

Measuring foreign impact: leading index construction using hierarchical dynamic factor model

Agne Reklaitė

To cite this article: Agne Reklaitė (2016) Measuring foreign impact: leading index construction using hierarchical dynamic factor model, *Baltic Journal of Economics*, 16:1, 21-32, DOI: 10.1080/1406099X.2016.1164976

To link to this article: <http://dx.doi.org/10.1080/1406099X.2016.1164976>



© 2016 The author(s). Published by Taylor & Francis



Published online: 24 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 329



View related articles [↗](#)



View Crossmark data [↗](#)

Measuring foreign impact: leading index construction using hierarchical dynamic factor model

Agne Reklaite

Department of Econometric Analysis, Vilnius University, Vilnius, Lithuania

ABSTRACT

In this paper a new method of constructing the leading economic index is presented. Its main advantage is the ability to distinguish domestic and foreign factors influencing the growth of economy and it is performed via dynamic hierarchical factor modelling. An application is carried out with Lithuanian data and the results indicate that foreign component corresponds to an economically and statistically significant amount of variance. Under this new methodology, a hypothesis that the effect of international trends on the growth of economy is increasing over time is validated. Results indicate that globalization effect can be quantified and monitored using the proposed decomposition.

ARTICLE HISTORY

Received 9 February 2015
Accepted 8 March 2016

KEYWORDS

Coincident and leading indicators; hierarchical dynamic factor model; small open economy; globalization

JEL CLASSIFICATION

C43; E32; C10

1. Introduction

The role of globalization is frequently noticed in various topics of economics. Recently, it is increasingly addressed as the underlying cause of diminishing accuracy of traditional domestically oriented macro-econometric models. An example of extended Conference Board methods (Drechsel and Scheufele, 2010) shows that more and more indicators have to be included into leading index construction to keep up with the accuracy of previously constructed models. This result could indicate that processes are becoming of more complicated structure impelled by increasing amount of information available for a single agent of economy and therefore affecting its decision-making. The accuracy of domestically oriented models deteriorates with time and this phenomenon is addressed by Fichtner, Rueffer, and Schnatz (2009). They find that it is caused by globalization, hence adding information about external environment improves the forecast performance.

The exploration of coincident and leading indicators began with works of Burns and Mitchell (1946), the designated indicators were later combined into composite indexes by NBER (*National Bureau of Economic Research*) economists Shiskin and Moore (1968). Their method was a groundwork for classical methods developed by the US Department of Commerce, which are based on weighted summing the growth rates of selected series of the leading indicators, and variable selection relies heavily on economic insight (The Conference Board, 2001). Many authors developed this approach and the underlying

CONTACT Agne Reklaite  agne.reklaite@mif.vu.lt; agne.reklaite@gmail.com

© 2016 The author(s). Published by Taylor & Francis

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

idea of combining indicators as a weighted sum is still in application (Auerbach, 1982; Issler and Vahid, 2003; OECD, 2012).

Stock and Watson (1989, 1991) offered a new framework for constructing coincident and leading indexes by applying a dynamic factor model. They stated that the coincident index measures the unobserved state of economy capturing co-movements of different sectors of economy, opposite to GDP which measures overall economic activity across all sectors. This method in time gained wide acceptance and was augmented in several ways: performing evaluation in frequency domain (Forni, Hallin, Lippi, and Reichlin, 2000), adding Markov-switching element (Chauvet, 1998; Diebold and Rudebusch, 1996; Kim, 1994; Kim and Nelson, 1998), moving to mixed-frequency models (Mariano and Murasawa, 2002). This among other methodologies is mostly targeted at large economies whose main drivers are embedded within economy itself. Applying it for a small open economy could provide erratic results in spite of the great influence of foreign economies, therefore some adjustments to it are highly preferred.

Applications of Stock–Watson framework on smaller economies share the inclusion of supranational indicators to represent international spill-over effects (Bandholz, 2005 for Poland and Hungary, Mapa and Simbulan, 2014 for Philippines, Schulz, 2007 for Estonia). The findings of Cubadda, Guardabascio, and Hecq (2013) show that a common factor explains a lot of co-movements of different European countries. Therefore including data of other countries could help acquire better accuracy in evaluating models.

