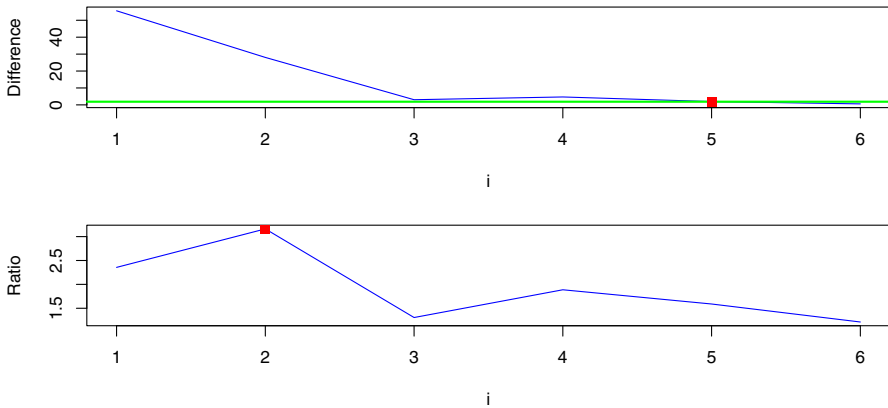
Figure 2 plots the behaviors between  $\ln x_t$  (deseasonalized) and each common factor extracted by PC and PLS. We observe that the first common factor, for each procedure, is similar to  $\ln x_t$  with contemporaneous linear correlations of 0.95 for PC and 0.96 for PLS. The second common factors are slightly correlated with  $\ln x_t$ , having contemporaneous correlations of 0.13 and 0.08 for PC and PLS, respectively. On the other hand,  $\tilde{F}_{3t}$  and  $\tilde{F}_{5t}$  have an inverse behavior, although the correlations with respect to  $\ln x_t$  are around  $-0.01$ . In the other cases, the estimated common factors are positively related to  $\ln x_t$  but the linear correlations are drastically small. These facts indicate that contemporaneously, only the first two common factors are associated with economic activity. It is interesting to mention that these two factors explain 65.73% of the total variability. Note that it is complicated to establish the predictive capacity between the common factors, and it is necessary to evaluate it in the FAVAR model. Furthermore, it is clear that the sample period includes the economic crisis of 2008–2009; however, Stock and Watson (2011) show that the PC estimator of the factors is consistent even with certain types of breaks or time variation in the factor loadings.

Figure 3 shows the weighted average contribution of each variable block in the common factors estimated by PC. Each color bar represents the contribution of the percentage of explanation of each variable group with respect to each common factor, denoted as follows:

<sup>12</sup> Note that the block of economic activity includes the components of the aggregate IGAE. Hence, this specific variable is not included in the block.

<sup>13</sup> Additionally, we estimate the Bai and Ng (2002) information criteria using data in levels and first-differenced data. In both cases, the three criteria tend to  $\hat{r} = r_{\max}$ .

<sup>14</sup> For the US economy, Alessi et al. (2010) determine six common factors using the database from Stock and Watson (2005).



**Fig. 1** Determination of the number of factors through the Onatski (2010) procedure,  $\hat{\lambda}_i - \hat{\lambda}_{i+1}$  (top panel) and ratio of eigenvalues (bottom panel),  $\hat{\lambda}_{i+1}/\hat{\lambda}_i$ , for  $i = 1, \dots, 6$

$$\hat{\mu}_{jg} = \sum_{i_g=1}^{N_g} |\hat{p}_{i_g jg}| / N_g \quad \text{for } j = 1, \dots, r, \text{ and } g = 1, \dots, G,$$

where  $\hat{p}_{i_g jg}$  is the loading weight of each group of variables;  $G$  is the number of blocks of variables; and  $N_g$  is the number of variables in each group. Specifically, the order of the groups of variables is computed as  $r^{-1} \sum_{j=1}^r \hat{\lambda}_j \hat{\mu}_{jg}$  where  $\hat{\lambda}_j$  is the variance contribution of each common factor. The block that most explains the common factors is the miscellaneous group, economic activity, and balance of trade blocks. On the other hand, prices, employment, and consumer confidence are the least relevant groups of variables. Note that this importance is in terms of the loading contribution and it is not interpreted as predictive power. Specifically, the first common factor is more correlated with the IGAE of tertiary activities (0.99), the second common factor with the economic situation with respect to last year (0.94), the third common factor with oil exports (0.78), the fourth common factor with edification (0.70), and the fifth common factor with the food industry (0.53).

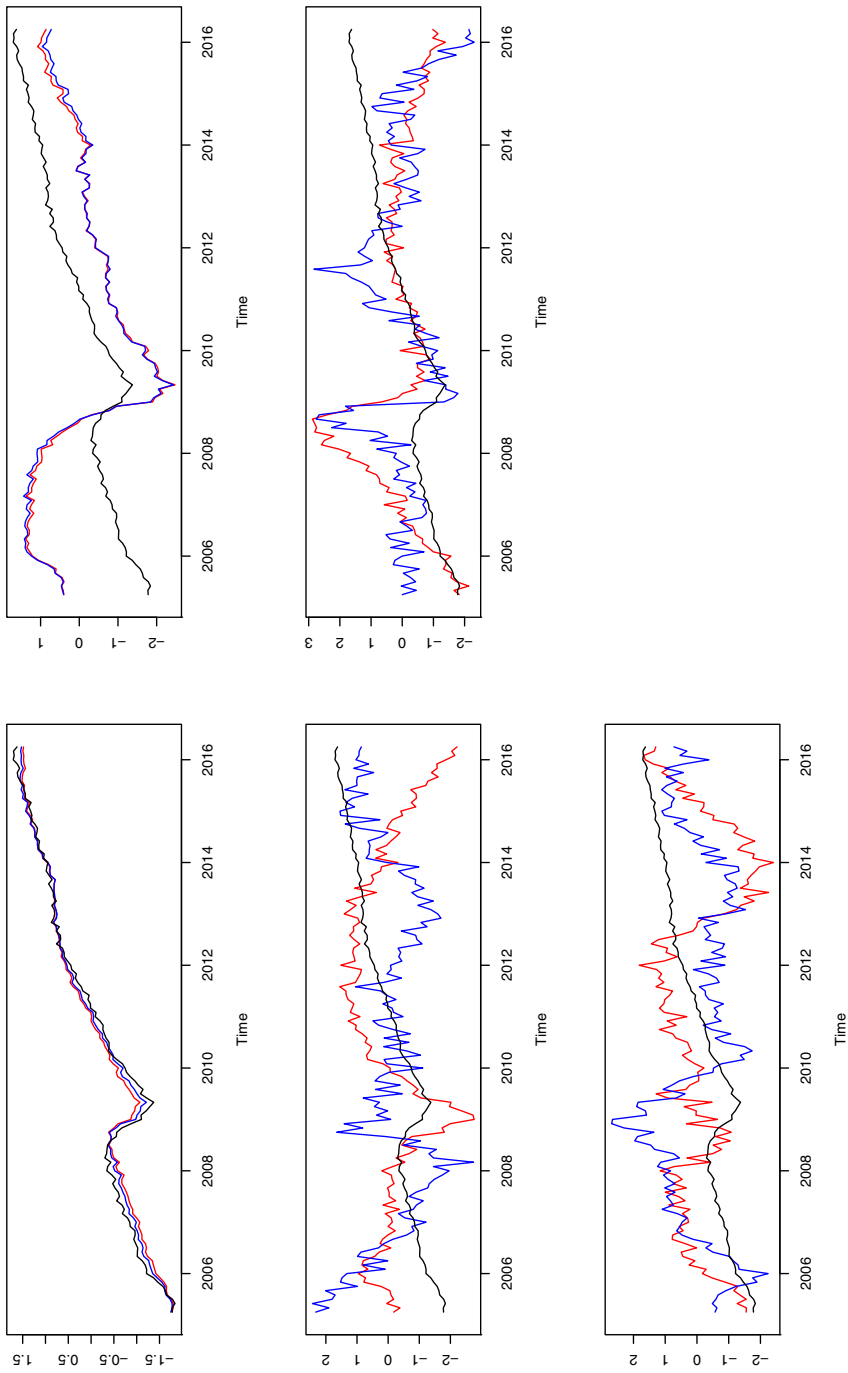
In order to evaluate if the common factors are the common trends of  $x_t$ , we carry out the cointegration exercise. The possible cointegration relationship is given by:

$$\begin{aligned} \hat{v}_t = \ln x_t &- 4.6369 - 0.0719\hat{F}_{1t} - 0.0098\hat{F}_{2t} \\ &\quad (0.0000) \quad (0.0000) \quad (0.0000) \\ &- 0.0003\hat{F}_{3t} + 0.0008\hat{F}_{4t} + 0.0002\hat{F}_{5t}. \\ &\quad (0.0000) \quad (0.0000) \quad (0.0000) \end{aligned} \tag{13}$$

