MACROECONOMIC ANALYSIS AND FORECASTING: AN EMPIRICAL INVESTIGATION

Doctorando

Carlos Cuerpo Caballero

Directora

Pilar Poncela Blanco

Departamento de Análisis Económico: Economía Cuantitativa

Facultad de Ciencias Económicas y Empresariales

Universidad Autónoma de Madrid

Junio 2017

A mi padre y a mi abuelo

AGRADECIMIENTOS

Me gustaría agradecer en primer lugar a mi directora, Pilar Poncela, que desde el principio entendió la dificultad de compaginar tesis y trabajo. Su apoyo ha sido un gran activo en todo este proceso. Su visión rigurosa y el grado de atención por el detalle son lecciones valiosas para alguien perdido en la vorágine del día a día. Su empuje en el tramo final ha supuesto un factor diferencial para llevar este proyecto a término.

Mi agradecimiento también al Departamento de Análisis Económico: Economía Cuantitativa de la Facultad de Ciencias Económicas y Empresariales de la Universidad Autónoma de Madrid, y en especial a su directora, Aránzazu de Juan, porque desde el principio hasta el final todo han sido facilidades y ayuda, tanto en el ámbito académico como en el administrativo.

Muchas personas han contribuido a hacer posible este proyecto, en especial compañeros de trabajo y amigos, como Luis González Calbet, Enrique Quilis, Ángel Cuevas, Jorge Prieto y Fernando Mateos, cuya colaboración y apoyo han hecho posible que no haya arrojado la toalla por el camino.

Mi agradecimiento más sincero también a Luis Regino Murillo y Juan Vega, que supieron incentivarme y fomentar mis ganas por formarme en mi etapa en la Universidad de Extremadura y cuya influencia ha permanecido muy presente desde entonces.

En el plano familiar, toda la gratitud del mundo a mis padres, Gregorio y Manoli. Su esfuerzo y sacrificios para que mi hermano y yo siempre dispusiéramos de los mejores medios y oportunidades han sido las llaves que nos han permitido abrir todas las puertas que nos hemos encontrado por el camino. Gracias a mi hermano, por ser un ejemplo a seguir, un referente, un apoyo y mi mejor amigo. Gracias a mi abuelo, por tener la suficiente visión como para mejorar la vida de todos los que hemos venido después. A mi tío, por creer en mí y estar siempre presente. A mi abuela, a mi tía Gloria y a mis primos Pablo y Enrique, por las risas, los abrazos y el apoyo incondicional.

Por último, gracias a ti, Ana, por entender y apoyar todos mis proyectos, aun a costa de muchos sacrificios. Por saber animarme y evitar que todo el esfuerzo hubiera sido en vano. Gracias por traer a Cristina a nuestras vidas; “la loca” es la que le da sentido a todo.

RESUMEN

Esta tesis analiza temas de especial relevancia para los responsables de política económica a raíz de la crisis financiera. En un primer paso, el enfoque se centra en la estimación de la holgura o brecha de producción de la economía en función de la acumulación de distintos desequilibrios macroeconómicos. El análisis presenta un enfoque novedoso centrado en la especificación del modelo y no en la selección previa del método de estimación. Los métodos multivariantes, junto con el filtro de Kalman, se consideran una opción de modelización apta que consigue un compromiso adecuado entre criterios difíciles de conjugar a priori; ajuste estadístico, fundamentación económica y replicabilidad de los resultados. El enfoque se ilustra con una aplicación para la economía española, seleccionando el mejor modelo entre combinaciones bivariantes del Producto Interior Bruto y 52 variables de acompañamiento. En un segundo paso, esta disertación evalúa la capacidad predictiva fuera de muestra de modelos estructurales y no estructurales utilizando datos de frecuencia trimestral correspondientes a los últimos 37 años para siete agregados macroeconómicos: PIB, consumo privado, inversión privada, empleo, deflactor del PIB, salarios reales y tipo de interés nominal. La capacidad predictiva se evalúa mediante un procedimiento recursivo a través de cuatro dimensiones diferentes: una dimensión temporal (de uno a ocho trimestres), una dimensión contextual (período de crecimiento suave y fase de recesión), una dimensión específica del país (resultados para España, zona euro y Estados Unidos) y una dimensión específica del modelo (comparación de modelos estructurales de equilibrio general con los modelos de referencia tradicionales, como los Vectores autor regresivos o VAR y los VAR Bayesianos). Finalmente, el tercer paso tiene como objetivo calibrar la importancia relativa de los canales de transmisión internacional de las perturbaciones económicas. Con el fin de medir de manera óptima la fuerza relativa de las interconexiones existentes entre países, el análisis circunscribe primero la transmisión de los choques a tres canales relevantes; flujos comerciales, exposiciones bancarias y contagio a través de la percepción de los agentes (reflejada en el co-movimiento de los rendimientos de los bonos soberanos). A continuación, se obtiene el esquema de ponderación óptimo dentro de un modelo VAR Global (GVAR), minimizando el error de predicción del PIB a corto plazo del modelo. Una vez

que se calibran los pesos relativos óptimos de los canales, se utilizan conjuntamente con los flujos bilaterales para construir un indicador ponderado que refleje el potencial de efectos desbordamiento o *spillover* entre países. Dependiendo del país de referencia, este indicador arroja el potencial de desbordamiento interno (qué países son relativamente más importantes para una economía específica y en qué medida) así como el externo (qué países son más dependientes de la evolución de una economía seleccionada) un país.

ABSTRACT

This dissertation analyzes topics of special relevance for policy makers in the aftermath of the financial crisis. In a first step, the focus lies in the estimation of the slack of the economy in pseudo-real time according to the accumulation of selected macroeconomic imbalances. The analysis presents a novel approach putting the focus on the specification of the model rather than on the prior selection of the methodology itself. Multivariate methods, coupled with Kalman filtering are considered as an adequate modeling choice reaching a compromise between three criteria that are difficult to reconcile a priori; statistical fit, economic soundness and replicability. The approach is illustrated with its application to the Spanish economy, by selecting the best model amongst bivariate combinations of Gross Domestic Product (GDP) and 52 accompanying variables. In a second step, this essay assesses the out-of-sample forecasting performance of structural and non-structural models with quarterly data covering the last 37 years for seven macroeconomic aggregates: GDP, private consumption, private investment, employment or total hours worked, the GDP deflator, real wages and the nominal interest rate. The forecasting performance is assessed using a recursive procedure through four different dimensions: a time dimension (from one to eight quarters ahead), a contextual dimension (smooth growth period and recession phase), a country-specific dimension (results for Spain, USA and the euro area) and a model-specific dimension (comparison of structural general equilibrium models against traditional benchmarks such as Vector Autoregressive or VAR models and Bayesian VARs). Finally, the third step aims at calibrating the relative importance of the channels for the international transmission of shocks. To optimally weight the relative strength of the existing interlinkages between countries, the analysis first circumscribes the transmission of shocks to three encompassing channels; namely trade flows, banking exposures and contagion via agents' perception (reflected in the co-movement of sovereign yields). Then the optimal weighting scheme is obtained within a Global VAR framework, by minimizing the short-term GDP forecast error of the model. Once the relative weights of the channels are calibrated, they are used together with the actual bilateral flows to construct a weighted indicator reflecting the potential for spillovers between countries. Depending on the reference country, this indicator yields

the inward (which countries are relatively more important for a specific economy, and to what extent) as well as the outward (which countries are more dependent on the evolution of a selected economy) spillover potential for a country.

CONTENTS

List of Tables

List of Figures

1 INTRODUCTION

1.1 Motivation

Policy-makers and academics live in very different worlds and face different constraints

Benoît Cœuré, 2014

The coming of a generation of young and brilliant economists in the late 70s (including E. Phelps, F. Kydland, E. Prescott, T.J. Sargent and N. Wallace amongst others) led by Robert Lucas started a scientific revolution *à la* Kuhn in the field of macroeconomics (see de Vroey and Malgrange, 2011). The business cycle took a dominant role as a research field. Cyclical fluctuations were originated by optimizing agents making optimal use of imperfect information in the face of economic shocks (Lucas and Rapping, 1969). Kydland and Prescott (1982) went a step further, bringing these models closer to the empirical evidence by assigning realistic values to the parameters (calibration techniques) and providing numerical solutions to the resulting systems of simultaneous equations (real business cycle models).[1]

The contributions by Lucas and the other leading New Classical economists inaugurated the Dynamic Stochastic General Equilibrium (DSGE) era, defined by dynamic economic relationships (with optimizing agents) and stochastic shocks affecting the economy, in a general equilibrium framework. This framework was flexible enough to overcome its initial weaknesses with progressive improvements on the margin. A decade after and ever since, New Keynesian prominent economists (e.g. G. Akerloff, Y. Yellen, O. Blanchard, G. Mankiw, B. Bernanke, M. Woodford, J. Gali or N. Kiyotaki) would keep the Real Business Cycle core of optimizing agents with microfounded decision problems in a dynamic general equilibrium context, but challenge some more

[1] Undoubtedly taking advantage of the on-goingcomputational advances and data availability.

peripheral assumptions, advocating for imperfect competition set-ups with nominal frictions and real rigidities, while focusing on monetary aspects and monetary policy implications. This second wave of enhanced DSGE models was coined the New Neoclassical Synthesis (Goodfriend, 2004) and widely accepted for business cycle modelling after decades of fights amongst Classical and Keynesian advocates. The Great Moderation (see Bernanke, 2004 and McConnell and Perez-Quiros, 2000) together with advances in econometric techniques that eased the estimation of large-scale DSGE models (especially Bayesian econometrics techniques) provided a fertile ground for the adoption of these models amongst policymakers, particularly by Central Banks as they serve a great variety of purposes such as policy guiding, historical analysis and counterfactual experiments.[2]

However, hard lessons were learnt over the last financial crisis, both in the main academic circles and in the policymakers' headquarters. Blanchard (2014) acknowledged that mainstream DSGE models were designed to provide a description of cyclical fluctuations in normal times only and not to cover tail events or "dark corner" situations. Policymakers were thus not properly equipped to weather the greatest crash since the Great Depression. Pre-crisis consensus was shattered and practitioners abandoned the straitjacket provided by the holistic DSGE paradigm and went back to a disaggregated approach, looking at different issues with specific instruments, trying to overcome some of the previous consensus shortcomings.

In this context, this dissertation aims at shedding light on some of the most contentious issues for policymakers, with three empirical chapters:

1. *Imbalances and the Business Cycle*: Estimation of the slack of the economy in pseudo-real time in view of existing macroeconomic imbalances.

2. *Forecasting along the Business Cycle*: Forecasting exercise with structural and data-driven models over different phases of the business cycle.

[2] The Great Moderation or prolonged period of low inflation rates and shrinking variance in economic growth was widely credited to the policies designed by central bankers according to DSGE guidelines (Landmann 2014).

3. *International Transmission of Shocks*: Calibration of the international transmission channels according to weighted trade, financial and contagion flows.

1.2 Imbalances and the Business Cycle

Policymakers strive to understand the dynamics of the business cycle and determine its specific location as a policy-relevant variable. The slack in the economy or output gap is, however, not observable and surrounded by considerable uncertainty.

Along the quest for the best output gap estimate, the literature has developed a myriad of estimation techniques over the last decades, ranging from data-driven univariate filters to structural general equilibrium models.[3] However, the horse race in search of an optimal output gap estimation methodology seems far from settled.

The uncertainty surrounding the output gap estimates has proven a challenging task, leading to unreliable estimates in real time, which is the policy-relevant time frame. Moreover, there is a lack of a well-defined metric or comparable benchmark for the different estimates.

This analysis presents an empirical approach overcoming these two limitations, based on a structural multivariate time series model and Kalman filtering, with an application to the Spanish economy.

This chapter defines a new selection algorithm based on a set of selection criteria defined along three dimensions: (i) statistical goodness (e.g. minimizing the end-point problem); (ii), economic soundness (e.g. "smell test" and consistency with selected stylized facts); and (iii) transparency or replicability.

The focus for the selection of a specific output gap estimate is thus diverted from the traditional comparison between different methodologies along the selected criteria. As a novelty, different specifications of the multivariate

[3] See for example Álvarez and Gómez-Loscos (2017), Alichi (2015) and Murray (2014) for a review of different estimation techniques.

unobserved components model are tested by combining output series with potential candidate variables sharing information about the business cycle. These include domestic (e.g. construction investment, capacity utilization, unemployment) open-economy (e.g. current account, exchange rate), financial (credit to non-financial corporations, M3) and price (e.g. GDP deflator, CPI, house prices) candidates. The selected approach allows for country-specific cycle definitions, generalizing the work in Borio *et al*. (2017) and Alberola *et al*. (2013).

Multivariate filters and the unobserved components multivariate Kalman filter technique represent a good compromise between these criteria. The multivariate framework allows for a country-specific approximation as it could accommodate specific cycles (financial, external, investment, fiscal, etc.) by considering additional variables related to the cycle.

1.3 Forecasting Along the Business Cycle

The ability of economists to forecast the main aggregate macroeconomic variables has undergone a complete revolution in the last 40 years. However, the absence of an agreed model to forecast the main economic aggregates at different time horizons remains an important challenge for econometric analysis, especially considering the recent financial crisis.