The findings of mentioned authors suggest that the component of foreign information in economic models is gaining more importance. Statistical explanation for this could be that the foreign component of these processes was always present but was discarded as insignificant because of its noise-like features. However, due to globalization indicators from different economies are becoming more similar and supranational element is becoming more apparent. This effect should be particularly visible for small open economies. Inspecting the foreign effect on the growth of the Coincident Economic Index (CEI) is more informative than doing it on growth of GDP because it reflects effects visible across many areas of economic activity since CEI is designed to capture co-movements of different economic indicators.

The main goal of this paper is to develop a method to quantify the impact of domestic and foreign variables on the future growth of focal economy as measured by the leading economic index, i.e. the predicted growth of the coincident index. In addition to our main goal, we are adding a practical task for application of this method: inspecting the trends of the impact of information of different origins – domestic and foreign. This brings us to the main hypothesis in this study: the effect of international trends on the growth of economy is increasing over time.

The idea that the leading economic index is constructed as a forecast of the CEI growth (Stock and Watson, 1989) was embraced and the method of forecasting was linear prediction on diffusion indexes by Stock and Watson (2002). This method was selected to construct the leading economic index and to validate the hypothesis of the practical task. The choice of the framework was motivated by its capacity to incorporate various predictors into the forecast in a simple and parsimonious way and because it is widely recognizable among economists. Another advantage of this method is the ability to extract a signal from multiple time series which reflects commonalities and therefore could indicate the major trends in the information available for the subjects in the economy. The CEI

growth was the forecast indicator and 2 factors – domestic and foreign – were used as predictors. The factors were evaluated by combining the selected indicators from domestic and supranational data in a structural way and building a dynamic hierarchical factor model following Moench, Ng, and Potter (2009).

The rest of this paper is organized as follows: first, the coincident index is evaluated. In the second section, the leading indicators are selected and dynamic hierarchical factor model is built. Afterwards the evaluated domestic and foreign leading indices are combined and hypothesis is validated. Finally, the conclusions are presented.

2. The coincident index

This section briefly presents how coincident index was constructed; the procedure replicated the previous paper by Reklaitė (2011) only the more recent data are used.

2.1. A single factor model

The CEI was evaluated using a single factor model applying Stock and Watson methodology (1989; 1991) and following the example by Gaudreault, Lamy, and Liu (2003).

$$\Delta X_t = \beta + \gamma(L)\Delta F_t + \mu_t, \quad (1)$$

$$D(L)\mu_t = \varepsilon_t, \quad (2)$$

$$\phi(L)\Delta F_t = \delta + \eta_t, \quad (3)$$

$$\Delta C_t = a + b\Delta F_t. \quad (4)$$

Here X is a vector of coincident variables: IM – turnover of manufacturing, RE – real estate price index, WT – turnover index of wholesale trade and IP – index of production. F_t is a factor, describing the unobserved state of economy at time t . The functions $\phi(L)$, $\gamma(L)$ and $D(L)$ are, respectively, scalar, vector and matrix lag polynomials. The error term μ_t is serially correlated and its dynamics are described in Equation (2). C_t is CEI. Error terms (ε_t , η_t) are assumed to be i.i.d. (0 , Σ), where Σ is diagonal. a and b are the de-normalization parameters.

The variables used for evaluation of this system for Lithuanian economy were selected following Reklaitė (2011) using The Conference Board (2001) recommendations and including a variable about real estate since it reflects general economic expectations and gives a big boost in accuracy – it helps to explain 2008–2009 crisis.