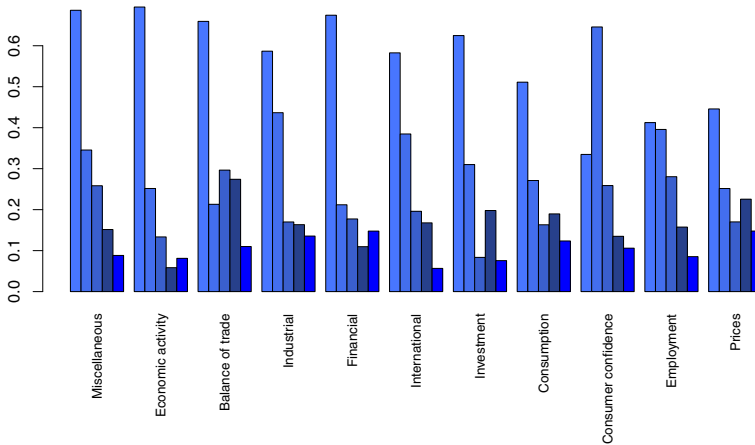
We estimate the ADF test with its respective  $p$  value (in parentheses), obtaining the following results:

$$\text{ADF test: } -3.4382 \text{ (0.01)}$$

Then, we can verify that the common trends of a large dataset of the economic variables are cointegrated with economic activity. Furthermore and following Bai



**Fig. 2** Top left panel  $\langle \ln x_t, F_{1t} \rangle$ , top right panel  $\langle \ln x_t, F_{2t} \rangle$ , middle left panel  $\langle \ln x_t, F_{3t} \rangle$ , the middle right panel  $\langle \ln x_t, F_{4t} \rangle$ , and bottom left panel  $\langle \ln x_t, F_{5t} \rangle$ . The blue color refers to PC and the red color to PLS (color figure online)



**Fig. 3** Weighted mean  $\hat{\mu}_{jg} = \sum_{i_g=1}^G |\hat{p}_{i_g jg}|/N_g$  for  $j = 1, \dots, r$ , and  $g = 1, \dots, G$ , where  $G$  is the number of blocks of variables and  $N_g$  is the number of variables in each group

and Ng (2004), first, we carry out a Panel Analysis of Non-stationarity in Idiosyncratic and Common Components (PANIC) on the idiosyncratic errors obtained using the “differencing and recumulating” method in order to disentangle the non-stationarity in this component. Second, we apply a PANIC to the idiosyncratic components estimated using data in levels as we have proposed in this study. In the first analysis, we obtain a  $p$  value of 0.1171 while in the second, we obtain a  $p$  value of 0.0000. Although in the first case  $\hat{\epsilon}_t$  is statistically non-stationary, the  $p$  value is around the uncertainty zone. Furthermore, we apply the variant of the ADF test proposed by Bai (2004) with the aim of detecting how many of the five common factors are non-stationary. As we expected, the tests show that the five common factors are non-stationary. Therefore, we conclude that the idiosyncratic terms are stationary and the common factors are non-stationary; hence, we can also argue that the elements of  $Y_t$  are cointegrated and the common factors are the common trends of  $Y_t$  and  $x_t$ .

Note that Eq. (13) is the static version of Eq. (7). The goal of this exercise is to determine whether  $F_t$  are the common trends of  $x_t$  and  $Y_t$ . We use this information to forecast the target variable with the FAVAR model presented in Sect. 2.

### 5.3 Evaluating the use of common trends to predict Mexican economic activity

It is important to have an empirical strategy to adequately predict the target variable. Consequently, with the aim of selecting the forecast model, we consider all possibilities of FAVAR models,  $\sum_{i=1}^r {}^r C_i \times 396$ , where  ${}^r C_i$  is the binomial coefficient  $\binom{r}{i}$  and 396 is obtained as the product of 11 seasonal dummies (3–12 and none), 3 deterministic specifications in the FAVAR model (none, constant and trend), and 12 lags (1–12). Therefore, the lag order, the seasonal dummies, the

deterministic component, and the factors are directly determined by minimum out-sample forecast error. The training sample covers from March 2015 to April 2016, such that we forecast 12 periods (1 year) for  $h = 2$ . This forecast period is one of relative economic and political stability in Mexico. For example, during this time frame, the annual growth rate of Mexico's GDP in any quarter was never lower than 2.3% and never higher than 2.8% (INEGI 2016). Moreover, the country's economic performance tends to be more volatile in times close to presidential elections in Mexico, and to a lesser degree in the US, which do not coincide with our forecast period. Statistically, this period represents around 10% of the number of observations and discounting the degrees of freedom, we are able to represent 25% of  $T - K$  where  $K$  is the number of parameters in the FAVAR model.

For each model, we compute the forecast error. The forecasts are dynamic, so that we update  $T + 1$  in each month. We selected the model that minimizes the Root Mean Square Error (RMSE).<sup>15</sup> Furthermore, we focus on the models that give a forecast error lower than a threshold. This threshold is determined directly using Eickmeier et al. (2014) as a reference. In their work they predict several macroeconomic variables using FAVAR models, FAVAR-tv, FAVAR-tv with stochastic volatility errors and univariate models. For one and two step ahead, forecasting the US GDP, the RMSE values in in-sample forecasts are 0.76 and 0.80 considering all periods for  $h = 1$  and  $h = 2$ , respectively. However, we select a threshold of 0.5 to obtain more accurate forecasts. Note that, if we predict  $\Delta \ln x_t$  for the first step ahead, the forecast error is  $e_{T+1} = (\ln x_{T+1} - \ln x_T) - (\ln x_{T+1}^f - \ln x_T) = \ln x_{T+1} - \ln x_{T+1}^f$ , i.e., it is equivalent to forecasting the level of the first step ahead. Therefore, we focus on the levels of the IGAE.

#### 5.4 Using common trends to predict Mexican economic activity

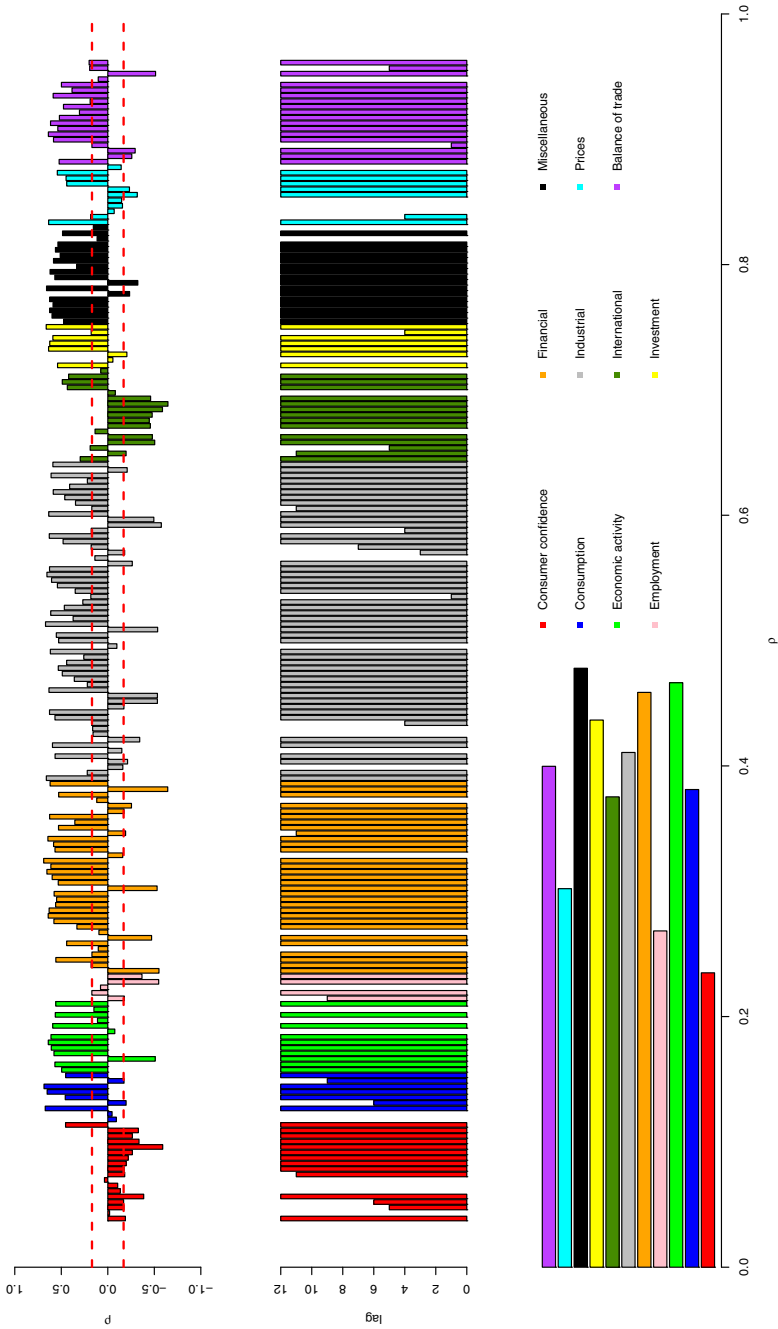
First, in order to descriptively evaluate the predictive capacity of each variable, we calculate the cross-correlation between  $y_{it-h}$  and  $x_t$  and consequently:

$$\rho_h^* = \text{corr}(y_{it-h}, x_t) | \{ \max(h \leq 12 : \text{Prob}(\text{corr}(y_{it-h}, x_t)) < \alpha) \},$$

where  $\alpha = 0.05$ . Figure 4 plots the results of the previous equation. Note that the top panel plots the  $\rho_h^*$  with the confidence interval. The middle panel shows their corresponding maximum significant lag,  $\max(h \leq 12 : \text{Prob}(\text{corr}(y_{it-h}, x_t)) < \alpha)$  and bottom panel presents the mean absolute correlations for each block of variables. Note that the block of variables most highly correlated with the future of  $x_t$  is the miscellaneous one. It is interesting to note that all significant variables of this block are positively correlated with the IGAE. Other interesting variable blocks are the economic activity and financial groups. On the other hand, the blocks least correlated with the future of  $x_t$  are consumer confidence, employment, and prices.