Initially, multivariate vector autoregressive (VAR) models became the workhorse in macroeconomic forecasting, following the work by Box and Jenkins (1970) and Sims (1980). These non-structural models had two main advantages. On the one hand, they were not subject to the Lucas (1976) critique, as their forecasts were not tied to a specific path of the policy variables. On the other hand, Unrestricted VAR (UVAR) models did not impose excessive identification restrictions, leading to a better in-sample fit. However, good in-sample fit did not grant a good out-of-sample forecasting performance, as indicated in Stock and Watson (1996) work. Increasing the number of variables could generate inaccurate estimates and bad predictive results due to over-fitting.

These limitations led to the development of two important lines of research. First, Bayesian VAR (BVAR) models (Doan, Litterman and Sims, 1984), restricting the parameter space by imposing Bayesian constraints based on prior information. Second, dynamic factor models (DFM) (see for example Sargent and Sims, 1977) assume that a few shocks can explain most of the dynamics of macroeconomic aggregates and express time series as the sum of two orthogonal components: a common one capturing the dynamic shared by all series and an idiosyncratic one, understood as a residual.

A parallel strand of work related to the development of structural DSGE models, pioneered by Real Business Cycle literature (Hansen and Sargent, 1980 and Kydland and Prescott, 1982), provided a connection between theory and data thanks to the state space representation of decision rules obtained from the solution of the models. However, their forecasting capability was not seriously put to the test until the contribution from Smets and Wouters (2004), who captured the statistical features of the main macroeconomic aggregates via Bayesian estimation techniques applied to large scale models.

Despite the vast literature on the predictive capabilities of competing models, there are several weaknesses, unexplored issues or inconclusive results. First, so far, the clear majority of forecast comparison exercises have focused on stable business cycle periods. Second, research on DSGE models is generally biased towards theoretical improvements without primarily aiming at forecasting gains. More refined models might help establishing a narrative of cyclical developments ex-post but might not improve their forecast accuracy with respect to misspecified models.

This chapter presents an empirical approach overcoming these two limitations, based on a comparative analysis of the out-of-sample forecasting performance of structural and non-structural models with quarterly data covering the 1980Q1 to 2016Q4 period for seven macroeconomic aggregates: Gross Domestic Product (GDP), private consumption, private investment, employment or total hours worked, the GDP deflator, real wages and the nominal interest rate.

The forecasting performance is assessed using a recursive procedure through four different dimensions: a time dimension (from one to eight quarters ahead), a contextual dimension (smooth growth period pre-crisis [2003Q1-2007Q2] as well as post-crisis [2012Q1-2016Q4] and recession phase

[2007Q3-2011Q4]), a country-specific dimension (results for Spain, USA and the Euro area) and a model-specific dimension (DSGE and non-structural models).

The main contribution of this chapter consists in comparing the forecasting performance of structural versus non-structural models, both in a smooth-growth period (2003Q1-2007Q3) and during a crisis (2007Q3-2011Q4) to overcome the first two issues, with the post-crisis period as a robustness exercise. This aspect is of relevance for policymakers as it could simplify their toolkit of forecasting models and make the selection of a specific approach state-dependent.

The empirical exercise covers three different economies, the euro area, Spain and the United States. This extensive approach overcomes the second difficulty as the estimated DSGE model implies varying degree of misspecification depending on the structural features and stylized facts of the selected economy.

Another contribution of this chapter consists in assessing the forecasting gains at different time horizons, to validate whether the introduction of theoretical restrictions implies a better performance in the medium run, when the informational gain for non-structural models from sticking closer to the data is exhausted and their mean reversion component takes over.

1.4 International Transmission of Shocks

Since the onset of the Economic and Monetary Union, its Member States (MS) have experienced an enhanced interdependence, with concerns from policymakers on potential spillover effects from idiosyncratic shocks.

The generation of cross-country spillover effects ultimately depends on the relative strength of existing transmission channels. The bilateral links between two countries are not homogeneous across channels and thus concentrating on a specific conduit will ultimately yield a biased picture of the potential for spillovers between two economies.

Moreover, the relative merits of each of the different channels change over time, along with the business cycle. From a cross-sectional perspective, the bilateral linkages between countries might appear strikingly different according to each one of these channels, at any point in time. The relative weight attached to each one of these channels will therefore have critical implications in the assessment of the existing potential for spillovers.

The optimal weighting scheme for the different channels is, however, difficult to grasp empirically and the literature generally opts for simplifying assumptions, such as focusing on one channel as in Pesaran *et al*. (2004).

This chapter opts for a novel approach, calibrating the relative weights of different transmission channels in a Global VAR framework, according to the short-term GDP forecast accuracy of the model.

To optimally weight the relative strength of the existing interlinkages between countries, the transmission of shocks is circumscribed to three channels; namely trade flows, banking exposures and contagion via agents' perception (reflected in the co-movement of sovereign yields).

Once the relative weights of the channels are calibrated, they are used together with the actual bilateral flows to construct a weighted indicator reflecting the potential for spillovers between countries. Depending on the reference country, this indicator yields the inward (what countries are relatively more important for a specific economy, and to what extent) as well as the outward (what countries are more dependent on the evolution of a selected economy) spillover potential for a country.

These results shed some light on the reallocation of systemic relevance amongst countries and can be useful when calibrating processes such as the on-going rebalancing within the euro area or Brexit concerns.

1.5 Organization

The rest of this dissertation is organized as follows. Chapter 2 presents a methodological approach for output gap estimation fulfilling a set of pre-defined selection criteria covering statistical as well as economic conditions. The methodology is illustrated for the Spanish economy. Chapter 3 conducts an out-of-sample forecasting exercise with structural (Dynamic Stochastic General Equilibrium) and non-structural (Dynamic Factor, VAR and BVAR) models. It is implemented over different stages of the business cycle and for short and medium term horizons, to Spanish, euro area as well as US data. Chapter 4 estimates a GVAR model to calibrate the potential for spillovers amongst the main euro area and OECD countries. Finally, Chapter 5 concludes and presents avenues for future research.

1.6 References

Álvarez, L.J., Gómez-Loscos, A. (2017). "A Menu on output gap estimation methods," Bank of Spain, Working Paper n. 1720.

Alberola, E., Estrada, A. and Santabarbara, D. (2013). "Growth beyond imbalances. Sustainable growth rates and output gap reassessment," Bank of Spain, Working Paper n. 1313.

Alichi, A. (2015). "A new methodology for estimating the output gap in the United Sates," IMF Working Paper Series, Working Paper n. 144.

Bernanke, B.S. (2004). "The Great Moderation," Remarks at the meetings of the Eastern Economic Association, Washington, DC, February 20, 2004.

Blanchard, O. (2014). "Where dangers lurks," *Finance and Development*, September, 28-31.

Borio, C., Disyatat, P. and Juselius, M. (2017), "Rethinking potential output: Embedding information about the financial cycle," *Oxford Economic Papers* 69(3), 655-677.

Box, G., and Jenkins, G. (1970). "Time series analysis: Forecasting and control," Rev. ed., San Francisco, Holden-Day, c1976.

De Vroey, M. and Malgrange, P. (2011). "The history of macroeconomics from Keynes's General Theory to the present," Discussion Papers IRES n. 2011028, Université Catholique de Louvain.

Doan, T., Litterman, R. and Sims, C. (1984). "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews* vol. 3(1), 1–100.

Goodfriend M. (2004). "Monetary policy in the new neoclassical synthesis: A primer," *Economic Quarterly*, Federal Reserve Bank of Richmond, issue Sum, 21-45.

Hansen, L. and Sargent, T. (1980). "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, vol. 2(1), 7–46.

Kydland, F. y Prescott, E. (1982). "Time to build and aggregate fluctuations," *Econometrica*, vol. 50(6), 1345-70.

Landmann, O. (2014). "Short-run macro after the crisis: The end of the 'new' neoclassical synthesis?," Discussion Paper Series, University of Freiburg, Department of International Economic Policy, n. 27.

Lucas, R.E. y Rapping, L.A. (1969). "Real wages, employment and inflation," *Journal of Political Economy*, vol. 77(5), 721-754.

McConnell, M. and Perez-Quiros, G. (2000). "Output fluctuations in the United States: What has changed since the early 1980's?," *American Economic Review*, vol. 90(5), 1464-1476.

Murray, J. (2014) "Output gap measurement: judgement and uncertainty", U.K. Office for Budget Responsibility, Working Paper n. 5.

Pesaran, M.H., Shuermann, T. and Weiner, S.M. (2004), "Modeling regional interdependencies using a global error-correcting macroeconometric model," *Journal of Business and Economic Statistics* 22, 129-162.

Sargent, T., and Sims, C. (1977). "Business cycle modeling without pretending to have too much a priori economic theory", Working Papers n. 55, Federal Reserve Bank of Minneapolis.

Sims, C. (1980). "Macroeconomics and reality," *Econometrica*, vol. 48(1), 1–48.

Smets, F., and Wouters, R. (2004). "Forecasting with a Bayesian DSGE model: An application to the euro area", *Journal of Common Market Studies* vol. 42(4), 841–867.

Stock, J., and Watson, M. (1996). "Evidence on structural instability in macroeconomic time series relations", *Journal of Business & Economic* Statistics, vol.14(1), 11–30.

2 Imbalances and the Business Cycle

2.1 Introduction

Policymakers strive to understand the dynamics of the business cycle and pinpoint its specific location as it decisively determines the outcome of policy decisions. The slack or output gap, defined as the amount of unemployed resources (i.e. the distance to its potential output) is, however, not observable and surrounded by considerable uncertainty.

The literature has developed a myriad of estimation techniques over the last decades, ranging from data-driven univariate filters to structural general equilibrium models.[4] The horse race in search of an optimal output gap estimation methodology seems far from settled. On the one hand, the uncertainty surrounding the output gap estimates has proven a challenging task, leading to unreliable estimates in real time, which is the policy-relevant time frame. On the other hand, confronting output gap estimates with optimality criteria (both statistical and practical ones) has generally led to inconclusive results, as the former might be ill-defined or even incompatible and thus a selection algorithm would become necessary.

As output gap estimates imply the decomposition of observables into unobserved components, there is a lack of a well-defined metric or comparable benchmark for the different estimates. The selection criteria can generally be split into three dimensions. First, statistical goodness (SG) referring to elements such as minimizing the end-point problem or providing information on the precision of the estimates. Second, economic soundness (ES) implying ex-ante consistency between selected stylized facts and the

---

[4] See for example Álvarez and Gómez-Loscos (2017), Alichi (2015) and Murray (2014) for a review of different estimation techniques.

method's underlying assumptions. And third, transparency (TR) requirements as seen from a user-specific perspective, reflecting accountability elements such as likelihood of replication or data needs.

Figure 1. Optimality necessary requirements

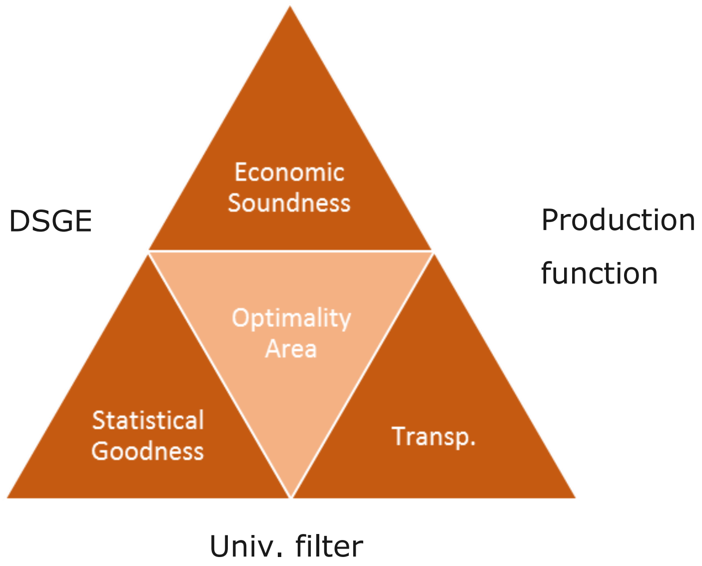


Figure 1 reflects potential tensions in fulfilling these criteria and represents trade-offs faced by some standard methodologies (DSGE models, univariate filters and the production function approach). The internal optimality area would represent methods fulfilling these three criteria (although in different degrees), which are considered a necessary methodological pre-requisite. They are not, however, sufficient conditions as ultimately the acceptance of a specific output gap estimate must pass the *smell test*, providing an acceptable country-specific narrative.