These series are quarterly seasonally adjusted ¹ data² covering period from 1998 1st quarter to 2013 3rd quarter. Since RE series started at the 4th quarter of the year 1998, the values of first 3 quarters were extrapolated backwards using Holt–Winters procedure. The initial data analysis showed that these four series are $I(1)$ processes, but they are not cointegrated.³

2.2. The coincident index evaluation

The evaluation is performed following Gaudreault et al. (2003) by differencing seasonally adjusted coincident series and normalizing them. Equations (1)–(3) form a state-space

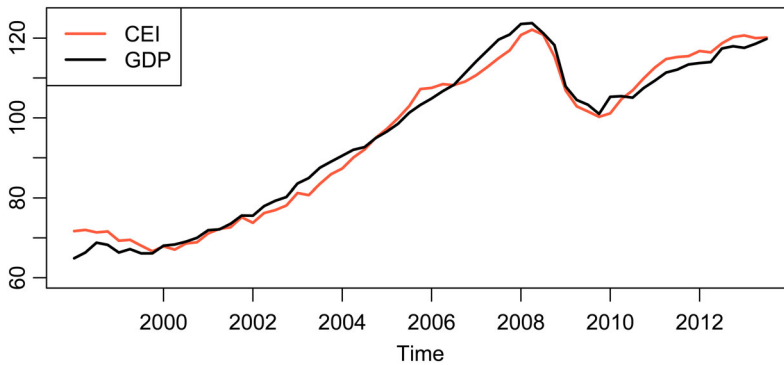


Figure 1. A comparison of evaluated CEI to gross domestic product (GDP).

model which is evaluated using maximum likelihood method and Kalman filter is used to extract the evaluated factor ΔF_t . Parameters a and b were evaluated by minimizing sum of squares: $\sum_{i=t}^T (C_t - GDP_t)^2$ (following Reklaitė, 2011). ΔF_t is going to be de-normalized $\Delta C_t = a + b\Delta F_t$ as defined by Equation (4) and CEI C_t is constructed:

$$C_t = \begin{cases} c, & t = 0; \\ c + \sum_{i=1}^t \Delta C_i, & t = 1, 2, \dots, T. \end{cases} \quad (5)$$

The expression in Equation (5) is evaluated using scaling constraint: CEI in 2005 is set to 100. The result is plotted with scaled GDP in Figure 1. It can be indicated from the graph that CEI reflects the state of economy in a very similar way as GDP (classical measure of economic activity).

3. The leading index

3.1. The leading indicators

According to Stock and Watson (1989) methodology, the leading index is constructed as a forecast of the CEI growth and it is usually performed as a separate task after having evaluated the CEI. They use the leading indicators as predictors to build the leading economic index. In this paper we are considering a much larger number of potential predictors therefore linear regression would not be feasible since there would be too many parameters to evaluate. Our intent is to use the linear forecast method by Stock and Watson (2002) which was originally developed for macroeconomic forecasting using diffusion indexes. This way we are going to use factors acquired from leading indicators rather than indicators themselves. The constructed prediction equation is of the form of Equation (6).

$$\Delta C_{t+2} = \alpha_1(L)G_{1,t} + \alpha_2(L)G_{2,t} + \beta(L)\Delta C_t + \varepsilon_t. \quad (6)$$

Here ΔC_{t+2} is future growth of CEI, $G_{1,t}$ and $G_{2,t}$ are factors acquired from domestic and foreign indicators, $\alpha_1(L)$, $\alpha_2(L)$, $\beta(L)$ are lag polynomials.

The initial domestic data set consisted of 283 time series of most Lithuanian quarterly economic indicators starting at least at 1998 (from the sectors of manufacturing and

production, labour, investment, international trade, retailing, public sector, business statistics, construction, transportation and agriculture). The initial supranational data set consisted of 1707 time series which geographically covered Lithuania's top 20 international trade partners,⁴ groups of countries such as EU, OECD, Euro area and a few largest economies on account that they might have influence to Lithuania through their global presence, such as USA and Japan. The economic indicators were from areas of national accounts, labour statistics, real effective exchange rate, saving and lending. The series were used in real terms where applicable, they were also seasonally adjusted⁵ and transformed to be stationary.

In order to achieve a straightforward interpretation, we are aiming for 1 domestic leading factor and 1 foreign leading factor. Therefore, it is important to use the time series that carry the most information about future growth of economy. Bai and Ng (2008) showed that using targeted predictors, i.e. selected subset from initial data set, gives better forecasting accuracy with the same number of factors than using the factors extracted from full data set. For this reason, we apply the leading indicators selection procedure. It is noteworthy that the selection is based on statistical properties of indicators therefore it slightly deviates from the leading indicator definition as used in OECD (2012) methodology; our definition is less restrictive.