<sup>15</sup> As we have mentioned in the introduction, the goal of this paper is to forecast Mexican economic activity. To this end, we have two measures: quarterly, GDP; and monthly, IGAE. In an early analysis, we prove that IGAE is perfectly correlated with GDP. In this case, given the monthly frequency data, it is more convenient to forecast the IGAE, with  $x_t$  as the target variable.





**Fig. 4** The top panel plots the  $\rho_h^*$  with the confidence interval. The middle panel shows their corresponding maximum significant lag,  $\max(h \leq 12 : \text{Prob}(\text{corr}(y_{t-h}, x_t) < 0.05))$ , and the bottom panel presents the mean absolute correlations for each block of variables

It is important to state that the PANIC carried out on the idiosyncratic errors and common factors shows that the elements of  $Y_t$  are cointegrated; hence, it is reasonable to expect that the correlations between  $y_{it}$  and  $x_t$  are not spurious.

Once the models are estimated, we review the forecast errors lower than the threshold in the training sample. Figure 5 plots the historical behavior of the RMSE for the selected predictive models. The top panel plots the RMSE for  $h = 1$ . Note that the PC gives slightly more accurate results than PLS. Furthermore, the dispersion of the RMSE for PLS is larger than PLS and neither approach presents outliers. The bottom panel shows the results for  $h = 2$ . We can see that for both procedures, the forecast errors are slightly increased with respect to  $h = 1$ . It is interesting to mention that for both  $h$  the tails from both procedures are intercepted. This graph is important because this behavior is expected for the following two predicted months.

Using the models from Fig. 5 we predict two steps ahead: May and June 2016. Note that we have  $n$  models, such that it is necessary to combine the forecasts. Thus we propose a weighted average, obtaining loadings similar to the PLS procedure by solving the following optimization problem:

$$\rho_1 = \max_{w,b} \text{cor}(X^f w, xb) \quad \text{subject to} \quad \text{Var}(X^f w) = \text{Var}(xb) = 1,$$

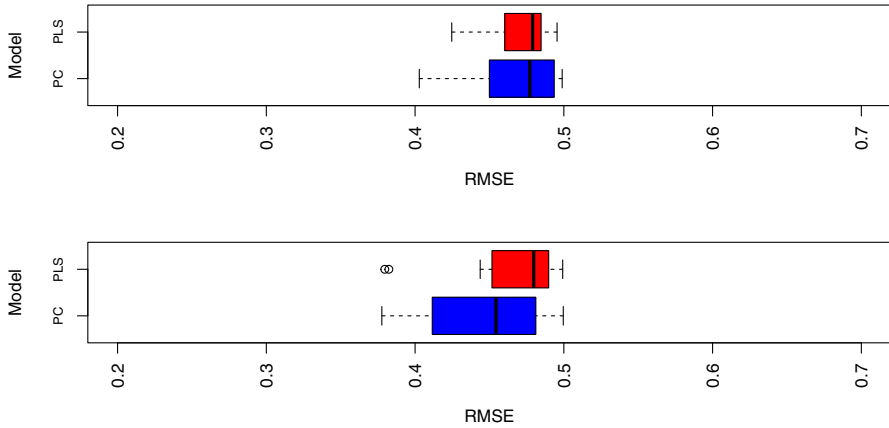
where  $w = (w_1, \dots, w_n)$ . In order to normalize the loading weights, we carry out the following scaling:  $w^* = n(\sum_{i=1}^n w_i)^{-1} w$ , such that the loading weights are between 0 and 1. Figure 6 shows the forecast density, the predictions, and the observed data. We plot the confidence interval to 95%. Note that for  $h = 1$ , PLS is more accurate than PC, while for  $h = 2$ , PLS is less accurate. The models are centered on the mean of the distribution. Moreover, the forecast density acquires the observed data. Furthermore, focusing on  $h = 2$ , the distribution of PC has two modes and the predicted data tend towards the center of the distribution. On the other hand, for PLS, it tends towards the median of the distribution.

An interesting question is: which common trends are helpful to reduce forecast error? To this end, we compute the following coefficients through OLS:

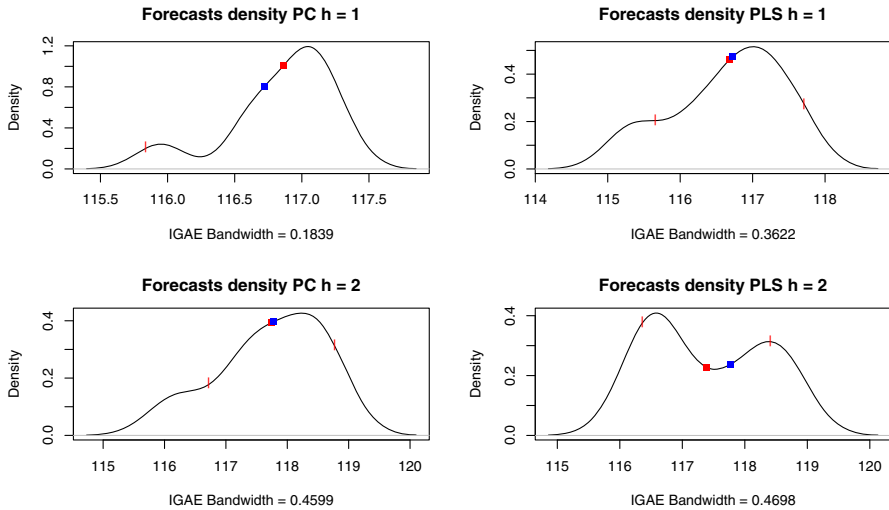
$$\hat{\partial} e_t | D_{it} = 1 < 0 \quad \text{and} \quad \text{Prob}(\hat{\partial} e_t | D_{it} = 1) < 0.10 \quad \text{for } i = 1, \dots, N_m,$$

where  $N_m$  is the number of models for each procedure. In other words, we carry out a linear regression between the forecast errors according to each procedure and dummy variables that specify the combination of common trends  $F_{1t}$ ,  $F_{2t}$ ,  $F_{3t}$ ,  $F_{4t}$ , and  $F_{5t}$  for both procedures. Figure 7 plots the result for each procedure in each  $h$ . We can see that, in PC,  $F_{1t}$  is a very important common factor to reduce the forecast error for  $h = 1$ , whereas  $F_{1t}$ ,  $F_{2t}$ ,  $F_{4t}$ , and  $F_{5t}$  are important for  $h = 2$ . Furthermore, in the PLS approach  $F_{1t}$ ,  $F_{2t}$ ,  $F_{4t}$ , and  $F_{5t}$  are relevant common factors for  $h = 1$  whereas  $F_{1t}$ , and  $F_{2t}$  are for  $h = 2$ . In fact, for both procedures the interaction of all common factors is important to reduce the forecast errors. This result completes the conclusions when we analyze Figs. 1 and 4, where all common factors are helpful to reduce the forecast errors.

In order to evaluate the forecast accuracy in “real time,” we predict two steps ahead (May and June 2016). Figure 8 shows the forecast accuracy of the following models:



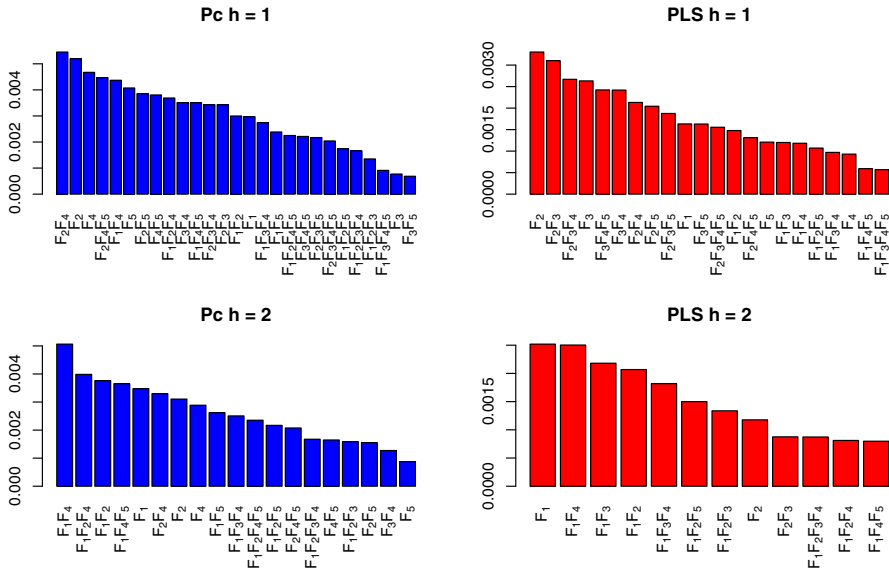
**Fig. 5** Box plots for the RMSE of out-sample forecasts for the selected models. The top panel is  $h = 1$  and the bottom panel  $h = 2$ . The blue color refers to PC and the red color to PLS (color figure online)