Multivariate filters and the unobserved components multivariate Kalman filter technique represent a good compromise between the different necessary criteria exposed in figure 1. First, the use of a multivariate framework allows for the consideration of additional economic relationships (Okun´s Law, Phillips Curve, etc.) going beyond univariate filters while at the same time imposing lighter economic priors than fully structural models and thus sticking more closely to the data. Second, the statistical properties of multivariate techniques clearly outperform other methods such as the production function approach, allowing for example for an integrated estimation of uncertainty. Third, multivariate approaches are generally not data-intensive and thus easily replicable and largely transparent, being more parsimonious than fully-fledged economic models. [5]

This chapter builds upon existing research on output gap measurement techniques and presents an approach for the selection of an output gap estimate that pivots around a multivariate unobserved components Kalman filter estimation, with an application to the Spanish economy. The focus for the selection of a specific output gap estimate is diverted from the traditional comparison between different methodologies along the three necessary criteria (ES, SG and TR). Instead, different specifications of the multivariate unobserved components model are tested according to pre-sepcified criteria, by combining output series with potential candidate variables sharing information about the business cycle, including domestic (capacity utilization, unemployment) open-economy (current account, exchange rate), financial (credit to non-financial corporations) and price (GDP deflator, CPI, house prices) candidates. The selected approach allows for country-specific cycle definitions, generalizing the work in Borio *et al*. (2017) and Alberola *et al*. (2013).

The chapter is structured as follows; section 2 reviews the estimation methodology, section 3 develops the selection criteria, section 4 and 5 present an application for Spain as a beta study and section 6 concludes.

2.2 Econometric methodology

This section develops the econometric approach used to estimate the output gap as well as the associated cyclical (or transitory) components. The chapter proceeds incrementally, starting with the basic univariate model and expanding it to the complete multivariate setting used for the estimation.

The econometric approach is based on the well-known Structural Time Series (STS) representation of a time series vector, see Clark (1987), Harvey (1989), Kuttner (1994), Kitagawa and Gersch (1996), Kim and Nelson (1999) and Durbin and Koopman (2001) among others. This method is rather general

[5] See for example Cotis *et al*. (2005) and references within for a complete discussion.

and flexible albeit keeping the number of parameters tightly controlled, in contrast with other econometric approaches (e.g. Vector of Autoregressions, VAR).

2.2.1 Univariate model

The structural decomposition provides an efficient way to estimate the output gap or, more generally, to decompose an observed time series as the sum of an arbitrary number of unobserved elements.

As a starting point, the (log-transformed) observed real Gross Domestic Product (GDP) can be decomposed as the sum of a non-stationary component and a stationary cycle as in [1]. The trend follows a random-walk plus time-varying drift, which is also stochastic and follows a random walk (see equations [2] and [3]). The cyclical dynamics is characterized by means of a second-order autoregressive process whose roots lie outside the unit circle (equation [4]).

[1] $$y_t = p_t + c_t$$

[2] $$p_t = g_{t-1} + p_{t-1} + v_t$$

[3] $$g_t = g_{t-1} + w_t$$

[4] $$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_t$$

Combining equations [1]-[4] the reduced-form MA model for y_t is given by:

[5] $$y_t = p_t + c_t = \underbrace{\underbrace{\frac{1}{(1-B)^2} v_{t-1}}_{\mathrm{I}(2)} + \underbrace{\frac{1}{(1-B)} w_t}_{\mathrm{I}(1)} + \underbrace{\frac{1}{(1-\phi_1 B - \phi_2 B^2)} e_t}_{\mathrm{I}(0)}}_{\mathrm{I}(2)}$$

Note that, in general, the structural model imposes an I(2) representation for the trend although, depending on the values of the variances of the shocks, this representation can collapse into an I(1) trend (with or without deterministic drift) or a linear trend plus noise. In this way, the model provides a flexible and parsimonious way to represent different non-stationary dynamics.[6]

Finally, the three shocks that drive the system are orthogonal Gaussian white noise innovations:

[6] $$\begin{bmatrix} v \\ w \\ e \end{bmatrix}_t \sim \text{iid}\, N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} v_v & 0 & 0 \\ 0 & v_w & 0 \\ 0 & 0 & v_e \end{bmatrix}\right)$$

The assumption of orthogonality can be relaxed at the price of making the identification of the shocks more difficult, see Clark (1987) for an in-depth analysis. In particular, to represent hysteresis the shocks that determine the long-term trend would be correlated with those that drive its short-term rate of growth, replacing [6] by a non-diagonal matrix:

[7] $$\begin{bmatrix} v \\ w \\ e \end{bmatrix}_t \sim \text{iid}\, N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} v_v & \gamma_{v,e} & 0 \\ \gamma_{v,e} & v_w & 0 \\ 0 & 0 & v_e \end{bmatrix}\right)$$

In the remaining of the chapter complete orthogonality among the shocks is assumed.

The structural model can be recast in state space format. The corresponding transition and measurement equations are given by:

[8] $$\underbrace{\begin{bmatrix} p_t \\ g_t \\ c_t \\ c_{t-1} \end{bmatrix}}_{S_t} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} p_{t-1} \\ g_{t-1} \\ c_{t-1} \\ c_{t-2} \end{bmatrix}}_{S_{t-1}} + \underbrace{\begin{bmatrix} v_t \\ w_t \\ e_t \\ 0 \end{bmatrix}}_{\zeta_t}$$

[9] $$y_t = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} p_t \\ g_t \\ c_t \\ c_{t-1} \end{bmatrix}}_{S_t} + \varepsilon_t$$

[6] In the Spanish case, GDP can be modeled following an I(1) structure plus a highly persistent Markov-switching drift, as shown in Cuevas and Quilis (2017). This specific structure can be linearly approximated by a random walk plus an evolving AR(1) drift.

The state space system can be cast in compact form as in [10]. Moreover, the state space shocks inherit the distributional assumptions of the structural shocks, and the assumption that the measurement errors are not related to the structural innovations and the inner orthogonality of the measurement errors (see equation [11]). The variance-covariance (VCV) matrices are given by equation [12]. Finally, the parameters of the model can be put together in a single vector, ϴ, see equation [13].

[10] $$\begin{aligned} S_t &= FS_{t-1} + \zeta_t \\ y_t &= HS_t + \varepsilon_t \end{aligned}$$

[11] $$\begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix} \sim iidN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \right)$$

[12] $$\begin{aligned} Q &= diag\begin{pmatrix} v_v & v_w & v_e & 0 \end{pmatrix} \\ R &= 0 \end{aligned}$$

[13] $$\theta = \begin{pmatrix} \phi_1 & \phi_2 & v_v & v_w & v_e \end{pmatrix}$$

Given some initial conditions for the state vector S_0 and assuming that the vector ϴ is known, the Kalman filter can be used to estimate the state vector and its corresponding standard error. In practice, the vector ϴ is not known and must be estimated from the sample. Fortunately, the state space format and the Kalman filter provide a feasible way to evaluate the likelihood function and, using numerical methods, to maximize it.

Once the ϴ parameters have been estimated, the Kalman filter is run to derive new initial conditions by means of backcasting (i.e., forecasting observations prior to the first observation). This process of backcasting can be done just by projecting forward the model using the reversed time series. In this way, a new set of initial conditions exerting a limited influence on the estimation of the state vector is derived by means of the Kalman filter. The complete algorithm can be stated as follows:

Estimation steps of the output gap using the Kalman filter

- Set initial parameters: Θ_0.
- Set initial conditions: S_0.
- Maximum likelihood estimation of Θ.
- Setting new initial conditions $S_{0,1}$.
- One-sided (concurrent) estimates of state vector.
- Two-sided (historical) estimates of state vector.

Before turning to the multivariate extension, some practical comments are in order:

- Initial conditions for the state vector are provided using a diffuse prior centered on zero with an arbitrarily large VCV matrix. Those initial conditions are required to run the Kalman filter and the optimization algorithm.
- The maximum likelihood estimation (MLE) is implemented numerically via the *fminunc* function from the Matlab optimization toolbox. The definition of the objective function incorporates the constraints that ensure the non-negativity of the variances and the stationary nature of the AR(2) parameters.
- The one-sided (or concurrent) estimates of the state vector are obtained running recursively the Kalman filter from t=1 to t=T (forward in time). This estimate considers only the information available from t=1 to t=h to estimate the state vector at time t=h and is very useful to analyze the state of the system on a real-time basis.
- On the other hand, the two-sided (or historical) estimates of the state vector are obtained running recursively the Kalman filter from t=T to t=1 (backward in time), using as initial conditions the terminal concurrent estimates. This process considers all the information available from t=1 to t=T to estimate the state vector at any time t=h, $1 \leq h \leq T$. This estimate is not useful for real-time analysis since it incorporates information not available a t=h to evaluate the state of

the system at that time and hence introduces some form of hindsight bias.

- However, two-sided estimates are optimal from a statistical perspective since they incorporate all the available information from t=1 up to time t=T to estimate the state vector in any intermediate point and, due to their symmetric nature. Note that this symmetry is due to the fact that the filter runs backward from estimates derived forward. In this way, two-sided filtering does not introduce any form of phase-shift in the estimates.

2.2.2 Multivariate model

The multivariate structural approach extends its univariate counterpart just by including additional variables whose stationary component is related to the output gap. This extension allows for the introduction of relevant macroeconomic stylized facts (as the Okun's Law, the Phillips Curve, etc.).

In this way, their observed values, properly filtered, provide additional information to estimate the output gap. The trend of the additional variables can be I(2) or I(1). For the sake of simplicity, let us consider two additional variables, one with an I(2) trend and the other with an I(1) trend.

The structural representation of the I(1) or I(2) variable is given by [14] or [15], respectively.

[14]
$$\begin{aligned} y_{1,t} &= p_{1,t} + c_{1,t} \\ p_{1,t} &= p_{1,t-1} + v_{1,t} \\ c_{1,t} &= \alpha_1 c_t + e_{1,t} \end{aligned}$$

$$\begin{bmatrix} v_{1,t} \\ e_{1,t} \end{bmatrix} \sim \text{iid}\, N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \nu_{v,1} & 0 \\ 0 & \nu_{e,1} \end{bmatrix}\right)$$

[15]
$$\begin{aligned} y_{2,t} &= p_{2,t} + c_{2,t} \\ p_{2,t} &= p_{2,t-1} + g_{2,t-1} + v_{2,t} \\ g_{2,t} &= g_{2,t-1} + w_{2,t} \\ c_{2,t} &= \alpha_2 c_t + e_{2,t} \end{aligned}$$

$$\begin{bmatrix} v_{2,t} \\ w_{2,t} \\ e_{2,t} \end{bmatrix} \sim iid\;\; N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \nu_{v,2} & 0 & 0 \\ 0 & \nu_{w,2} & 0 \\ 0 & 0 & \nu_{e,2} \end{bmatrix}\right)$$

The transition equation for the extended model, together with its corresponding measurement counterpart are given by equations [16] and [17], below.

[16]
$$\underbrace{\begin{bmatrix} p_t \\ g_t \\ c_t \\ c_{t-1} \\ p_{1,t} \\ p_{2,t} \\ g_{2,t} \end{bmatrix}}_{S_t} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_1 & \phi_2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} p_{t-1} \\ g_{t-1} \\ c_{t-1} \\ c_{t-2} \\ p_{1,t-1} \\ p_{2,t-1} \\ g_{2,t-1} \end{bmatrix}}_{S_{t-1}} + \underbrace{\begin{bmatrix} v_t \\ w_t \\ e_t \\ 0 \\ v_{1,t} \\ v_{2,t} \\ w_{2,t} \end{bmatrix}}_{\zeta_t}$$

[17]
$$\underbrace{\begin{bmatrix} y_t \\ y_{1,t} \\ y_{2,t} \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 & 1 & 0 & 0 \\ 0 & 0 & \alpha_2 & 0 & 0 & 1 & 0 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} p_t \\ g_t \\ c_t \\ c_{t-1} \\ p_{1,t} \\ p_{2,t} \\ g_{2,t} \end{bmatrix}}_{S_t} + \begin{bmatrix} 0 \\ e_{1,t} \\ e_{2,t} \end{bmatrix}$$

Equation [18] represents the state space equations in compact form. The notation F(ϕ) and H(α) emphasizes the allocation of dynamical parameters (ϕ) and static parameters (α) in the transition equation and the measurement equation, respectively. The variance-covariance (VCV) matrices of the extended model are given by [19]. The parameters of the model can be put together in a single vector, as in [20].

[18]
$$\begin{aligned} S_t &= F(\phi)S_{t-1} + \zeta_t \\ Y_t &= H(\alpha)S_t + \varepsilon_t \end{aligned}$$

[19]
$$\begin{aligned} Q &= \mathrm{diag}\begin{pmatrix} v_v & v_w & v_e & 0 & v_{v_1} & v_{v_2} & v_{w_2} \end{pmatrix} \\ R &= \mathrm{diag}\begin{pmatrix} 0 & v_{e_1} & v_{e_2} \end{pmatrix} \end{aligned}$$

[20]
$$\theta = \begin{pmatrix} \phi_1 & \phi_2 & v_v & v_w & v_e & \alpha_1 & v_{v_1} & v_{e_1} & \alpha_2 & v_{v_2} & v_{w_2} & v_{e_2} \end{pmatrix}$$

The estimation algorithm is the same as in the univariate case, see box 1, once properly adapted to deal with the extended model [14]-[18].

2.3 Selection criteria

As mentioned before, the potential output of the economy cannot be measured directly, consequently there is no observable target or benchmark for comparison. This makes it difficult to evaluate alternative specifications.

To operationalize the optimality requirements specified previously, this section defines a set of criteria covering the relevant dimensions against which to gauge the different estimates. These criteria are split into two categories. First, the statistical-based ones define the necessary conditions. Second, the more economically and policy-oriented ones, underline the sufficient conditions.