The first stage of selecting the leading series was based on two criteria:

- (1) Granger causality (pairwise testing for lag depth 2 with significance level $\alpha = 0.05$).
- (2) Correlation between series $\Delta X_{i,(t-l)}$ and coincident index ΔC_t should be greater with lags $l > 0$.

Only the series that met both criteria were included into the following stages of modelling. After the first selection stage was completed the data set which consisted of 4 domestic and 16 foreign indicators included several collinear time series, e.g. 6 time series of labour productivity in different European countries and the EU were selected. Even though the collinearity does not cause technical problems for factor model evaluation, it can cause a certain imbalance since the factor might hinge to the series that have multiple collinear counterparts.

Hierarchical clustering was applied in order to identify the groups of indicators that are collinear. Afterwards the 'soft-thresholding' method was applied (Bai and Ng, 2008). The indicators from largest cluster were included in least angle regression (Efron, Hastie, Johnston, and Tibshirani, 2004) where the predicted variable was future growth of CEI and ranked according to their predictive power. Next, the least informative indicators were discarded so that the largest cluster diminishes to the size of second-largest cluster. More details on variable selection are provided in Appendix 2.

The finalized leading indicators data set was composed of a domestic block which consisted of 4 time series and the foreign block which was formed from 12 series. The number of series constituting the foreign data block is larger in spite of much bigger initial data pool.

The selected indicator set (the full list is given in Appendix 1) includes Lithuania's profitable share of enterprises, which was the leading indicator from the domestic leading model (Reklaite, 2011) which reflects dynamics in customer purchasing power, labour productivity and efficiency in management. Foreign direct investment to Lithuania is among

selected indicators mostly due to direct causal relationship between investment and future growth of economy; livestock and poultry represent the potential output in the agricultural sector, therefore its presence among selected indicators reveals the importance of agriculture to Lithuanian economy. Lithuania's investment abroad does not have the direct effect on the growth of the economy but it might be a good proxy indicator for business confidence and interest rates.⁶ The foreign block included several indicators of consumer and business confidence and a few indicators of labour productivity from European countries, a couple of indicators of GDP components from Portugal, Japan and France. The rest of selected leading indicators are net saving of USA and gross saving of Cyprus. These indicators reflect fluctuations in financial market: USA was selected with regard to its size and enormous impact on international financial sector while Cyprus was selected due to its large offshore banking industry (relative to GDP) and sensitivity to shocks in the finance sector. These results suggest that it might be useful to consider including more financial indicators to initial data set.

3.2. The hierarchical factor model

The method for evaluating the factors is a three-level dynamic hierarchical factor model. This method allows to impose a certain structure and estimate separate factors for domestic and foreign variables. The equations constituting the three-level hierarchical model are the following (one equation for each hierarchy level):

$$X_{bit} = \Lambda_{G,bi}G_{bt} + e_{Xbit}, \quad (7)$$

$$G_{bt} = \Lambda_{F,b}F_t + e_{Gbt}, \quad (8)$$

$$\psi_F(L)F_t = \varepsilon_{Ft}, \quad (9)$$

X_{bit} are leading series, which were transformed to be stationary and scaled (with zero mean and unit variance), index b denotes the block (either domestic or foreign), i – index of time series, t denotes time index. Λ_G and Λ_F are loadings, G_{bt} are block-level factors, F_t is a common factor. Equation (9) describes stationary AR(1) process.⁷ e_{Xbit} , e_{Gbt} and ε_{Ft} have zero mean and their variances $\Sigma_X = cov(e_{Xbit})$ and $\Sigma_G = cov(e_{Gbt})$ are assumed to be diagonal.

Since the likelihood function of this model is too complicated for consistent evaluation via the maximum likelihood method the Bayesian approach was used. The evaluation of this model was carried out following the procedure by Moench et al. (2009), via Markov Chain Monte Carlo (MCMC) using the Gibbs sampling technique (Carter and Kohn, 1994), under assumption of Gaussian innovations.