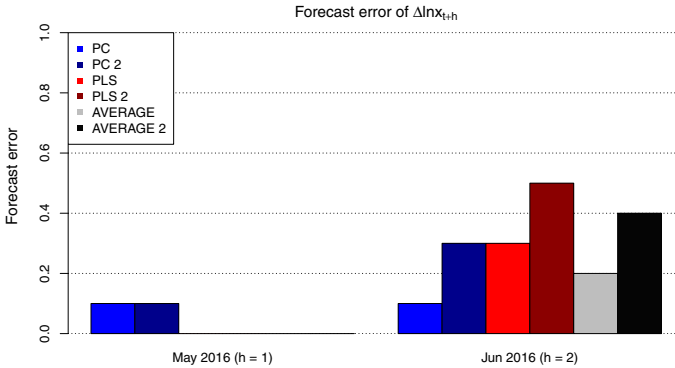
**Fig. 6** Forecast densities, forecast points (red color), and observed data (blue color) (color figure online)

(i) PC, (ii) PC using only factors that contribute to reducing the forecast error denoted as PC (2), (iii) PLS, (iv) PLS (2), (v) the average between the first and third models, and (vi) the average between the second and fourth models. We can see that for  $h = 1$  PLS, PLS (2) and the forecast average of PC and PLS are the most accurate model with 0 forecast error for May 2016. For  $h = 1$ , PC gives a forecast error of 0.1 in June 2016. In conclusion, and with reference to Fig. 5, the results are as expected given that PC and PLS give forecast errors lower than the selected threshold.

It is interesting to mention the historical behavior of the models. Hence, we first analyze the RMSE of the benchmark models: the Autoregressive Integrated Moving Average (ARIMA) and the macroeconomic diffusion index (Stock and Watson 2002b). Figure 9 plots the results for the out-sample training. For  $h = 1$ , we observe



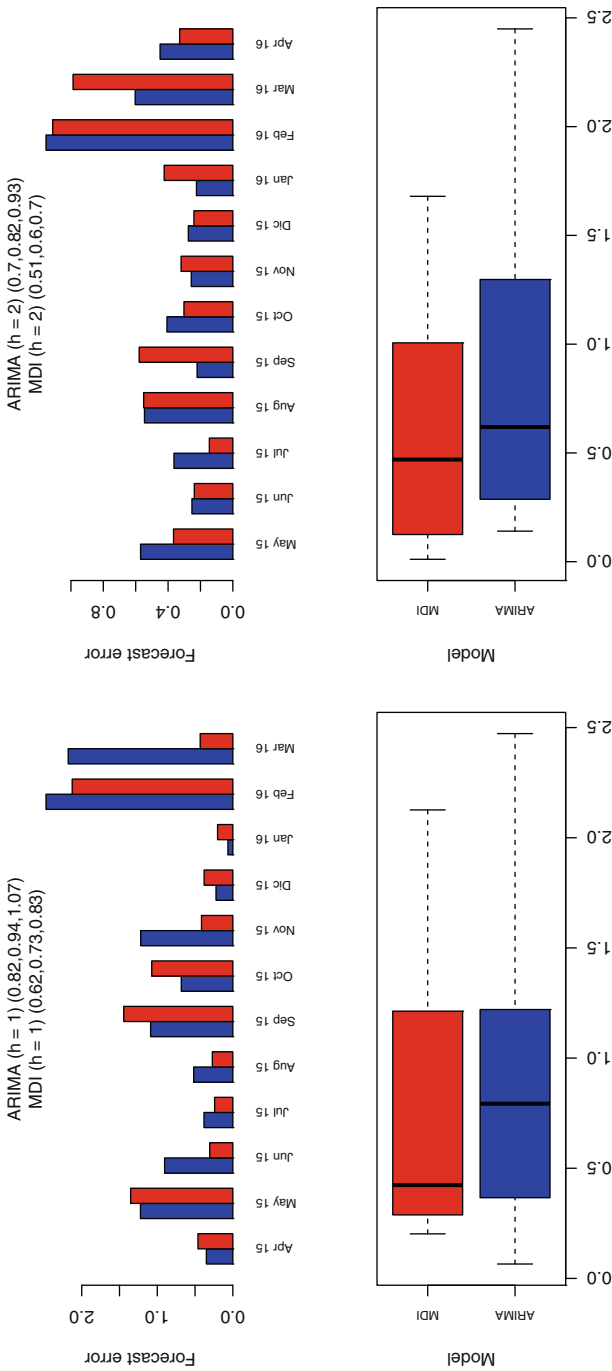
**Fig. 7** Effects of variables on forecast errors:  $\partial e_t|D_{it} = 1 < 0$  and  $\text{Prob}(\partial e_t|D_{it} = 1) < 0.10$ . The blue color refers to PC and the red color to PLS (color figure online)



**Fig. 8** Observed forecast error for May 2016 ( $h = 1$ ) and June 2016 ( $h = 2$ ). PC (blue color), PC 2 (deep blue color), PLS (red color), PC 2 (deep red color), average of forecasts (gray color), and average 2 of forecasts (black color). The number 2 indicates that the models only considered the variables that satisfy  $\partial e_t|D_{it} = 1 < 0$  and  $\text{Prob}(\partial e_t|D_{it} = 1) < 0.10$  (color figure online)

that the RMSE interval is between 0.82 and 1.07 for the ARIMA model, while for the macroeconomic diffusion index it is between 0.62 and 0.83. For  $h = 2$  the errors are slightly reduced in both cases; however, the macroeconomic diffusion index has better results. Hence, note that the inclusion of the factor in linear models reduces the forecast error. Then, we would expect the FAVAR models to show a small RMSE.<sup>16</sup>

<sup>16</sup> Alternatively, we can consider RMSE in terms of benchmark models like several authors. However, we decide to use the RMSE to easily verify whether the FAVAR models give RMSEs lower than the selected threshold of 0.5.



**Fig. 9** Forecast Errors for benchmark models: April 2015–2016. ARIMA (red color) and Macroeconomic diffusion index (blue color). The top left panel plots  $h = 1$  and the bottom right panel plots  $h = 2$  (color figure online)

Figure 10 plots step-by-step the forecast errors of the FAVAR models considering the approach presented in this work. Note that the RMSE interval of PC for  $h = 1$  is between 0.27 and 0.67 and for  $h = 2$  between 0.29 and 0.59. The RMSE mean values are 0.47 and 0.44 for each  $h$ , respectively. Note that in both  $h$ , February 2016 is the outlier forecast error. On the other hand, for PLS and  $h = 1$ , the forecast errors are between 0.34 and 0.60 and for  $h = 2$ , the confidence interval is between 0.3 and 0.64. In this case, the mean of RMSE is 0.47 for each  $h$ , respectively. Note that the improvement with respect to the ARIMA model and macroeconomic diffusion index is relevant, above all when the factors are estimated using PC. Note that the forecast errors are very similar between PC and PLS.

A question of interest is: are the models consistent through the training sample? We can obtain different models in each step ahead, in which case the consistency of the predictors may be questionable. Taking into account only the selected models, we can observe in Fig. 11 that all models are robust in  $h = 1$ . In fact, 76% of the models are within the threshold and for  $h = 2$ , the behavior of the predictions is similar to May 2016. Note that the predictions for May and June 2016 are carried out with information up to April 2016. Hence, we present the forecast in  $h = 1$  for May 2016 and  $h = 2$  for June 2016. The model is not updated as in the previous forecasts. It is interesting to note that February 2016 was the most complicated month to predict; however, the robustness of the selected models is reasonable. The reason why February 2016 was a complicated month to predict can be explained by the fact that this month has 29 days, and the seasonal variables and the dynamics between the variables do not account for this effect.

## 6 Conclusions and further research

In this paper we estimated the common trends of the Mexican economy using a large dataset of macroeconomic variables. Using cointegration concepts, we estimated the common trends using  $I(1)$  variables through PC, determining the number of common factors according to the Onatski (2010) and the Ahn and Horenstein (2013) procedures. Alternatively, we estimated the common factors using PLS.

We find that 211 macroeconomic and financial variables can be summarized in at most  $\hat{r} = 5$  common trends, which are cointegrated with Mexican economic activity. Furthermore, we use the common trends in a FAVAR model with the aim of predicting Mexican economic activity. We statistically evaluate the predictive capacity of the common trends, where we can see that each common factor reduces the forecast error. Additionally, we observe that the forecast error is reduced with respect to the ARIMA and factor-augmented regressions, such that the common trends are useful to predict Mexican economic activity.

An important conclusion is that the macroeconomic common trends can be used in more sophisticated models in order to reduce the forecast errors. Additionally, note that in this study we used data in levels in order to estimate the common factors. This is empirical evidence on the use of DFMs when we do not transform the data to stationarity.