Group 1, necessary conditions:

- Criteria 1: Statistical significance of the coefficients, focusing on the loadings of the observables on the cycle;
- Criteria 2: Average relative revision, defined as the average distance between one-sided and two-sided estimates, relative to the maximum amplitude of the output gap estimate;

- Criteria 3: Average relative uncertainty surrounding the cycle estimates, as the average standard error relative to the maximum amplitude.

Group 2, sufficient conditions:

- Criteria 4: Amplitude and profile alignment with consensus figures (range given by a panel of official institutions) and in agreement with commonly accepted business cycle chronology (e.g. ECRI dating);[7]
- Criteria 5: Stability of the one-sided cycle estimate, as this would mimic the practitioner's need for updated estimates as new data is added in real time.

2.4 Let the data speak: an application to Spain

Let us now turn to the implementation of the methodology above described to the Spanish economy. Alternative attempts are described in Doménech and Gómez (2006), Doménech *et al*. (2007) and Estrada *et al*. (2004). In particular, our approach is affine to the first one.

2.4.1 Data set and data processing

The selection of potential candidate variables follows an encompassing approach, aiming at capturing the build-up of potential imbalances across all relevant dimensions: (i) domestic economy; (ii) external sector; (iii) prices; (iv) labour market, and (v) financial and monetary conditions, as can be seen in Table 1. This set of indicators is easily replicable for different countries and at the same time encompassing enough to reflect a great variety of economic cycles.

In relation with data processing, all the variables have to be corrected from seasonal and calendar effects to get a signal free of possible distortive elements that helps to calculate more accurately the cyclical component of

---

[7] Economic Cycle Research Institute recession dating: https://www.businesscycle.com/ecri-business-cycles/international-business-cycle-dates-chronologies

the economy. In the case of the series from the Quarterly National Accounts, they are already published corrected of such effects. For the remaining time series, Tramo-Seats is used (Gómez and Maravall, 1996, Caporello and Maravall, 2004).[8]

Formally:

[21] $$x_{j,t} = V(B,F,\theta_{i,j})xr_{j,t}$$

where $xr_{j.t}$ is the raw indicator and $x_{j.t}$ the corrected indicator; V() is the Wiener-Kolmogorov filter symmetrically defined on the backward and forward operators B and F and $\theta_{i,j}$ are the parameters of the filter derived consistently with those of the ARIMA model for $xr_{j.t}$, see Gómez and Maravall (1998) for a detailed exposition of the model-based approach used by Tramo-Seats.

All series have been extended and/or completed until the first quarter of 1980, considering their specificities (sources, concepts, different statistical bases, mixed frequencies, etc.). The sample ends in 2016Q4.

Overall, the necessary processing could be summarized by backward linking retropolation and temporal disaggregation when needed.[9] Moreover, additional benchmarking techniques are implemented whenever the seasonal adjustment process breaks the temporal consistency with respect to the annual reference.

Finally, there are three main issues to set before performing the estimation of the different combinations: (i) the cyclical behavior of the selected variables, accompanying the GDP; (ii) their order of integration; and (iii) unit specification.

---

[8] The use of symmetric filters for seasonal adjustment introduces an additional source of revisions in the output gap estimates.

[9] Based on the most common procedures implemented by the National Accounts such as Fernandez (1981) Chow-Lin (1971) and Boot-Feibes-Lisman (1967).

Table 1. Data set

Variable	Unit	Source
GDP	Volume index (base 2010=100)	INE
Internal demand		
Investment, Construction	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Investment, Equipment	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Productive Capacity Utilization	%	MINETUR
External sector		
Real Effective Exchange Rate	Index 1999 I=100	Bank of Spain
Current Account Balance	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Gross National Savings	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Prices		
CPI, General	(i) Price index (base 2011=100); (ii) growth rate, % change	INE
GDP Deflator	(i) Price index (base 2010=100); (ii) growth rate, % change	INE
Compensation per employee	Euros per employee	INE
Housing prices	Euros per square meter	MFOM
Labour market		
Unemployment Rate	%	
Employment, full-time equivalent	Thousands	INE
Hours worked per employee	Units	INE
Compensation of employees	(i) Volume index (base 2010=100); (ii) M€;	INE
Financial and Monetary sector		
Credit to Non-Financial Corporations	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Credit to Households	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Broad Money (M3 aggregate)	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Narrow Money (M1 aggregate)	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Fiscal Variables		
Public Debt, Excessive Deficit Procedure	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	Bank of Spain
Net Lending (+), Net Borrowing (-): General Government	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Taxes on Production and Imports	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Taxes on Income and Wealth	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Social Contributions	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	INE
Unemployment Benefits	(i) Volume index (base 2010=100); (ii) M€; (iii) %GDP	MEYSS

* Total number of variables included: 52

Source of data: INE: National Statistics Institute; BDE: Bank of Spain; MFOM: Ministry of Public Works; MINETUR: Ministry of Industry, Energy and Tourism; MEYSS: Ministry of Employment and Social Security.

2.4.2 Selection results

The selection of the relevant variables follows a reductionist approach per the criteria specified above, starting with the necessary conditions. Every variable is modelled in a bivariate framework together with real GDP.

In the first place, the candidates not passing the significance test are removed, as can be seen in Table 2. Two sets of variables are left out in this first round, most labour market series and somewhat surprisingly, financial variables. Although highly intertwined in the latest crisis, financial and domestic demand variables tend to follow different cyclical patterns. Indeed, the literature has identified longer financial cycles, particularly as the deleveraging process of overindebted economies takes time and is still present after the economy is fully on track.

Table 2. Selected variables according to criteria 1 to 5

Variable	Transformation	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5
GDP		t-statistic	ARR	ARU	Profile	Stability
Internal demand						
Investment, Construction	Volume index (base 2010=100)	5.32	~~0.28~~			
	% GDP	8.88	0.10	0.38	YES	YES
Investment, Equipment	Volume index (base 2010=100)	6.30	0.24	~~0.41~~		
	% GDP	4.87	0.11	0.18	NO	
Productive Capacity Utilization	%	3.22	0.02	0.11	NO	
External sector						
Real Effective Exchange Rate	Index 1999 I=100	~~-0.66~~				
Current Account Balance	Volume index (base 2010=100)	~~0.00~~				
	% GDP	-6.98	0.13	0.34	YES	YES
Gross National Savings	Volume index (base 2010=100)	~~0.97~~				
	% GDP	~~-1.64~~				
Prices						
CPI, General	Price index (base 2011=100)	25.22	0.25	~~0.87~~		
	Growth rate, % change	~~0.03~~				
GDP Deflator	Price index (base 2011=100)	~~1.34~~				
	Growth rate, % change	~~0.02~~				
Housing prices	Euros per square meter	2.26	~~0.29~~			
Labour market						
Unemployment Rate	%	-7.59	0.06	0.23	YES	YES
Employment, full-time equivalent	Thousands	3.17	~~0.28~~			
Hours worked per employee	Units	~~-0.27~~				
Compensation per employee	Euros per employee	~~1.59~~				
Compensation of employees	Volume index (base 2010=100)	~~1.84~~				
	M€	~~1.71~~				
Financial and Monetary sector						
Credit to Non-Financial Corporations	Volume index (base 2010=100)	~~-0.14~~				
	M€	~~0.23~~				
	% GDP	~~-1.44~~				
Credit to Households	Volume index (base 2010=100)	~~-0.19~~				
	M€	~~0.43~~				
	% GDP	~~-1.58~~				
Broad Money (M3 aggregate)	M€	~~0.12~~				
	% GDP	5.43	0.13	~~1.54~~		
Narrow Money (M1 aggregate)	M€	~~-0.22~~				
	% GDP	~~-1.55~~				
Fiscal Variables						
Public Debt, Excessive Deficit Procedure	Volume index (base 2010=100)	-2.57	~~0.31~~			
	M€	-6.93	~~0.29~~			
	% GDP	-8.23	0.25	0.36	NO	
Net Lending (+), Net Borrowing (-): General Government	Volume index (base 2010=100)	~~0.00~~				
	M€	~~-0.01~~				
	% GDP	~~1.11~~				
Taxes on Production and Imports	Volume index (base 2010=100)	~~0.47~~				
	M€	1.92	0.20	~~0.92~~		
	% GDP	-3.95	0.10	0.07	NO	
Taxes on Income and Wealth	Volume index (base 2010=100)					
	M€	~~0.06~~				
	% GDP	2.14	0.11	0.12	NO	
Social Contributions	Volume index (base 2010=100)					
	M€	1.90	~~0.26~~			
	% GDP	-5.43	0.10	0.20	NO	
Unemployment Benefits	Volume index (base 2010=100)					
	M€	~~-0.75~~				
	% GDP	-8.84	0.04	0.18	NO	
Net Income	Volume index (base 2010=100)					
	M€	6.41	~~0.26~~			
	% GDP	-5.12	0.16	0.16	NO	

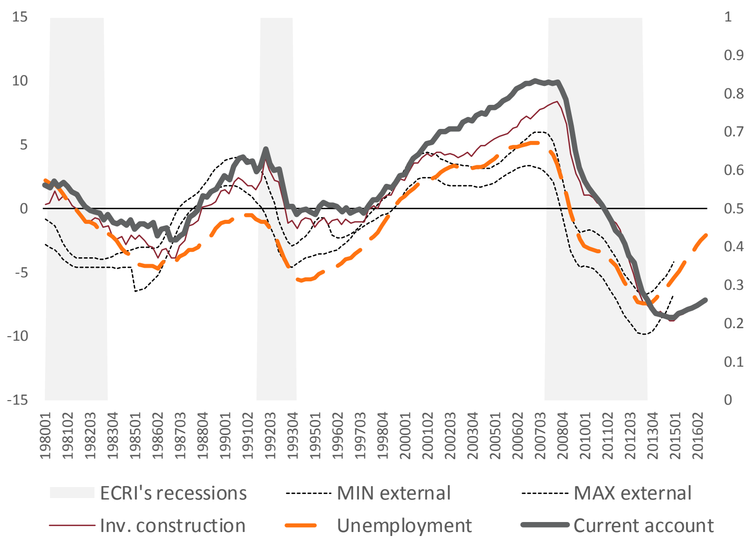
Source of data: author's estimations. Note: strikethrough text means the variables is discarded.

The average revision indicator provides the second screening for the remaining variables. This indicator reflects the average gap between the filtered (one-sided) and smoothed (two-sided) estimates of the output gap, normalized by the maximum range of the filtered estimation. Variables experimenting large revisions relative to their volatility are thus penalized (e.g. public debt, housing prices). The defining threshold is set at 0.25, to include two thirds of the remaining sample. Third, goodness of fit is assessed in relative terms as the ratio between the average standard error and the maximum range of the filtered estimate. Again, the threshold is set to keep two thirds of the competing variables (at 0.4). Prices and monetary variables are discarded at this stage as can be seen in Table 2.

Once the necessary conditions are checked out, the fourth criterion looks at the amplitude and profile of the output gap estimates. Small cycles, as defined by a small amplitude (lower than 4 pp.) are first left out. These include productive investment and most of the remaining fiscal variables (net income, social security contributions, direct and indirect taxes). A closer look at the specific profiles and ECRI dating allows for a further screening by removing unemployment benefits (as it does not properly identify the beginning of the last cycle) and capacity utilization (as it advances the recovery after the last cycle and points to positive output gap figures already in 2016).

Only three candidates made it all the way down to the fourth criteria: (i) the unemployment rate; (ii) the current account balance over GDP; and (iii) investment in construction over GDP. Figure 2. shows the three output gap estimates, together with the ECRI dating and the range of external institutions.