Data series are structured into blocks $b = 1, 2$. Each series i in a given block b is decomposed into an idiosyncratic component e_{Xbit} and a common component $\Lambda_{G,bi}(L)G_{bt}$ which it shares with other variables in the same block. Each block-level factor G_{bjt} has a serially correlated block-specific component e_{Gbjt} and a common component $\Lambda_{F,bj}(L)F_t$ which it shares with all other blocks. Finally, the economy-wide factor F_t is assumed to be serially correlated.

In this model, variables within a block can be correlated through F_t and the e_{Gbjt} 's, but variables between blocks can be correlated only through F_t .

Estimation procedure by MCMC:

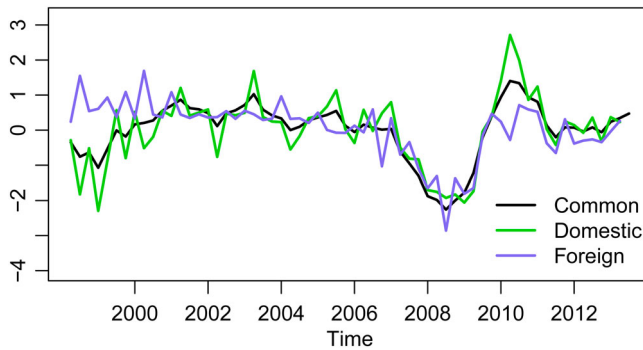


Figure 2. Evaluated common, domestic and foreign leading factors from the hierarchical factor model.

$$\text{Let } \Lambda = (\Lambda_G, \Lambda_F), \Psi = (\Psi_F, \Psi_G, \Psi_X), \Sigma = (\Sigma_F, \Sigma_G, \Sigma_X).$$

- (1) Organize data into blocks to yield X_{bt} , $b = 1, 2$. Use principal components to initialize $\{G_t\}$ and $\{F_t\}$. Use these to produce initial values for Λ , Ψ and Σ .
- (2) Conditional on Λ , Ψ , Σ and $\{F_t\}$ draw $\{G_t\}$ taking into account time varying intercepts.
- (3) Conditional on Λ , Ψ , Σ and $\{G_t\}$ draw $\{F_t\}$.
- (4) Conditional on $\{G_t\}$ and $\{F_t\}$, draw Λ , Ψ and Σ
- (5) Return to 2.

10,000 iterations were made, and first 500 were dropped out as a ‘burn-in’. The domestic and foreign leading factors were evaluated calculating the expectation from posterior distributions. The estimations were carried out using *dIm* (Petris, 2010) package of statistical software *R*.

Another round of simulations was carried out to compare the results. 1,00,000 iterations were made and first 50,000 were discarded. The results are almost identical (mean absolute difference in acquired factors was 0.0034, which is very low since the variance of factors is set to 1). The resulting factors are plotted in Figure 2.

The results indicate that even though the extracted domestic and foreign factors are a bit noisy, they depicted the economic crisis and recovery in 2007–2011 pretty well. As expected, domestic and foreign factors have similarities with common factor (domestic factor $G_{1,t}$ correlates with common factor by 0.90, foreign factor $G_{2,t}$ correlation with common factor F_t is 0.67).

3.3. Structure validation

In order to validate the imposed structure another factor model was built which had 2 factors in a single block, i.e. domestic and foreign leading series were pooled together and 2 dynamic factors were evaluated from that pool. The correlation matrix of factors from structural approach $G_{1,t}$, $G_{2,t}$ and factors from non-structural approach $F_{1,t}$ and $F_{2,t}$ is in Table 1.

It can be identified that even without the imposed block structure, the factors from structural approach correlate with 2 factors from non-structural approach by 0.96 and 0.72. This means that the information of series from 2 different blocks naturally form 2

Table 1. Correlations between factors acquired from structural and non-structural approach.