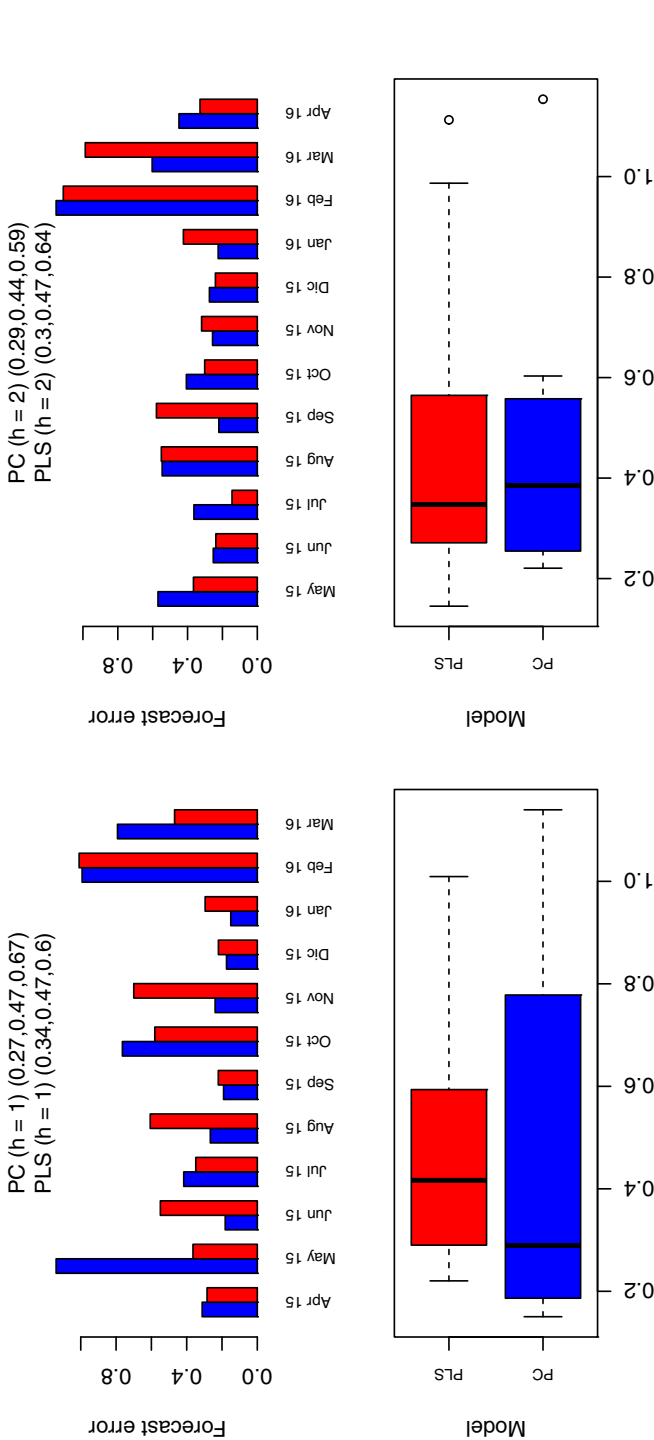
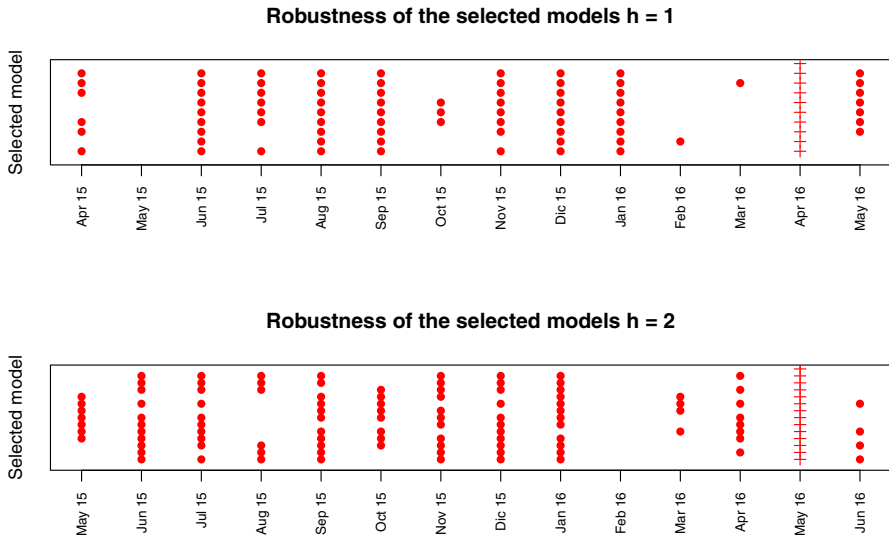


Fig. 10 Forecast Errors for FAVAR models: April 2015–2016. PC (red color) and PLS (blue color). The top left panel plots  $h = 1$  and the bottom right panel plots  $h = 2$  (color figure online)



**Fig. 11** Robustness of selected models. Red points indicate that the model satisfies the empirical threshold in the specific forecast month. A red cross indicates that there is no information to compute the statistic (color figure online)

Economically, the more relevant groups of variables in the determination of the common trends are the miscellaneous group, the economic activity, balance of trade, and the industrial sector. The first group of variables basically comprises tourism information and variables related to the automotive industry. In this context, the behaviors of the external sector and internal demand are very important to predict the behavior of the global economy. It is reasonable to expect that external uncertainty and disincentives in the internal market can be dangerous for future economic growth. In other words, these groups of variables indicate the future movements of Mexican economic activity.

A future line of research is to use the common trends in factor error correction models (FECM), and FAVAR with time-varying coefficients, among many other models, combining the predictions with different combinations of forecasting methods. Furthermore, we consider a unit root test that takes into account the structural break that occurred between 2008 and 2009 in order to refine the procedure to select  $I(1)$  variables. Consequently, it is interesting to study the non-stationarity of the estimated common factors in the presence of breaks. Additionally, we will take into account the possibility of sparse cointegration for forecasting proposes.

**Acknowledgements** Partial financial support from the CONACYT CB-2015-01-252996 is gratefully acknowledged.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



**Annex: Variables in the dynamic factor model**

#	Short	Long	Block	Log	SA	T
1	EXP TOT	Total absolute value exports	Balance of trade	Yes	Yes	1
2	EXP PET TOT	Oil exports	Balance of trade	Yes	Yes	1
3	EXP PET CRU	Crude oil exports	Balance of trade	Yes	Yes	1
4	EXP PET OTR	Other oil exports	Balance of trade	Yes	Yes	1
5	EXP NO PET	Non-oil exports	Balance of trade	Yes	Yes	1
6	EXP NO PET AGR	Farming exports	Balance of trade	Yes	Yes	1
7	EXP NO PET EXTRAC	Extractive exports	Balance of trade	Yes	Yes	1
8	EXP MAN AUT	Automotive exports	Balance of trade	Yes	Yes	1
9	EXP NO MAN RES	Manufacture exports	Balance of trade	Yes	Yes	1
10	IMP TOT PET	Total oil imports	Balance of trade	Yes	Yes	1
11	IMP CON TOT	Consumption goods imports	Balance of trade	Yes	Yes	1
12	IMP CON PET	Consumption goods oil imports	Balance of trade	Yes	No	1
13	IMP CON NO PET	Consumption goods non-oil imports	Balance of trade	Yes	Yes	1
14	IMP INT PET	Intermediate goods oil imports	Balance of trade	Yes	Yes	1
15	IMP BK	Capital goods imports	Balance of trade	Yes	Yes	1
16	REM TOT	Total remittances	Balance of trade	Yes	Yes	1
17	REM MO	Money orders remittances	Balance of trade	Yes	Yes	1
18	REM TRANS ELECT	Electronic transfers remittances	Balance of trade	Yes	Yes	1
19	REM EFEC	Cash and in-kind remittances	Balance of trade	Yes	Yes	1
20	I CONF	Consumer confidence index	Consumer confidence	Yes	Yes	1
21	I CONF COMP HOG 12	Index: compared to this household's economic situation 12 months ago, how do you think your situation is at the moment?	Consumer confidence	Yes	Yes	1
22	I CONF FUT COMP PAI 12	Index: how do you foresee this household's economic situation in 12 months' time, compared to the current situation?	Consumer confidence	Yes	Yes	2