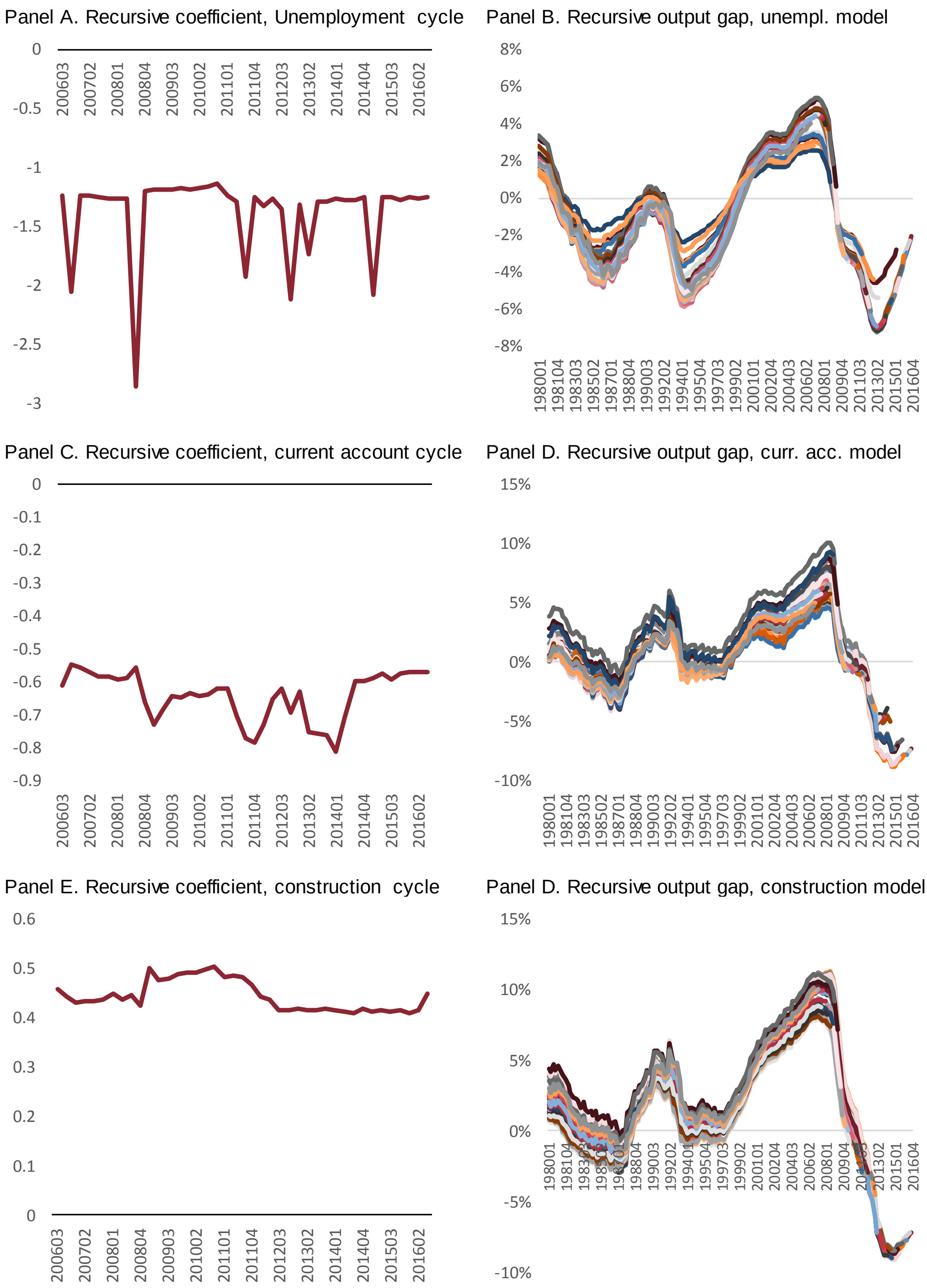
Figure 2. Selected output gap estimates



Source of data: author's estimations, IMF, OECD, European Commission, MINECO.

Finally, the stability of the estimates is assessed via a backward test covering the last 40 quarters, and results are obtained for the cyclical parameter as well as for the output gap estimates (see Figure 3).

Figure 3. Backtest, selected variables



Source of data: author's estimations.

As can be seen in the left-hand panels of Figure 3, parameter stability remains rather high, although with some discontinuities in the unemployment coefficient. This reassuring result would ensure robustness in the estimates as new data becomes available. This pseudo-real time exercise translates into updated output gap estimations as can be seen in the right-hand side of Figure 3. A general pattern emerges in all three cases as new data is added to the sample, the peak of the last cycle is revised upwards and the trough is equally revised downwards, thus amplifying the extent of the crisis and delaying the closure of the output gap. This is particularly relevant when it comes to the current account bivariate model, pointing towards structural gains associated with the latest current account developments.

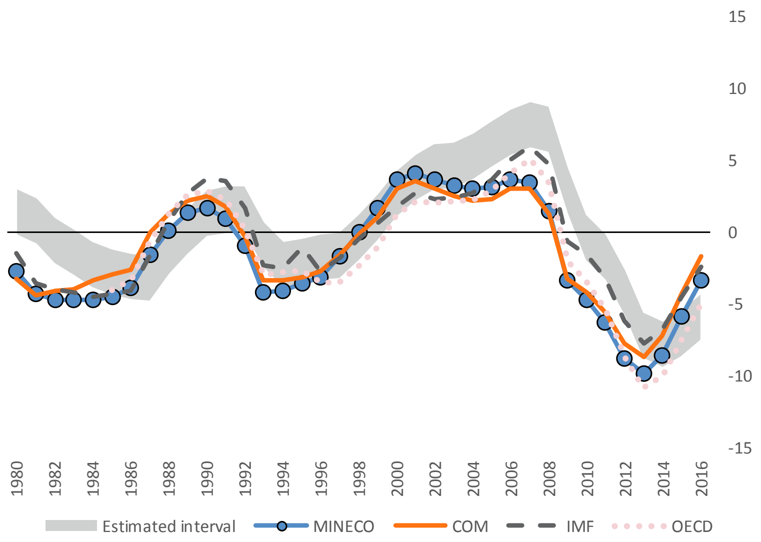
2.5 An estimate for Spain

The identification of the relevant imbalances for the definition and estimation of the cycle pints towards three clear candidates covering relevant areas of the Spanish economy: (i) the unemployment rate; (ii) the current account balance (% GDP); and (iii) the construction investment ratio (% of GDP).

The final step of the selection algorithm consists in summing up all the information gathered via the three bivariate models into an estimate of the Spanish output gap, ideally considering uncertainty considerations. This chapter opts for the simplest and more replicable option in these two dimensions. The central estimate is obtained from an unweighted average of the three bivariate models. The uncertainty surrounding the central estimate is equally derived from the unweighted average of the revision (as the gap between the concurrent and smoothed estimate) of the bivariate models.

The results are depicted in Figure 4 together with the estimates of the Spanish Ministry of Economy and the main international references; i.e. the OECD, the IMF and the European Commission. Both the profile and amplitude of the cycles over the last 35 years look rather coincident and the outside references tend to fall within the estimated confidence interval. The estimates tend to slightly differ, however, in their assessment of the last cycle. First the amplitude of the estimated gap is larger, as the confidence interval for the peak 2008 remains above the comparing estimations. At the same time, the profile of the upturn is slightly upward sloping, reflecting an increasing overheating until the beginning of the crisis. The end-of-the-sample figures also differ as the estimated interval points towards a larger slack in 2016.

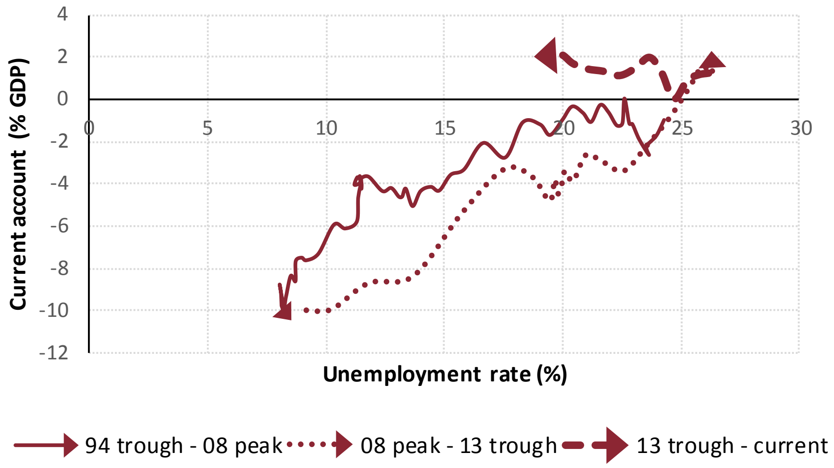
Figure 4. Output gap estimate and surrounding uncertainty



Source of data: author's estimations, IMF, OECD, European Commission, MINECO.

The interpretation of these differences represents another supporting argument in favor of the selected approach and estimation. On the one hand, it easily compares with two-sided estimations (such as the OECD reference), while keeping the benefits of being estimated only with concurrent information. On the other hand, the economic narrative also supports the interpretation of the current slack in the economy being larger than previously thought. The current growth pattern is proving to be resilient and balanced. Growth is more export-oriented and deleveraging in the private sector is co-existing with a robust productive investment and strong employment creation without generating inflationary or wage pressures. The correction of the macro imbalances has a significant structural component. As can be seen in Figure 5, unemployment is cut back with historically low real growth figures, while this has not generated additional imbalances or tensions in terms of current account deficit or construction sector investment.

Figure 5. Selected variables over the last cycle

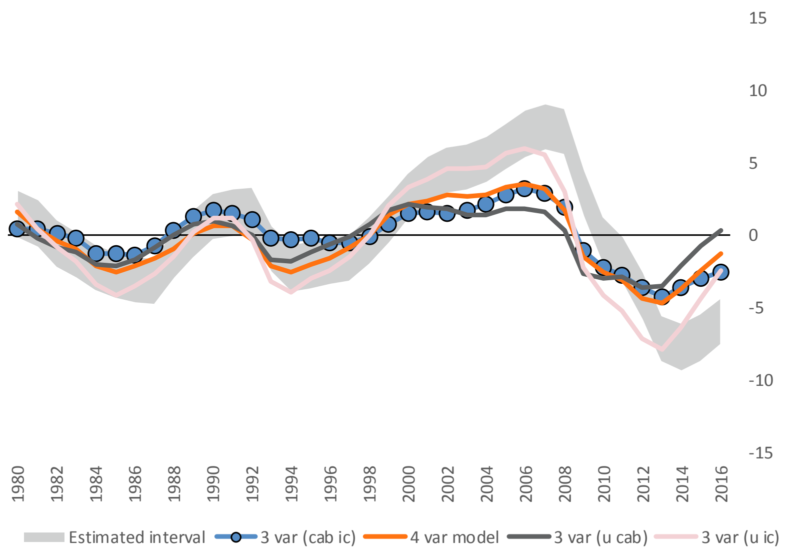


Source of data: INE, Bank of Spain.

2.5.1 Robustness check: alternative combinations

Flexibility is at the core of the proposed modelling approach. The existing set-up allows for an incremental approximation as more variables are added to the estimated model. Taking advantage of this possibility, all the 3 and 4 variable model combinations are estimated and gauges against our preferred alternative.[10]

Figure 6. Output gap estimates, enhanced models



Source of data: author's estimations.

Figure 6 shows how the diagnosis resembles the previous comparison with external estimations. The main divergences arise when considering the last cycle. Indeed, enlarging the model undeniably shrinks the output gap amplitude and provides a significantly smaller estimation of the existing slack. Although this fact deserves further research, our experience with the econometric procedure used in this chapter reveals a tendency of the loadings to become smaller in absolute value with respect of the loadings to the

[10] The full model, with four variables (GDP, unemployment (u), current account balance (cab) and construction investment (ic)), together with 3 different trivariate models: (i) GDP, unemployment and current account balance; (ii) GDP, unemployment and construction investment; and (iii) GDP, current account balance and construction investment.

corresponding bivariate model when the number of variables is expanded. This reduction in the size of the loadings is mirrored by a less volatile estimate of the output gap. Our view is that the numerical optimization that generates the maximum likelihood estimates operates as in a regression setting, considering the correlation among the variables.[11]

[11] The correlation among them is due to their sharing of a common cycle linked to the output gap) as multicollinearity, then reducing the individual loadings.

2.6 Conclusion

Over the last decades, the estimation of the slack in the economy has become an essential piece of analysis for policymakers, both on the monetary and the fiscal policy side. Output gap estimation techniques have flourished accordingly, although there is no consensus on a best-performing methodology, as the selection criteria often imply important trade-offs.

This chapter presents a novel approach putting the focus on the specification of the model rather than on the prior selection of the methodology itself. Ideally, an agreeable method should achieve three necessary conditions: economic soundness, statistical goodness and transparency. On top of this, a sufficient condition is given by the *smell test*, often implemented by policymakers. In practice, fulfilling these conditions can prove to be challenging.

Multivariate methods, coupled with Kalman filtering are generally considered amongst those reaching an acceptable level of compromise between these dimensions and thus are selected as a starting point, allowing for a combination of an economically-sound specification with a well-tested and flexible econometric procedure. The method fulfils the necessary criteria and allows for enough flexibility to get a country-specific approximation to the sufficient (smell test) criteria as it could accommodate specific cycles (financial, external, investment, fiscal, etc.)

This somewhat eclectic approach is illustrated with its application to a data set for the Spanish economy, by selecting the best model amongst bivariate combinations of GDP and 52 accompanying variables.

Some preliminary conclusions can be drawn at this stage. First, there are some technical aspects that are important to be taken care of before jumping into the selection of the variables specification, such as: (i) modeling of GDP; (ii) cyclical prior of the accompanying series; (iii) transformation of the series (nominal vs. real, ratios vs. logs, etc.). Second, there is no clear algorithm for the selection of the variables to be included in the final specification. Should it be an incrementalistic approach or rather a *brute force* consideration of all the alternative combinations? Third, this chapter has opted for the definition of necessary vs. sufficient conditions, although other combinations or weighting of the criteria might be possible.

Finally, future extensions of this work include an attempt at answering some of these open questions and providing a full assessment of the methodology in more complex data environments as well as technical improvements adding to the existing selection criteria, for example by estimating the contribution of the observables to the estimation of the output gap.

2.7 REFERENCES

Álvarez, L.J., Gómez-Loscos, A. (2017). "A menu on output gap estimation methods," Bank of Spain, Working Paper n. 1720.

Alberola, E., Estrada, A. and Santabarbara, D. (2013). "Growth beyond imbalances. sustainable growth rates and output gap reassessment," Bank of Spain, Working Paper n. 1313.