	$G_{1,t}$	$G_{2,t}$	$F_{1,t}$	$F_{2,t}$
$G_{1,t}$	1	0.41	0.05	0.96
$G_{2,t}$	0.41	1	0.72	0.43
$F_{1,t}$	0.05	0.72	1	0.05
$F_{2,t}$	0.96	0.43	0.05	1

different factors. The structural approach lets us name those factors and give them interpretation which could be very difficult to justify in the case of non-structural factors.

3.4. Varying factor load evaluation

In order to capture the load of domestic and foreign indicators on the future growth of Lithuanian economy, a linear model following the idea of Stock and Watson (2002) was considered in the form of regressing the growth of coincident index on both leading factor estimates (10):

$$\Delta C_{t+2} = \alpha_1 G_{1,t} + \alpha_2 G_{2,t} + \varepsilon_{t+2}. \quad (10)$$

Here $G_{1,t}$ is a the domestic leading factor, and $G_{2,t}$ is the foreign leading factor. The expression from Equation (6) was reduced to Equation (10) based on statistical significance of parameters in linear regression.

Since we are more interested in the dynamics of the α parameters, Equation (10) had to be modified to include time-varying coefficient on factors. Therefore, a dynamic linear model was built

$$\Delta C_{t+2} = (1 - \alpha_t)G_{1,t} + \alpha_t G_{2,t} + \varepsilon_{t+2}, \quad (11)$$

$$\alpha_{t+1} = \psi\alpha_t + u_t. \quad (12)$$

The hypothesis that we are trying to validate is that the proportion of economic growth forecast explained by foreign indicators is increasing over time. Under this specification our hypothetical statement means that parameter α_t should be increasing over time.

The constraint that parameters α_1 and α_2 from Equation (10) should sum to one was added in order have fewer parameters to evaluate because the data set is not big enough to provide sufficient information to evaluate 2 varying parameters to the desired precision. This specification expresses our interest in the foreign impact relative to the domestic one. Another measure that had to be taken was rescaling of evaluated factors $G_{1,t}$ and $G_{2,t}$ in order to comply the requirement that the new model has to explain the same amount of variance as a constant parameter model (10).

The parameters of this model were evaluated by maximum likelihood assuming i.i.d. Gaussian innovations ε_{t+2} and u_t . The parameters estimated from regression (10) were used to set initial state α_t . The plot of dynamic coefficient α_t (extracted with Kalman filter) is in Figure 3.

It can be identified from Figure 3 that parameter α_t is increasing, which means that Lithuanian economy is more and more intertwined with foreign economies. This result also validates our hypothesis about the increasing amount of explained forecast by foreign indicators. It leads to a conclusion that globalization can be measured and its effect on focal economy is increasing in magnitude over time.

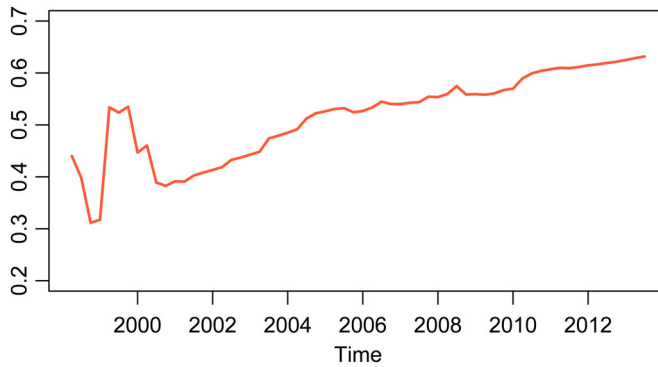


Figure 3. Evaluated parameter series α_t – the load of foreign factor impact on future growth of economy.

4. Conclusions

In this paper the issue of foreign influence, especially the globalization effect on the growth of focal economy is addressed. As a result, the method to quantify the impact of domestic and foreign variables on the future growth of focal economy as measured by the leading economic index is developed. Hierarchical dynamic factor model was used to expand the conventional framework for constructing leading economic indicators. Using this structural approach, the domestic and foreign drivers of economy were distinguished and their effects quantified.