continued

#	Short	Long	Block	Log	SA	T
23	I CONF COMP PAI 12	Index: how do you consider the country's economic situation today compared to 12 months ago?	Consumer confidence	Yes	Yes	1
24	I CONF FUT COMP PAI 12	Index: how do you foresee the country's economic situation in 12 months' time?	Consumer confidence	Yes	Yes	1
25	I CONF POS FUT	Index: comparing the current economic situation with that of a year ago, how do you consider at the present moment the possibilities that you or any of the members of this household	Consumer confidence	Yes	Yes	1
26	I CONF B	Consumer confidence index (balance)	Consumer confidence	Yes	Yes	2
27	COMP 12 B	Other complementary index: how would you describe your economic situation compared to 12 months ago?	Consumer confidence	Yes	Yes	1
28	FUT 12 B	Other complementary index: and how do you think your economic situation will be in 12 months, compared to the current situation?	Consumer confidence	Yes	Yes	1
29	POSI 12	Other complementary index: at this moment, are you more able to buy clothes, shoes, food, etc., than a year ago?	Consumer confidence	Yes	Yes	1
30	POSI 12 VAC	Other complementary index: do you consider that during the next 12 months you or any of the members of this household will afford to go on vacation?	Consumer confidence	Yes	Yes	1
31	POS AHO	Other complementary index: are you currently able to save some of your income?	Consumer confidence	Yes	Yes	1
32	COMP C 12	Other complementary index: how do you foresee your economic conditions to save in 12 months' time compared to current conditions?	Consumer confidence	Yes	Yes	1
33	COMP PRE	Other complementary index: compared with the previous 12 months, how do you think prices will behave in the country in the next 12 months?	Consumer confidence	Yes	Yes	1
34	COMP 12 EMP	Other complementary index: do you think that employment in the country in the next 12 months is going to increase, remain the same, or decrease?	Consumer confidence	Yes	Yes	1
35	AUT 12	Other supplemental index: are any members of this household or you planning to buy a new or used car in the next 2 years?	Consumer confidence	Yes	Yes	1
36	PLA AUT 2	Other complementary index: are any members of this household or you planning to buy, build or remodel a home in the next 2 years?	Consumer confidence	Yes	Yes	1

continued

#	Short	Long	Block	Log	SA	T
37	PLA CONST 2	Other complementary index: (balance) compared with the previous 12 months, how do you think prices will behave in the country in the next 12 months?	Consumer confidence	Yes	Yes	1
38	IVF	Physical volume index total	Consumption	Yes	Yes	2
39	IVFBST	IVF goods and services of national origin total	Consumption	Yes	Yes	2
40	IVFB	IVF goods and services of national origin goods	Consumption	Yes	Yes	1
41	IVFS	IVF goods and services of national origin services	Consumption	Yes	Yes	2
42	IVFBI	IVF imported goods	Consumption	Yes	Yes	1
43	IVFAT	IVF total accumulative	Consumption	Yes	Yes	1
44	IVFNB	IVFAT goods and services of national origin goods	Consumption	Yes	Yes	1
45	IVFNS	IVFAT goods and services of national origin services	Consumption	Yes	Yes	2
46	IVFBIMP	IVFAT imported goods	Consumption	Yes	Yes	1
47	IGAE A 1	IGAE primary activities	Economic activity	Yes	Yes	1
48	IGAE A 2	IGAE secondary activities total	Economic activity	Yes	Yes	1
49	IGAE A 21	IGAE secondary activities 21 mining	Economic activity	Yes	Yes	1
50	IGAE A 22	IGAE secondary activities 22	Economic activity	Yes	Yes	1
51	IGAE A 31–33	IGAE secondary activities 31–33	Economic activity	Yes	Yes	1
52	IGAE A 3	IGAE tertiary activities total	Economic activity	Yes	Yes	1
53	IGAE A 43–46	IGAE tertiary activities 43–46 trade	Economic activity	Yes	Yes	1
54	IGAE A 48 49 51	IGAE tertiary activities 48–49–51. Transport, mail, and storage; mass media information	Economic activity	Yes	Yes	2
55	IGAE A 52–53	IGAE tertiary activities 52–53 financial and insurance services; real estate and rental services of movable and intangible goods	Economic activity	Yes	Yes	1
56	IGAE A 54–56	IGAE tertiary activities 54–55–56 professional, scientific and technical services; corporate; business support services and waste management and remediation services	Economic activity	Yes	Yes	2
57	IGAE A 61–62	IGAE tertiary activities 61–62 educational services; health and social work services	Economic activity	Yes	Yes	1
58	IGAE A 71–81	IGAE tertiary activities 71–81 cultural and sporting recreation services, and other recreational services; other services except government activities	Economic activity	Yes	Yes	2

continued

#	Short	Long	Block	Log	SA	T
59	IGAE 72	IGAE tertiary activities 72 temporary accommodation and food and beverage preparation services	Economic activity	Yes	Yes	1
60	PEO	Economically active population: occupied population	Employment	Yes	Yes	1
61	PED	Economically active population: unemployed population	Employment	Yes	Yes	1
62	PO S AGR	Occupied population by agriculture economic activity sector	Employment	Yes	Yes	1
63	PO CONS	Occupied population by construction economic activity sector	Employment	Yes	Yes	1
64	PO MAN	Occupied population by manufacturing industry economic activity sector	Employment	Yes	Yes	1
65	TIIE	Interbank interest rates, bank collection costs, and CETES performance Interbank interest rate (TIIE)	Financial	Yes	Yes	1
66	CETES	Interbank interest rates, bank funding costs, and CETES performance cost of time deposits of liabilities denominated in US dollars (CCP-dollars)	Financial	Yes	Yes	1
67	TC	Exchange rate of peso against dollar and UDIS value Interbank (sale)	Financial	Yes	Yes	1
68	I TC R	Mexican peso real exchange rate index	Financial	Yes	Yes	1
69	I TC R VAR	Mexican peso real exchange rate index annual change	Financial	No	No	1
70	IPCBMV	Index of prices and quotes of the Mexican stock exchange maximum	Financial	Yes	No	1
71	TIIL	International interest rates: Libor rate	Financial	Yes	No	1
72	M1	M1 money and coins held by the public	Financial	Yes	Yes	2
73	M1 CB	M1 foreign currency checking accounts at resident banks	Financial	Yes	Yes	2
74	M1 DAP	M1 demand deposits of savings and loan companies	Financial	Yes	Yes	1
75	M2 CIRT	Internal funding of resident banks total	Financial	Yes	Yes	1
76	M2 CIBR	Domestic funding from resident banks	Financial	Yes	No	1
77	M2 CIBR ME	Internal collection of resident banks in foreign currency	Financial	Yes	Yes	1
78	M2 VP GF	Securities issued by the Federal Government total	Financial	Yes	Yes	1
79	IPAB	IPAB total	Financial	Yes	Yes	1
80	IPAB S	IPAB securities held by the siefores	Financial	Yes	Yes	1
81	IPAB EPP	IPAB securities held by public and private companies	Financial	Yes	Yes	1
82	M2 SIEFORES	Other public securities held by the siefores	Financial	Yes	Yes	1

continued

#	Short	Long	Block	Log	SA	T
83	M2 EPP	Other public securities owned by public and private companies	Financial	Yes	Yes	1
84	M2 V P T	Private securities total	Financial	Yes	Yes	1
85	M2 V P SIEFORES	Private securities held by the siefores	Financial	Yes	No	1
86	M2 V P EPP	Private securities owned by public and private companies	Financial	Yes	Yes	2
87	FON F SIEFORES	Savings funds for withdrawal from SIEFORES total	Financial	Yes	Yes	1
88	FON F SIEFORES BM	Retirement funds at the Banco de Mexico total	Financial	Yes	Yes	1
89	SIEFORES	Retirement funds at the Banco de Mexico total	Financial	Yes	Yes	1
90	M3	Domestic financial assets held by non-residents total	Financial	Yes	No	2
91	M3 BRT	Collection of resident banks total	Financial	Yes	Yes	1
92	M3 BRMN	Collection of resident banks in national currency	Financial	Yes	Yes	1
93	M3 BRME	Collection of resident banks in foreign currency	Financial	Yes	Yes	1
94	M3 VP T	Government securities held by non-residents total	Financial	Yes	No	2
95	M3 GF	Securities issued by the Federal Government total	Financial	Yes	Yes	2
96	M3 IPAB	Public securities held by non-residents securities issued by the IPAB	Financial	Yes	Yes	1
97	M4	Acquisition of branches and agencies of Mexican banks abroad total	Financial	Yes	Yes	1
98	M4 DR	Deposits of residents total	Financial	Yes	Yes	1
99	M4 DNR	Deposits of non-residents total	Financial	Yes	No	1
100	IVFI 212	Mining of metallic and non-metallic minerals, except oil and gas	Industrial	Yes	Yes	1
101	IVFI 213	Mining-related services	Industrial	Yes	Yes	1
102	IVFI 23	Construction total	Industrial	Yes	Yes	2
103	IVFI 237	Construction of civil engineering works	Industrial	Yes	Yes	2
104	IVFI 238	Specialized construction works	Industrial	Yes	Yes	1
105	IVFI 311	Manufacturing 311 industry	Industrial	Yes	Yes	2
106	IVFI 312	Manufacturing 312 beverages and tobacco industry	Industrial	Yes	Yes	1
107	IVFI 313	Manufacturing 313 manufacture of textiles and textile finishing	Industrial	Yes	Yes	1
108	IVFI 314	Manufacturing 314 manufacture of textiles, except apparel	Industrial	Yes	Yes	1
109	IVFI 315	Manufacturing 315 manufacture of clothing	Industrial	Yes	Yes	1