Alichi, A. (2015). "A new methodology for estimating the output gap in the United Sates," IMF Working Paper Series, Working Paper n. 144.

Boot, J.C.G., Feibes, W. and Lisman, J.H.C. (1967). "Further methods of derivation of quarterly figures from annual data," *Applied Statistics*, vol. 16(1), 65-75.

Borio, C., Disyatat, P. and Juselius, M. (2017), "Rethinking potential output: Embedding information about the financial cycle," Oxford Economic Papers 69(3), 655-677.

Caporello, G. and Maravall, A. (2004). "Program TSW. Revised manual," Bank of Spain, Occasional Paper n. 0408.

Chow, G. and Lin, A.L. (1971). "Best linear unbiased distribution and extrapolation of economic time series by related series," *Review of Economic and Statistics*, vol. 53(4), 372-375.

Clark. T. (1987). "The cyclical component of U.S. economic activity," *Quarterly Journal of Economics*, vol. 102(4), 797-814.

Cotis, J.P., Elmeskov, J. and Mourougane, A. (2005). "Estimates of potential output: Benefit and pitfalls from a policy perspective," in L. Rechling (ed) Euro area business cycle: stylized facts and measurement issues, CEPR London.

Cuevas, A. and Quilis E. (2017). "Non-linear modelling of the Spanish GDP," AIReF Working Paper, n. 01.

Doménech, R. and Gómez, V. (2006). "Estimating potential output, core inflation, and the NAIRU as latent variables," *Journal of Business and Economic Statistics*, vol. 24(3), 354-365.

Domenech R., Estrada, A. and González-Calbet, L. (2007). "Potential growth and business cycle in the Spanish economy: implications for fiscal policy," Working Paper n. 05, International Economics Institute, University of Valencia.

Durbin, J. and Koopman, S.J. (2001). "Time Series Analysis by State Space Methods," Oxford University Press.

Estrada, A., Hernández de Cos, P. and Jareño J. (2004). "Una estimación del crecimiento potencial de la economía española," Documentos Ocasionales, n. 0405 Banco de España.

Fernández, R.B. (1981). "Methodological note on the estimation of time series", *Review of Economic and Statistics*, vol. 63(3), 471-478.

Gómez, V. and Maravall, A. (1996). "Programs TRAMO (Time Series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instruction for the User," Bank of Spain Working Paper n. 28.

Gómez, V. and Maravall, A. (1998). "Guide for Using the Programs TRAMO and SEATS," Bank of Spain Working Paper n. 05.

Harvey, A.C. (1989). "Forecasting, Structural Time Series Models and the Kalman Filter," Cambridge University Press.

Press, Cambridge.

Kim C.J., and Nelson C.R. (1999). "State space models with regime switching," MIT Press, Cambridge, U.S.A.

Kitagawa, G. and Gersch, W. (1996). "Smoothness Priors Analysis of Time Series," Springer, Berlin.

Kuttner, K. N. (1994). "Estimating potential output as a latent variable," *Journal of Business and Economic Statistics*, 12(3), 361-368.

Murray, J. (2014). "Output gap measurement: Judgement and uncertainty,", U.K. Office for Budget Responsibility, Working Paper n. 5.

3 FORECASTING ALONG THE BUSINESS CYCLE

3.1 Introduction

The ability of economists to forecast the main aggregate macroeconomic variables has undergone a complete revolution over the last 40 years. The stagflation period that affected the developed economies during the late 70s, along with theoretical dissatisfactions with Keynesian principals shaped a new way of understanding forecasting and led to the pioneering contributions of Sargent and Sims (1977) and Sims (1980): non-structural models that minimized their theoretical roots. Non-structural models were not subject to changes in the dominant theoretical paradigm and could also escape the Lucas (1976) critique, as their forecasts were not tied to a specific path of the policy variables. Pivoting on Box and Jenkins (1970) contributions and Sims (1980) multivariate extension, Vector Autoregressive (VAR) models became the workhorse in macroeconomic forecasting. Indeed, Unrestricted VAR (UVAR) models do not impose excessive identification restrictions, leading to a better in-sample fit.

However, good in-sample fit did not grant a good out-of-sample forecasting performance, as indicated in Stock and Watson (1996) work. To avoid omitted variables biases and allow for the identification of structural shocks through the model innovations, researchers could have a tendency towards increasing the number of variables in the analysis. This strategy would generate inaccurate estimates and bad predictive results due to over-fitting, as the number of parameters to estimate increases with the square of the number of variables included in the model. VAR modeling limitations led thus to the parallel development of two major lines of research:

- The restriction of the parameter space by imposing Bayesian constraints through a priori information. The first Bayesian VAR (BVAR) models were based on purely statistical beliefs. For example, the famous ”Minnesota prior”, developed by Doan, Litterman and Sims (1984) and Litterman (1986), restricts the higher lags coefficients to near to zero values. The development of Monte Carlo type algorithms, such as the Gibbs-Sampler, developed by Geman and Geman (1984) or the Metropolis *et al*. (1953) algorithm, generalized in Hastings (1970), as well as their application to the field of economics have positioned BVAR models as benchmarks in forecasting major macroeconomic aggregates, as indicated in Zarnowitz and Braun (1993). For example, see Kinal and Ratner (1986) BVAR application to forecast New York data, extended by Sims (1992) for the U.S. economy or Amisano and Serati (2002) for the euro area.

- Sargent and Sims (1977) findings that a few shocks can explain most of the dynamics of macroeconomic aggregates set the scenario for Dynamic Factor Models (DFM). Following this evidence, Geweke (1977) assumes that all time series can be expressed as the sum of two orthogonal components: a common one, which captures the dynamic shared by all series, and an idiosyncratic component, understood as a residual. Information technologies developments and the availability of real-time data have facilitated the exploitation of DFM’s potential, as stated in Forni *et al*. (2005a), Forni *et al*. (2009), and Stock and Watson (2002), among others, where the orthogonality assumption among the idiosyncratic components is relaxed and the number of series considered in the models can be very large.

Economic theory would not lag for a long time. Dissatisfaction with the theoretical basis of non-structural models and the need to generate forecasts conditional on economic policy as a guide to policymakers, stimulated the birth of structural models, explicitly grounded in the optimizing behavior of economic agents. The so-called new classical macroeconomics school came to acknowledge the need for micro-founded, dynamic, stochastic macroeconomic models that could escape from the Lucas critique in policy guiding.

The contributions by Hansen and Sargent (1980) and Kydland and Prescott (1982) started the literature of Dynamic Stochastic General Equilibrium (DSGE) models. They provided a connection between theory and data thanks to the state space representation of the decision rules obtained from the solution of the models. Since then, the parallel evolution of macroeconomic theory towards a widespread paradigm, the New Neoclassical Synthesis (see Goodfriend, 2004 for an introduction or Gali and Gertler, 2007 for a historical overview) and econometric techniques that ease the estimation of large-scale DSGE models, especially Bayesian econometrics techniques (see Greenberg ,2007 for an introduction and Geweke, 2005 and Canova, 2007 for a deeper discussion) have provided useful tools and DSGE models have been adopted by many Central Banks as they serve a great variety of purposes such as policy guiding, historical analysis and counterfactual experiments.

However, as discussed in Sims (2002a), forecasting exercises have traditionally been backed mainly by large-scale macro-econometric models and expert judgment analysis, without an explicit theoretical structure or a consistent treatment of expectations. Interestingly, since the contribution of Smets and Wouters (2004), economists have begun to look seriously at structural DSGE models as effective tools in forecasting. These authors capture the statistical features of the main aggregate variables by applying Bayesian estimation techniques to large scale models. Ever since, the forecasting performance of DSGEs has been tested against traditional UVAR, BVAR benchmarks (see Smets and Wouters, 2005 for a closed economy application and Adolfson *et al.*, 2007 or Christofel *et al.*, 2010) for an open economy case), more sophisticated set-ups such as dynamic factor models *à la* Stock and Watson (2002) (see Wang, 2009) and even against experts judgment with real-time data (for example in Adolfson *et al.*, 2007, Edge, Kiley and LaForte, 2009, Kolasa *et al.*, 2009 and Rubaszek and Skrzypczynski, 2008).

Following this literature, DSGE forecasting ability is comparable to that of the competing models, especially in the medium to long-term horizon, as suggested in Monti (2008) and Wang (2009). These results have inspired new estimation and forecasting practices to combine the best of the various methodologies, and minimize the issues resulting from poorly specified models.

Generically, the solution of a DSGE model can be cast into a VAR representation, which is used to assess and validate the DSGE empirically.

Ingram and Witheman (1994) goes one step further and builds a BVAR model whose priors consist of a real business cycle model, away from the traditional statistical Minnesota prior, obtaining accuracy gains in out-of-sample forecasts of the main US aggregates. In this line, Del Negro and Schorfheide (2003, 2004) and Del negro *et al*. (2004) use a DSGE model to build the prior density function for a VAR (DSGE-VAR) and develop an estimation procedure to obtain the posterior distribution of the DSGE structural parameters by minimizing the divergence between the UVAR estimates and those from the VAR representation of the DSGE (Kullback-Leibler divergence). An application of this methodology can be found in Hodge *et al.* (2008) for the Australian economy. These authors have obtained competitive results in the prediction of GDP and inflation, both against non-structural models as well as purely theoretical models. Waggoner and Zha (2012) further extend the methodology through a combination between a BVAR and a DSGE model, assigning state-dependent probabilities to the two models (to their likelihood functions).[12]

Giannone *et al.* (2006) criticize, however, the VAR representation of the DSGE model, as the presence of measurement errors in observed variables would contaminate the inference results, especially in the short term. Moreover, they find that observed variables usually follow a factor-type structure. They therefore decompose the spectral density of the model into two components, a common and an idiosyncratic one. The DSGE representation through a factor structure leads to higher inference accuracy and to the ability of handling high-frequency data in real-time. In this line, in a pioneering article, Boivin and Giannoni (2006) relax the assumption that DSGE theoretical concepts are measured adequately by a single series and estimate a structural model with a wide range of data, relating the factors of the resulting system with the endogenous variables of the DSGE model. Following this approach, Baeurle (2008) finds significant forecasting accuracy improvements at all time horizons. Schorfheide *et al.* (2010), simplify the Boivin and Giannoni (2006) approach by introducing a two steps estimation

[12] The Federal Reserve Bank of New York launched in 2015 an innovative experiment by publishing its DSGE forecasts (see del Negro *et al*., 2017).

process which creates a link between the DSGE model variables and some other non-modeled variables that can therefore be forecast indirectly through auxiliary regressions.

The bulk of the literature restricts its attention to economies that are linearized around a steady state or long-run equilibrium, yielding approximate decision rules and likelihood functions, from which inferences and forecasts are constructed. The work of Fernandez-Villaverde and Rubio-Ramírez (2005 and 2007) noted the importance of quadratic terms in the approximation to decision functions, since second-order errors may have first order effects on the likelihood function. The consequences of non-linearities in terms of forecasting have been evaluated by Pichler (2007), which assesses the trade-off between the sampling error introduced by the non-linear filters and the error due to the linear approximation of the model, which turns out to be more impeding in terms of out-of-sample forecasting.

Moreover, another related field of research seeks to incorporate the know-how or experts views into the DSGE framework. Monti (2008) and Giannone *et al.* (2009) successfully integrate the soft information or expert judgments of high frequency information in real time in a DSGE model, obtaining GDP forecasting gains, as monthly data is incorporated into the model.

Despite the growing literature on the predictive capabilities of DSGE models and their variants, there are a number of weaknesses, unexplored issues or inconclusive results. First, so far, the clear majority of forecast comparison exercises have focused on stable business cycle periods (except for Waggoner and Zha, 2012). There is a consensus regarding the misbehavior of model forecasts during the latest financial crisis, which has revealed the need to rethink the academic agenda. It is obvious that any economic disruption with respect to the past implies a great challenge in out-of-sample forecasting exercises. However, to what extent will these special circumstances affect the different models?. Second, the literature covering DSGE models is generally biased towards theoretical improvements and it is worth checking if forecasting gains follow these refinements. Enriched models allow a finer description of the existing transmission channels but might not prove to improve misspecified models' forecast accuracy.

Sticking to a linearized environment, this chapter conducts a comparative analysis of the forecasting performance of structural versus non-structural models, both in a smooth-growth period, both pre-crisis [2003Q1-2007Q3] and post-crisis [2012Q1-2016Q4] and during a crisis [2007Q3-2011Q4] to overcome the first two issues. Robustness is assessed by applying the exercise to Spanish, euro area 16 as well as to US data. Our sample covers the period 1980Q1-2016Q4. The forecasting performance will be assessed with a recursive procedure, checking the out-of-sample RMSE from 1 to 8 periods.

Section 2 covers the main methodological issues while section 3 specifies the data collection and treatment as well as the model estimations results[13]. Section 4 presents the results and main findings. Section 5 conducts an additional check by looking at the post-crisis period and finally, section 6 concludes.

3.2 Methodology

3.2.1 Non-Structural Models

3.2.1.1 Reference models

To introduce the notation and establish the framework for comparison of the DSGE and DFM models, this section first defines UVAR and BVAR models.