In order to apply this new method, as a practical task a hypothesis was formed: due to globalization the proportion of economic growth forecast explained by foreign indicators is increasing over time. Thus, a dynamic linear model was built to evaluate time-varying effect of foreign and domestic indicators and it was applied on Lithuanian data. Lithuanian example showed that foreign series correspond to an amount which is increasing over time. This confirms not only that incorporating foreign data is useful, but also that in this framework the globalization effect is visible and it can be monitored using dynamic linear models. These conclusions state that the hypothesis was validated and foreign information corresponds to an amount of forecast explained that is increasing over time.

The strong feature of the proposed new method is flexibility in ways to impose the structure and restrictions. Also this method could be used to evaluate weights of various indicators by using different divisions: foreign/domestic, regional/global, real variables/price variables, etc. It could even be expanded to include more levels in hierarchy (e. g. foreign block could consist of sub-blocks using geographical division). Since every simulation takes time the largest drawback of this method is the time-consuming process of selecting the best specification.

Notes

1. The seasonal adjustment was applied by national statistical agencies.
2. *IM*, *WT*, *IP* series were acquired from Statistics Lithuania. The source of *RE* series is State Enterprise Centre of Registers.

3. Dickey–Fuller test failed to reject the null hypothesis about unit root existence and Johansen test did not provide evidence about cointegration.
4. The number 20 was selected on the account that Lithuania’s top 20 trade partners on average cover 90% of exports and 92% of imports and the rest of partners were discarded as having insignificant influence.
5. The seasonal adjustment procedure used was X-13ARIMA-SEATS developed by US Census Bureau (<http://www.census.gov/srd/www/x13as/>).
6. Both of these indicators were not considered due to insufficient observations.
7. The higher order AR(p) processes were considered but modelling showed that coefficients for lags 2 and greater were statistically insignificant.

Acknowledgments

I would like to thank Virmantas Kvedaras and Marijus Radavicius for valuable discussions and guidance. I am also grateful to anonymous reviewers for helpful comments and constructive suggestions.

Disclosure Statement

No potential conflict of interest was reported by the author.

References

- Auerbach, A. J. (1982). The index of leading indicators: Measurement without theory, thirty-five years later. *The Review of Economics and Statistics*, 64, 589–595.
- Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146, 304–317.
- Bandholz, H. (2005). *New composite leading indicators for Hungary and Poland* (Working Paper, No. 3), Ifo Institute.
- Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. New York, NY: NBER.
- Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81, 541–553.
- Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review*, 39, 969–996.
- Cubadda, G., Guardabascio, B., & Hecq, A. (2013). A general to specific approach for constructing composite business cycle indicators. *Economic Modelling*, 33, 367–374.
- Diebold, F. X., & Rudebusch, G. D. (1996). Measuring business cycles: A modern perspective. *The Review of Economics and Statistics*, 78, 67–77.
- Drechsel, K., & Scheufele, R. (2010). *Should we trust in leading indicators? Evidence from the recent recession* (Discussion paper No. 10), IWH.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32, 407–499.
- Fichtner, F., Rueffer, R., & Schnatz, B. (2009). *Leading indicators in a globalised world* (Working paper No. 1125), European Central Bank.
- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *The Review of Economics and Statistics*, 82, 540–554.
- Gaudreault, C., Lamy, R., & Liu, Y. (2003). *New coincident, leading and recession indexes for the canadian economy: An application of the stock and watson methodology* (Working paper No. 12). Canada: Department of Finance.
- Issler, J. V., & Vahid, F. (2003). *The missing link: Using NBER recession indicator to construct coincident and leading indices of economic activity* (Economic Essay No. 492). Rio de Janeiro: Graduate School of Economics.
- Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60, 1–22.