continued

#	Short	Long	Block	Log	SA	T
110	IVFI 316	Manufacturing 316 tanning and finishing of leather, manufacture of leather and leather substitutes materials	Industrial	Yes	Yes	1
111	IVFI 321	Manufacturing 321 wood industry	Industrial	Yes	Yes	1
112	IVFI 322	Manufacturing 322 paper industry	Industrial	Yes	Yes	1
113	IVFI 323	Manufacturing 323 printing and related industries	Industrial	Yes	Yes	1
114	IVFI 324	Manufacturing 324 manufacture of petroleum products and coal	Industrial	Yes	Yes	1
115	IVFI 325	Manufacturing industries 325 chemical industry	Industrial	Yes	Yes	1
116	IVFI 326	Manufacturing 326 Plastic and rubber industry	Industrial	Yes	Yes	1
117	IVFI 327	Manufacturing 327 manufacture of non-metallic mineral products	Industrial	Yes	Yes	1
118	IVFI 331	Manufacturing industries 331 basic metal industries	Industrial	Yes	Yes	1
119	IVFI 332	Manufacturing 332 manufacture of metal products	Industrial	Yes	Yes	1
120	IVFI 333	Manufacturing 333 manufacture of machinery and equipment	Industrial	Yes	Yes	1
121	IVFI 334	Manufacturing 334 manufacture of computer equipment, communication, measurement and other electronic equipment, components and accessories	Industrial	Yes	Yes	1
122	IVFI 335	Manufacturing 335 manufacture of electrical fittings, electrical apparatus, and electric power generation equipment	Industrial	Yes	Yes	1
123	IVFI 336	Manufacturing 336 manufacture of transport equipment	Industrial	Yes	Yes	1
124	IVFI 337	Manufacturing 337 manufacture of furniture, mattresses, and blinds	Industrial	Yes	Yes	1
125	IVFI 339	Manufacturing 339 other industrial manufacturing	Industrial	Yes	Yes	1
126	IVFIA	Accumulated total of industrial activity	Industrial	Yes	Yes	1
127	IVFIA 21	Accumulated mining total mining	Industrial	Yes	Yes	1
128	IVFIA 212	Accumulated mining of metallic and non-metallic minerals, other than oil and gas	Industrial	Yes	Yes	1
129	IVFIA 213	Accumulated mining-related services	Industrial	Yes	Yes	1
130	IVFIA 22	Accumulated total generation, transmission and distribution of electricity, water and gas supply by pipelines to the final consumer	Industrial	Yes	Yes	1
131	IVFIA 222	Accumulated Water supply and piped gas supply to the final consumer	Industrial	Yes	Yes	1
132	IVFIA 23	Accumulated construction total	Industrial	Yes	Yes	1
133	IVFIA 236	Accumulated building	Industrial	Yes	Yes	1
134	IVFIA 237	Accumulated construction of civil engineering works	Industrial	Yes	Yes	1

continued

#	Short	Long	Block	Log	SA	T
135	IVFIA 238	Accumulated specialized construction work	Industrial	Yes	Yes	1
136	IVFIA 31–33	Accumulated manufacturing industry total	Industrial	Yes	Yes	1
137	IVFIA 311	Accumulated 311 food industry	Industrial	Yes	Yes	1
138	IVFIA 312	Accumulated 312 beverages and tobacco industry	Industrial	Yes	Yes	1
139	IVFIA 313	Accumulated manufacturing industry 313	Industrial	Yes	Yes	1
140	IVFIA 314	Accumulated manufacturing industry 314	Industrial	Yes	Yes	1
141	IVFIA 315	Accumulated manufacturing industry 315	Industrial	Yes	Yes	1
142	IVFIA 316	Accumulated manufacturing industry 316 tanning and finishing of leather	Industrial	Yes	Yes	1
143	IVFIA 321	Accumulated manufacturing industry 321	Industrial	Yes	Yes	1
144	IVFIA 322	Accumulated manufacturing 322 Paper industry	Industrial	Yes	Yes	1
145	IVFIA 323	Accumulated manufacturing 323 printing and related industries	Industrial	Yes	Yes	1
146	IVFIA 324	Accumulated manufacturing 324 manufacture of petroleum products and coal	Industrial	Yes	Yes	1
147	IVFIA 325	Accumulated manufacturing industries 325 chemical industry	Industrial	Yes	Yes	1
148	IVFIA 326	Accumulated manufacturing 326 plastic and rubber industry	Industrial	Yes	Yes	1
149	IVFIA 327	Accumulated manufacturing 327 manufacture of non-metallic mineral products	Industrial	Yes	Yes	1
150	IVFIA 331	Accumulated manufacturing 331 basic metal industries	Industrial	Yes	Yes	1
151	IVFIA 332	Accumulated manufacturing 332 manufacture of metal products	Industrial	Yes	Yes	1
152	IVFIA 333	Accumulated manufacturing 333 manufacture of machinery and equipment	Industrial	Yes	Yes	1
153	IVFIA 334	Accumulated manufacturing 334 manufacture of computer, communication, measurement and other electronic equipment, components, and accessories	Industrial	Yes	Yes	1
154	IVFIA 335	Accumulated manufacturing 335 manufacture of accessories, electrical apparatus, and electric power generation equipment	Industrial	Yes	Yes	1
155	IVFIA 336	Accumulated manufacturing 336 manufacture of transport equipment	Industrial	Yes	Yes	1
156	IVFIA 337	Accumulated manufacturing 337 manufacture of furniture, mattresses, and blinds	Industrial	Yes	Yes	1
157	IVFIA 339	Accumulated manufacturing 339 other manufacturing	Industrial	Yes	Yes	1
158	IPI EUA	Indices of United States industrial production	International	Yes	Yes	1
159	TD EUA	United States unemployment rates	International	Yes	No	2
160	OM EUA	United States monetary offer	International	Yes	No	2
161	TI 3M	Interest rate USA 3 months	International	Yes	Yes	1

continued

#	Short	Long	Block	Log	SA	T
162	TI 6M	Interest rate USA 6 months	International	Yes	Yes	1
163	TI 1M	Interest rate USA 1 year	International	No	Yes	2
164	TI 2A	Interest rate USA 2 years	International	Yes	No	1
165	TI 3A	Interest rate USA 3 years	International	Yes	No	1
166	TI 5A	Interest rate USA 5 years	International	Yes	No	1
167	TI 10A	Interest rate USA 10 years	International	Yes	No	1
168	TI 20	Interest rate USA 20 years	International	Yes	No	1
169	TI FF	Interest rate USA federal funds	International	Yes	No	1
170	I REMUN EUA	Manufacturing remunerations USA	International	Yes	Yes	2
171	RESER EUA	Total reserves (not including gold) USA	International	Yes	No	1
172	EXP EUA	Total exports USA	International	Yes	No	1
173	IMP EUA	Total imports USA	International	Yes	No	1
174	BC EUA	Balance of trade USA	International	No	No	1
175	INV T	Investment total	Investment	Yes	Yes	1
176	INV CON	Investment residential	Investment	Yes	Yes	2
177	INV CON NR	Investment non-residential	Investment	Yes	Yes	2
178	INV MET	Investment total machines and equipment	Investment	Yes	Yes	1
179	INV ME	Investment machines and equipment	Investment	Yes	Yes	1
180	INV MEO	Investment machines, equipment, and others	Investment	Yes	Yes	1
181	INV MET	Investment transporting equipment (imported)	Investment	Yes	Yes	1
182	INV MAEO	Investment machines, equipment and others (imported)	Investment	Yes	Yes	1
183	TUR VS	Total balance	Miscellaneous	Yes	Yes	1
184	TUR IT	Total income	Miscellaneous	Yes	Yes	1
185	TUR TI	International tourists	Miscellaneous	Yes	Yes	1
186	TUR ET	Total expenditure	Miscellaneous	Yes	Yes	1
187	TUR TIT	Total expenditure (international tourists)	Miscellaneous	Yes	Yes	1
188	TUR VIT	Total income (volume)	Miscellaneous	Yes	Yes	1
189	TUR ITIT	Total income (international tourists)	Miscellaneous	Yes	Yes	1
190	TUR ET V	Total expenditure (volume)	Miscellaneous	Yes	Yes	1
191	TUR ET TUT	Total expenditure (international tourists)	Miscellaneous	Yes	Yes	1
192	TUR GM IT	Expenditure average (income total)	Miscellaneous	Yes	Yes	1
193	TUR GM I TUT	Expenditure average international tourists (income total)	Miscellaneous	Yes	Yes	1
194	TUR GM ET	Expenditure average (expenditure total)	Miscellaneous	Yes	Yes	1
195	TUR GM E TIT	Expenditure average international tourists (expenditure total)	Miscellaneous	Yes	Yes	1
196	AUT	Production total automotive	Miscellaneous	Yes	Yes	1
197	AUT CAM T	Sales to the public. Total automotive	Miscellaneous	Yes	Yes	1
198	AUT CAM TT	Sales to the public. Total buses	Miscellaneous	Yes	Yes	1



continued

#	Short	Long	Block	Log	SA	$T$
199	AUT CAM TN	Sales to the public. Total buses (national)	Miscellaneous	Yes	Yes	1
200	AUT CAM TI	Sales to the public. Total buses (imported)	Miscellaneous	Yes	Yes	2
201	INPC	National index of consumer prices	Prices	Yes	Yes	1
202	INPC 4	INPC clothing, footwear, and accessories	Prices	Yes	Yes	2
203	INPC 5	INPC furniture, household appliances, and accessories	Prices	Yes	Yes	2
204	INPC 5	INPC health and personal care	Prices	Yes	Yes	2
205	INPC 6	INPC transportation	Prices	Yes	Yes	2
206	INPC S	INPC underlying	Prices	Yes	Yes	2
207	INPC SM	INPC underlying goods	Prices	Yes	Yes	2
208	INPP 21	National index of producer prices (INPP) mining without oil	Prices	Yes	No	1
209	INPP 22	INPP generation, transmission and distribution of electricity, water and gas supply by pipelines to the final consumer	Prices	Yes	Yes	1
210	INPP 23	INPP construction	Prices	Yes	No	1
211	INPP 48-49	INPP transportation, post and storage	Prices	Yes	Yes	2