Vector Autoregressive Models

All time series are generated from the linear combination of three elements: their own lagged values (dynamics), lagged values of the other variables (cross-dynamics) and innovations or specific shocks. A simplified system of order 1 could be specified in the following form:

$$\begin{bmatrix} z_{1,t} \\ z_{2,t} \\ \vdots \\ z_{k,t} \end{bmatrix} = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots & \varphi_{1,k} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots & \varphi_{2,k} \\ \vdots & \vdots & \ldots & \vdots \\ \varphi_{k,1} & \varphi_{k,2} & \cdots & \varphi_{2,1} \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \\ \vdots \\ z_{k,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{k,t} \end{bmatrix}, \qquad [1]$$

[13] Detailed results on the model estimations and tests undertaken are available upon request.

where $u_t = [u_{1,t} \cdots \quad u_{k,t}]' \sim N(0, \Sigma_u)$. Furthermore, to implement a univariate analysis a diagonal structure is imposed on the matrix of parameters as well as on the variance covariance (VCV) matrix.

To allow for the estimation of the $z_t = [z_{1,t} \cdots \quad z_{k,t}]'$system via Kalman filtering techniques, a "State Space" structure is imposed on it. By further generalizing for *p* lags, the transition equation follows,

$$\begin{bmatrix} z_t \\ z_{t-1} \\ \cdots \\ z_{t-p+1} \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_{p-1} & \varphi_p \\ I & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ z_{t-2} \\ \cdots \\ z_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \qquad [2]$$

where now and φ_i is the kxk matrix of coefficients that relates z_t to its i-th lag, being i = 1, ..., p. In matrix notation, this transition equation describes the dynamics of the state vector Z_t,

$$Z_t = \Gamma Z_{t-1} + U_t, \qquad [3]$$

where $Z_t = \left[z'_{\ t} \quad z'_{\ t-1} \cdots \quad z'_{\ t-p}\right]'$, $U_t = [u'_{\ t} \quad 0' \dots \quad 0']'$and Γ is defined in an obvious way from [2].

By adding the corresponding measurement equation,

$$Y_t = \mathrm{H} Z_t + e_t, \qquad [4]$$

with Y_t being the observables vector and assuming zero mean Gaussian errors and both noises U_t and e_t uncorrelated for all lags, the state space representation is complete and ready for estimation via the Kalman filter, which will evaluate its Likelihood Function (LF) and can therefore be used to estimate the unknown parameters $\zeta = (P \quad H \quad \Sigma_u \quad \Sigma_e)$ *via* maximum likelihood algorithms. Notice that, to estimate a pure VAR(p) model, $e_t = 0_{np,1}$ and $H = [I \quad 0 \quad 0]$, while adding measurement noise in equation (4) imposes a VARMA structure for the observed Y_t series.

As seen in the introduction, UVAR models can lead to overfitting problems that directly influence the forecasting results. To minimize this issue, this Bayesian estimation methods can be implemented to VAR models.

Bayesian VAR models

The Bayesian approach allows us to combine beliefs with observed historical data to reduce UVAR dimensionality problems. By using a priori information (either statistical or economic), an initial value is provided for the parameters while specifying its trust on it at the same time, and observed data is then used as a device to review the prior beliefs.

The ”Minnesota prior” developed by Doan, Litterman and Sims (1984) and Litterman (1986), is taken as a reference point for the specification of the mean vector of the parameters. That is, all the variables of the system will follow a random walk with drift. However, as the original Minnesota prior was designed for non-stationary data, the prior was adapted following Lütkepohl (2005) whenever stationary series were used. The vector Z_t is made up of k variables and their corresponding p lags. The β coefficients associated to the mean can be cast as $\beta = vec(\varphi)$, with $\varphi = (\mu_t \quad \varphi_1 \quad \cdots \quad \varphi_p)$. The ”Litterman prior” will restrict β mean and variance covariance matrix $\beta \sim N(\beta^*, \mathrm{V}_\beta)$.

The prior density of the first moment of the coefficients will be equal to one for the first lag of all the coefficients and zero otherwise. The prior density of the second moment of the constant is considered as diffuse, leaving its estimation to the data. As for the rest of the coefficients of the variance covariance matrix, their `a priori will depend on a vector of hyperparameters $\Pi = (\pi_1 \quad \pi_2 \quad \pi_3)$, which will set three dimensions; the general dynamics, *g* that can follow a harmonic or a geometric decay process ruled by π_3, the first order own dynamics, with π_1 representing the trust on the prior density over the mean and finally the cross dynamics, where the out-of-sample specification of the dynamic interaction between the series will depend on π_2.

3.2.2 Dynamic Factor Models

Following Peña and Poncela (2004), the of observable series Y_t are defined as an N-dimensional vector. Every time series can be written as a linear combination of r factors capturing the common dynamics and m specific components:

$$Y_t = \mathrm{P} f_t + n_t, \qquad [5]$$

with f_t being a vector of common factors, of dimension $r\times 1$, P their loading matrix and n_t an $N\times 1$ vector of specific components. The common factors follow a VAR representation:

$$\Phi(B) f_t = a_t, \qquad [6]$$

where $\Phi(B) = I - \Phi(1)B - \cdots - \Phi(p)B^p$ is an $r \times r$ polynomial matrix, B is the lag operator ($By_t = y_{t-1}$) and $a_t \sim N_r(0, \Sigma_a)$ and is serially uncorrelated. The model can be written in state space form as in the case of the VAR (see equations 3 and 4), with the $s \times 1$ state vector containing the common factors and their lags.

Forecasting is also done by applying Kalman equations to get first the state vector forecast h periods ahead and its VCV matrix and then use the measurement equation to get the observables forecasts altogether with their second moment.

3.2.3 Structural Models

3.2.3.1 DSGE Models

Once approximated by a log-linearization process around the steady state, DSGEs solution gives the laws of motion that can be cast into a state space form as the transition equation:

$$Z_t = f(Z_{t-1}, W_t; \theta), \qquad [7]$$

where Z_t represents endogenous variables, W_t is the innovations vector and θ includes all structural parameters. To fulfil this process, Sims (2002b) algorithm is used.[14] This method requires a specific initial matrix representation of the model, such as

$$\Gamma_0 Z_t = \Gamma_1 Z_{t-1} + C + \psi V_t + \Pi \eta_t, \qquad [8]$$

where V_t represents the structural innovations, while η_t introduces expectational errors. Its reduced-from representation is given by:

$$Z_t = \Theta_c + \Theta_0 Z_{t-1} + \Theta_1 W_t, \qquad [9]$$

where the Θ_c and Θ_0 and Θ_1 matrices depend on the structural parameters and summarize the dynamic behavior of the model, with a generic representation following equation (6). Moreover, the model estimation needs a measurement equation to complete the state space representation. Again,

---

[14] Which MATLAB version *gensys.m* is available at his personal website, http://www.princeton.edu/sims/.

the observed variable Y_t will be a linear function of the state variables, with measurement errors e_t as in equation (4). If perturbations are assumed Gaussian white noises, the Kalman Filter will evaluate the LF of our model, as for DFM and VAR models, by specifying the parameter vector to estimate ζ.

The model is estimated using Bayesian techniques. The posterior density is made up of two components, the LF and a prior distribution over the structural parameters,

$$\mathrm{g}(\zeta|Y_t) = \mathrm{L}(Y_t|\zeta)\, g(\zeta), \qquad [10]$$

and is obtained by Metropolis-Hastings numerical approximation methods, more specifically the Random Walk Metropolis (RWM) algorithm[15].

3.2.3.2 DSGE-VAR Models

Following Del Negro and Schorfheide (2004), prior information from Smets and Wouters (2005) is incorporated into to a VAR representation to proceed with the Bayesian estimation, by minimizing the divergence between an UVAR representation of the observable variables and the DSGE-VAR model [Kullback-Leibler discrepancy]. The *a priori* density function will have a hierarchical structure, as the DSGE model depends on unknown structural parameters. The density function will therefore be given by a marginal distribution of the θ parameters and a conditional distribution of the VAR (φ, Σ) parameters, given θ,

$$\mathrm{p}(\varphi, \Sigma, \theta) = \mathrm{p}(\varphi, \Sigma|\theta)\, p(\theta), \qquad [11]$$

Symmetrically, the joint posterior density function will depend on the posterior density of the VAR parameters and the marginal posterior density of the θ,

$$\mathrm{p}(\varphi, \Sigma, \theta|Z) = \mathrm{p}(\varphi, \Sigma|Z, \theta)\, p(\theta|Z). \qquad [12]$$

All in all, the DSGE-VAR representation implies many restrictions on the VAR parameters. To allow for possible misspecification problems, a new parameter

[15] See the Dynare website (http://www.dynare.org) for a more detailed explanation on the RWM approximation method used.

λ rescales the VCV of the VAR and assesses the weight of the structural versus the non-structural component.

3.3 Data and estimation results

Both structural and non-structural models are estimated for Spain, the euro area and the US, using seven macroeconomic aggregates: Gross Domestic Product, GDP (Y), Private Consumption (C), Private Investment (I), Employment (L) or Total Hours worked (H), GDP deflator (Pr), Wages (W) and the interest rate (i). The data sample covers the period ranging from 1980:Q1 to 2016:Q4, thus including the latest subprime crisis and the post-crisis period. GDP and its components as well as wages are expressed in real terms and then divided by a population index to obtain per capita variables.

More specifically, for the US, in order to remain close to the original Smets and Wouters (2005) analysis, Real Gross Domestic Product, Nominal Personal Consumption Expenditures, Fixed Private Domestic Investment and the Implicit price deflator of GDP come from the US Department of Commerce - Bureau of Economic Analysis databank (BEA) while the Index of average weekly hours for the Nonfarm Business sector (NFB) and their Real Hourly compensation are taken from the Bureau of Labour Statistics (BLS). Moreover, hours are adjusted with a civilian employment index to consider the limited coverage of the NFB sector.

Table 3 and its graphical representation (figure 7) present more detailed information and the specific transformations of the US series for structural as well as non-structural models.

Table 3. US data, basic information

Acronym	Units	Transformation	Source
Y	B. of dollars	$\Delta(100*\ln(Y))$	BEA databank
C	B. of dollars	$\Delta(100*\ln(C/Pr))$	BEA databank
I	B. of dollars	$\Delta(100*\ln(I/Pr))$	BEA databank
H	Index	$\Delta(100*\ln(H))$	BLS databank
Pr	Index	$\Delta(100*\ln(Pr))$	BEA databank
W	Index	$\Delta(100*\ln(W/Pr))$	BLS databank
i[a]	Percentage	i/4	Federal

[a] The short-term interest rate is the Federal Funds rate

For the euro area, all time series are taken from the latest update of the Area Wide Model (AWM) database at the ECB, originally developed in *Fagan et al.* (2001).[16] Private Consumption and Total Gross Investment are deflated with their own deflator. According to Smets and Wouters (2005), total employment data was used due to the lack of availability on hours worked. To compensate for this in the estimation of the DSGE model, an auxiliary equation will link labour services with observed employment following a hiring mechanism à la Calvo, as suggested in Smets and Wouters (2003).

Table 4 and its corresponding graph (figure 8) present more detailed information and the specific transformations of the series for structural as well as non-structural euro area models.

Table 4. Euro area data, basic information

Acronym	Units	Transformation	Source	
Y	M. of ECU/euro	Δ(100*ln(Y))	QNA Eurostat	
C	M. of ECU/euro	Δ(100*ln(C))	QNA Eurostat	
I	M. of ECU/euro	Δ(100*ln(I))	QNA Eurostat	
H	Thousands	Δ(100*ln(H))	ECB	Monthly
Pr	Index	Δ(100*ln(Pr))	ECB	Monthly
W	M. of ECU/euro	Δ(100*ln(W/Pr))	ECB	Monthly
i[a]	Percentage	i/4	ECB	Monthly

[a] Three months interest rate

Lastly, data for the Spanish economy comes mainly from the National Statistics Institute (NSI) Quarterly National Accounts (QNA) except for the interest rate, provided by the Bank of Spain. Data on employment full-time equivalents require an auxiliary equation to the DSGE model, following the euro area specification. Table 5 and its accompanying graph (figure 9) provide a full description and representation of the Spanish data.

[16] The sample for the euro area ends in 2015q4 as the AWM database presents an annual update every month of September.

Table 5. Spain data, basic information

Acronym	Units	Transformation	Source
Y	M. of euro (2010)	Δ(100*ln(Y))	QNA INE
C	M. of euro (2010)	Δ(100*ln(C))	QNA INE
I	M. of euro (2010)	Δ(100*ln(I))	QNA INE
H	Thousands	Δ(100*ln(H))	QNA INE
Pr	Index	Δ(100*ln(Pr))	QNA INE
W	M. of euro	Δ(100*ln(W/Pr))	QNA INE
i[a]	Percentage	i/4	Banco de

[a] Non-transferable three month deposits

As can be seen in figures 7 to 9, stationarity is an issue for the employment, inflation and interest rate series. Their differenced versions are represented in dashed lines and used in the estimation of the non-structural models to obtain the best performing one in terms of forecasting accuracy[17]. Moreover, as the original Minnesota prior was designed for non-stationary data, the code was adapted following a procedure similar to Lütkepohl (2005) whenever stationary series were used. Four lags were selected for the VAR estimation following standard criteria and they were kept also for the Bayesian counterpart for consistency reasons.