- Kim, C.-J., & Nelson, C. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *The Review of Economics and Statistics*, 80, 188–201.
- Mapa, D., & Simbulan, M. C. (2014). *Analyzing and forecasting movements of the Philippine economy using the dynamic factor models (DFM)* (Paper No. 54478). Munich: Munich Personal RePEc Archive.
- Mariano, R. S., & Murasawa, Y. (2002). *A new coincident index of business cycles based on monthly and quarterly series* (Working paper No. 14). New York: Penn Institute for Economic Research.
- Moench, E., Ng, S., & Potter, S. (2009). *Dynamic hierarchical factor models* (Report No. 412), Federal Reserve Bank of New York.
- OECD System of Composite Leading Indicators, (2012). Retrieved November 9, 2014, from <http://www.oecd.org/std/leading-indicators/41629509.pdf>.
- Petris, G. (2010). An R package for dynamic linear models. *Journal of Statistical Software*, 36(12), 1–16.
- Reklaite, A. (2011). Coincident, leading and recession indexes for the Lithuanian economy. *The Baltic Journal of Economics*, 11, 91–107.
- Schulz, C. (2007). *Forecasting economic growth for Estonia: Application of common factor methodologies* (Working paper No. 9). Tallinn: Bank of Estonia.
- Shiskin, J., & Moore, G. H. (1968). *Composite indexes of leading, coinciding and lagging indicators 1948–67* (Supplement to NBER report One). Cambridge, MA: NBER.
- Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual*. Cambridge, MA: MIT Press, pp. 351–394.
- Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident economic indicators (Chapter 4). *Leading economic indicators: New approaches and forecasting records*. Cambridge, MA: Cambridge University Press.
- Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147–162.
- The Conference Board. (2001). *The handbook of cyclical indicators*. Retrieved November 7, 2014, from http://www.conference-board.org/pdf_free/economics/bci/BCI-Handbook.pdf.

Appendix 1

A.1. Selected leading variables

Block	Country	Variable
Domestic	Lithuania	Lithuania's investment abroad
Domestic	Lithuania	Foreign direct investment to Lithuania
Domestic	Lithuania	Livestock and poultry
Domestic	Lithuania	A profitable share out of the total number of enterprises
Foreign	Japan	Final consumption expenditure of general government
Foreign	Portugal	Final consumption expenditure of households, total
Foreign	Japan	Household and NPISH final consumption expenditure
Foreign	France	Real Gross Domestic Product per capita
Foreign	Cyprus	Gross saving
Foreign	USA	Net saving
Foreign	Finland	Real labour productivity per hour worked
Foreign	Finland	Real labour productivity per person employed
Foreign	Latvia	Real labour productivity per person employed
Foreign	Denmark	Consumer Confidence Index
Foreign	France	Business Confidence Index
Foreign	France	Consumer Confidence Index

Note: The source of domestic series is Statistics Lithuania, of foreign series – Eurostat.

Appendix 2

A.2. Detecting and removing collinear series

After the hard thresholding was performed the foreign data set consisted of 16 time series. Unfortunately, this set included collinear indicators. Six time series in this set were on real labour productivity and it was likely that they carry very similar information. In order to identify the collinear time series, the data were scaled and Euclidean distance (A1) was calculated between each pair of time series.

$$d(x_i, x_j) = \|x_i - x_j\|_2 = \left(\sum_t |x_{i,t} - x_{j,t}|^2 \right)^{1/2} \quad (\text{A1})$$

Afterwards the hierarchical clustering was performed and cluster dendrogram was inspected.

The generated dendrogram in Figure A1 reveals a cluster of six time series which happened to be the same series on labour productivity mentioned earlier. The second largest cluster of indicators at the selected level, as marked by a red dashed line, is of size 2 and is formed of series with codes 1017 and 1464. In order to diminish the large cluster to the size of two time series, the least angle regression algorithm was applied using future growth of CEI as the variable of interest. The results revealed that time series with codes 1817 and 1853 are most informative. They correspond to real labour productivity per person employed in Latvia and in Finland. The other four time series from the large cluster were removed and the resulting data set was used in further steps of modelling.

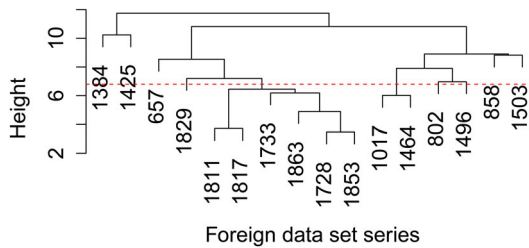


Figure A1. Cluster dendrogram of foreign time series selected after hard thresholding. The labels note the time series code (number).