Log denotes logarithmic transformation, SA means seasonal adjustment, and  $T$  if 1 denotes that  $y_t \sim I(1)$  and 2 that  $\Delta y_t \sim I(1)$

## References

- Ahn S, Horenstein A (2013) Eigenvalue ratio test for the number of factors. *Econometrica* 81(3):1203–1227
- Alessi L, Barigozzi M, Capasso M (2010) Improved penalization for determining the number of factors in approximate factor models. *Stat Probab Lett* 80(1):1806–1813
- Bai J (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71(1):135–171
- Bai J (2004) Estimating cross-section common stochastic trends in nonstationary panel data. *J Econom* 122(1):137–183
- Bai J, Ng S (2002) Determining the number of factors in approximate factor models. *Econometrica* 70(1):191–221
- Bai J, Ng S (2004) A PANIC attack on unit roots and cointegration. *Econometrica* 72(4):1127–1177
- Bai J, Ng S (2007) Determining the number of primitive shocks in factor models. *J Bus Econ Stat* 25(1):52–60
- Bai J, Ng S (2008) Large dimensional factor analysis. *Found Trends Econom* 3(2):89–163
- Banerjee A, Marcellino M, Masten I (2013) Structural factor error correction models: cointegration in large-scale structural FAVAR models. Bocconi University and CEPR, Working Paper
- Banerjee A, Marcellino M, Masten I (2014) Forecasting with factor-augmented error correction models. *Int J Forecast* 30(3):589–612
- Barigozzi M, Lippi M, Luciani M (2016) Non-stationary dynamic factor models for large datasets. *Finance and Economics Discussion Series* 024
- Boivin J, Ng S (2006) Are more data always better for factor analysis? *J Econom* 132(1):169–194
- Bräuning F, Koopman SJ (2014) Forecasting macroeconomic variables using collapsed dynamic factor analysis. *Int J Forecast* 30(3):572–584
- Breitung J, Choi I (2013) Factor models. In: Hashimzade N, Thorthon MA (eds) *Handbook of research methods and applications in empirical macroeconomics*. Edward Elgar, Cheltenham (UK)

- Caruso A (2015) Nowcasting Mexican GDP. ECARES Working Paper 2015-40
- Chamberlain G, Rothschild M (1983) Arbitrage factor structure, and mean-variance analysis of large asset markets. *Econometrica* 51(1):1281–1304
- Corona F, Poncela P, Ruiz E (2017a) Determining the number of factors after stationary univariate transformations. *Empir Econ* 53:351–372
- Corona F, Poncela P, Ruiz E (2017b) Estimating non-stationary common factors: implications for risk sharing. UC3M Working Papers Statistics and Econometrics 17(09)
- Dougherty S (2015) Boosting growth and reducing informality in Mexico. OECD Economics Department Working Papers 1188
- Duy TA, Thoma MA (1998) Modeling and forecasting cointegrated variables: some practical experience. *J Econ Bus* 50(1):291–307
- Eickmeier S, Ziegler C (2008) How good are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *J Forecast* 27(3):237–265
- Eickmeier S, Lemke W, Marcellino M (2014) Classical time-varying FAVAR models—estimation, forecasting and structural analysis. *J R Stat Soc* 178(3):493–533
- Engle RF, Granger CWJ (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55(1):251–276
- Ergemen YE, Rodriguez-Caballero CV (2016) A dynamic multi-level factor model with long-range dependence. CREATES Research Paper 23
- Forni M, Hallin M, Lippi M, Reichlin L (2000) The generalized dynamic-factor model: identification and estimation. *Rev Econ Stat* 82(4):540–554
- Fuentes J, Poncela P, Rodriguez J (2015) Sparse partial least squares in time series for macroeconomic forecasting. *J Appl Econom* 30(4):576–595
- Geweke J (1977) The dynamic factor analysis of economic time series. In: Aigner DJ, Goldberger AS (eds) *Latent variables in socio-economic models*. North-Holland, Amsterdam
- Giannone D, Reichlin L, Small D (2008) Nowcasting: the real-time informational content of macroeconomic data. *J Monet Econ* 55(4):665–676
- Gonzalo J, Granger CWJ (1995) Estimation of common long-memory components in cointegrated systems. *J Bus Econ Stat* 13(1):27–35
- Hallin M, Liska R (2007) Determining the number of factors in the general dynamic factor model. *J Am Stat Assoc* 102:603–617
- Hanson G (2010) Why isn't Mexico rich? *J Econ Lit* 48(4):987–1004
- Hindrayanto I, Koopman SJ, de Winter J (2016) Forecasting and nowcasting economic growth in the euro area using factor models. *Int J Forecast* 32(4):1284–1305
- INEGI (2016) Producto interno bruto de México durante el tercer trimestre de 2016. [http://www.inegi.org.mx/saladeprensa/boletines/2016/pib\\_pconst/pib\\_pconst2016\\_11.pdf](http://www.inegi.org.mx/saladeprensa/boletines/2016/pib_pconst/pib_pconst2016_11.pdf)
- Johansen S (1988) Statistical analysis of cointegration vectors. *J Econ Dyn Control* 12(1):231–254
- Johansen S (1991) Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica* 59(1):1551–1580
- Kapetanios G (2010) A testing procedure for determining the number of factors in approximate factor models with large datasets. *J Bus Econ Stat* 28(3):397–409
- Kehoe T, Ruhl K (2010) Why have economic reforms in Mexico not generated growth? *J Econ Lit* 48(4):1005–1017
- Kunst R, Neusser K (1997) Cointegration in a macroeconomic system. *J Appl Econ* 5(1):351–365
- Lahiri K, Monokroussos G, Yongchen Z (2016) Forecasting consumption: the role of consumer confidence in real time with many predictors. *J Appl Econ* 31(7):1254–1275
- Lütkepohl H (2006) *New introduction to multiple time series analysis*. Springer, New York
- Marcellino M, Stock JH, Watson MW (2003) Macroeconomic forecasting in the euro area: country specific versus euro wide information. *Eur Econ Rev* 47(1):1–18
- Onatski A (2010) Determining the number of factors from empirical distribution of eigenvalues. *Rev Econ Stat* 92(4):1004–1016
- Panopoulou E, Vrontos S (2015) Hedge fund return predictability: to combine forecast or combine information? *J Bank Financ* 56:103–122
- Peña D, Poncela P (2004) Forecasting with nonstationary dynamic factor models. *J Econom* 119(1):291–321
- Reijer AHJ (2005) Forecasting Dutch GDP using large scale factor models. Netherlands Central Bank Working Paper

- Rodriguez-Pose A, Villarreal E (2015) Innovation and regional growth in Mexico: 2000–2010. *Growth Change* 46(2):172–195
- Sargent TJ, Sims CA (1977) Business cycle modeling without pretending to have too much a priori economic theory. In: Sims CA (ed) *New methods in business cycle research*. Federal Reserve Bank of Minneapolis, Minneapolis
- Schumacher C (2007) Forecasting German GDP using alternative factor models based on large datasets. *J Forecast* 26(4):1167–1179
- Schwellnus C (2011) Macroeconomic and structural policies to further stabilise the Mexican economy. OECD Economics Department Working Papers 906
- Stock JH, Watson MW (1988) Testing for common trends. *J Am Stat Assoc* 83(1):1097–1107
- Stock JH, Watson MW (2002a) Forecasting using principal components from a large number of predictors. *J Am Stat Assoc* 97(1):1169–1179
- Stock JH, Watson MW (2002b) Macroeconomic forecasting using diffusion indexes. *J Bus Econ Stat* 20(1):147–163
- Stock JH, Watson MW (2005) Implications of dynamic factor models for VAR analysis. NBER Working Paper 11467
- Stock JH, Watson MW (2011) Dynamic factor models. In: Clements MP, Hendry DF (eds) *Oxford handbook of economic forecasting*. Oxford University Press, Oxford
- Vahid F, Engle RF (1993) Common trends and common cycles. *J Appl Econom* 8(1):341–360
- Wilms I, Croux C (2016) Forecasting using sparse cointegration. *Int J Forecast* 32(4):1256–1267
- World Bank (2016a) Mexico overview. World Bank, Washington, DC. <http://www.worldbank.org/en/country/mexico/overview#3>
- World Bank (2016b) World development indicators. World Bank, Washington, DC. <http://data.worldbank.org/data-catalog/world-development-indicators>