[17] Detailed descriptions of the data and its transformations are available upon request.

Figure 7. US series

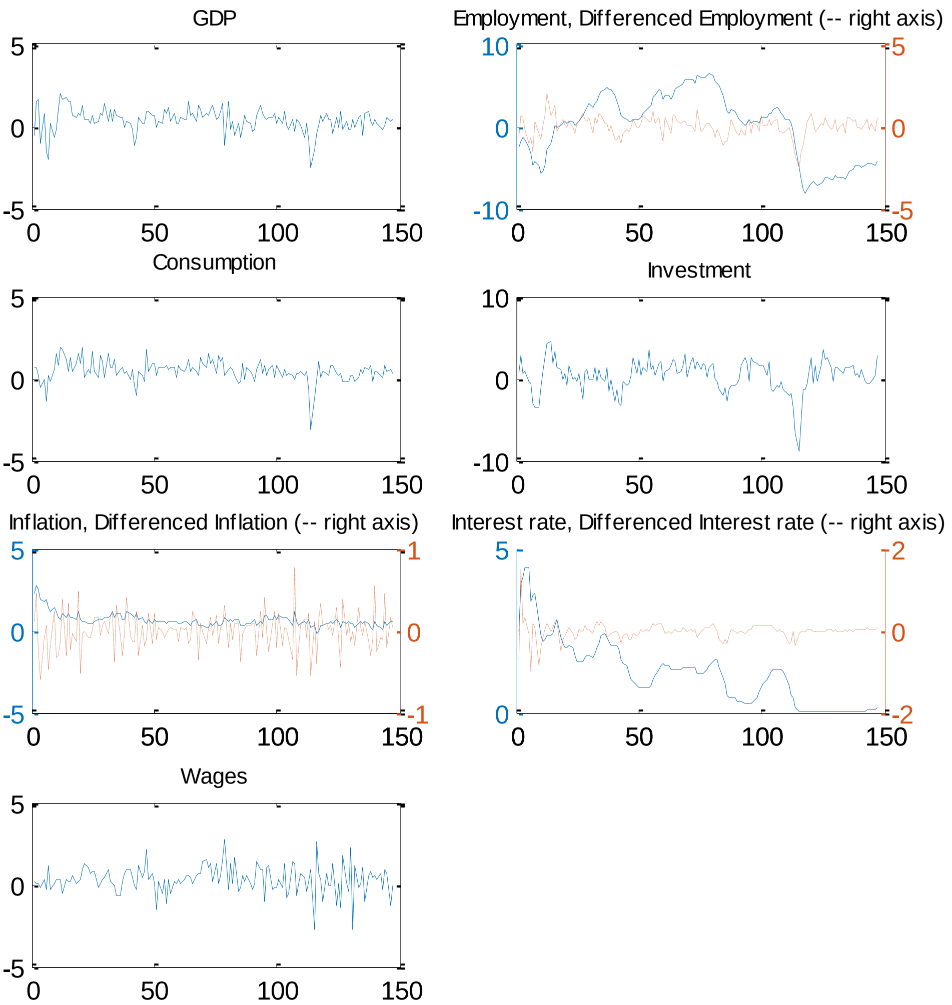


Figure 8. Euro area series

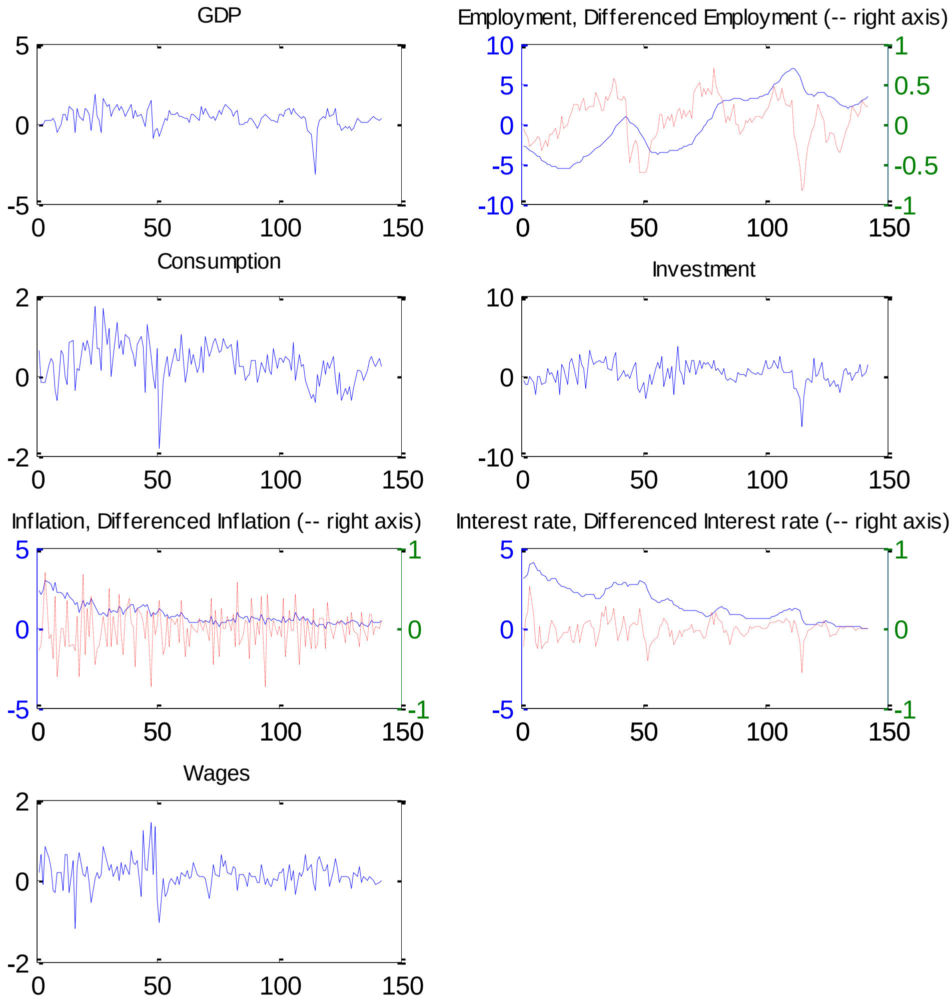
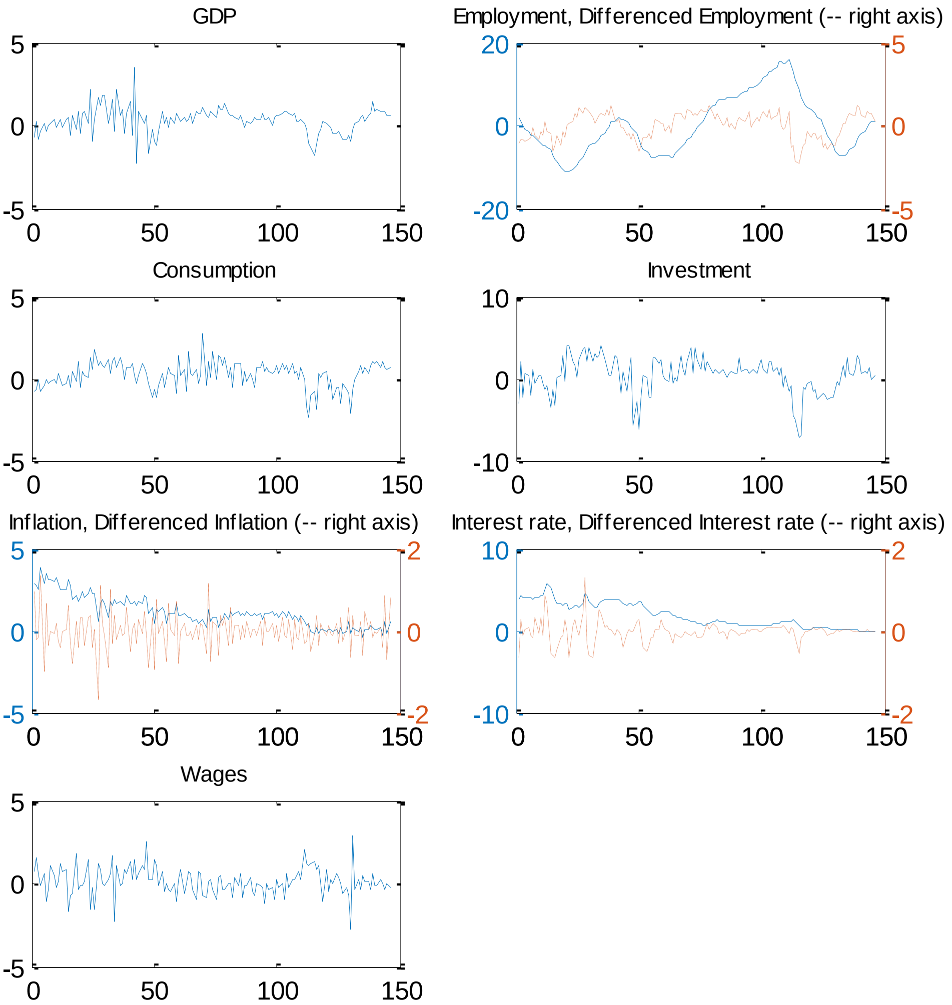


Figure 9. Spain series

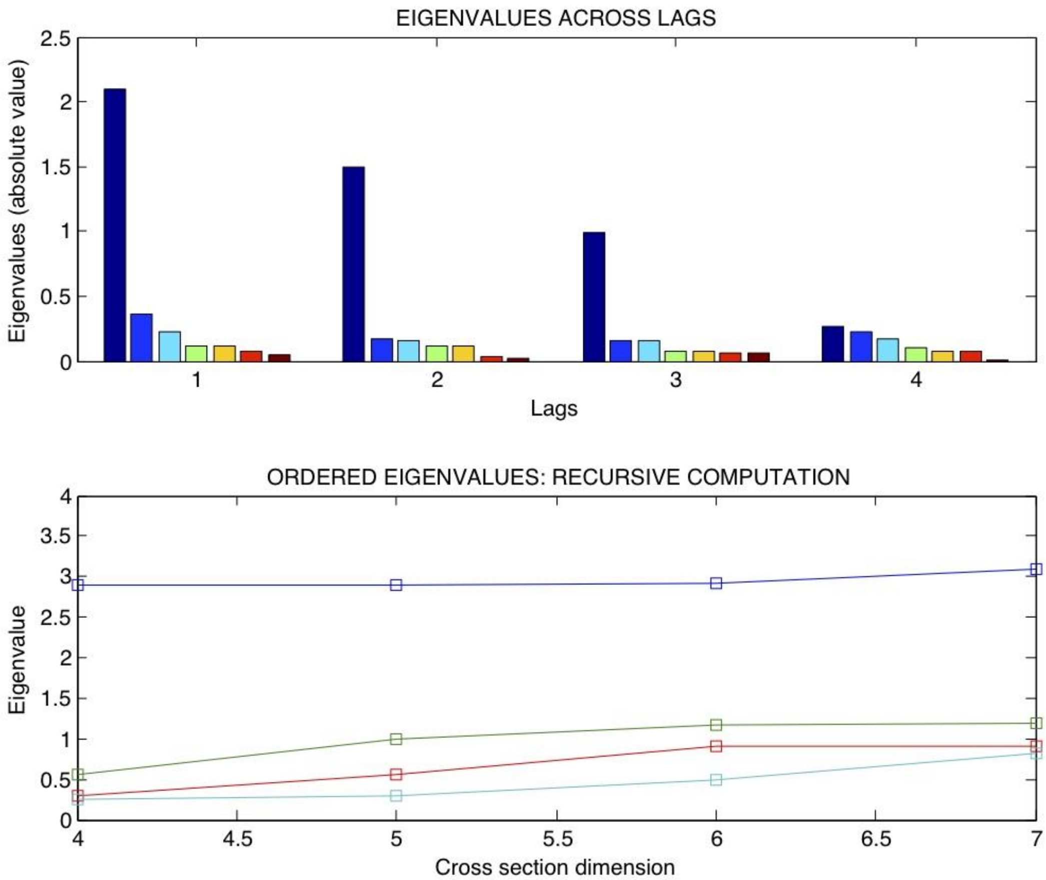


Turning to the identification of the common factors, the first panel of figure 10 highlights how one eigenvalue of the cross-correlation matrix stands out when compared across lags for the US case. Moreover, following Peña and Box (1987) procedure, one factor was identified for each of the three countries as the corresponding stability requirements for the first associated eigenvector are met as well. Interestingly, performing Forni *et al.* (2005b) methods, initially designed for large scale factor models, yields the same identification results as can be seen in the second panel of figure 10, with the eigenvalues computed recursively. By looking at the factor loadings, the extracted factor rests mainly on real series (GDP, consumption, investment and employment). However, nominal interest rate also stands as an important driver of this ”real” factor for the US and the euro area as can be seen in table 4. The impact of interest rate in the Spanish factor is much lower, possibly pointing to the importance of the ”one size fit all” monetary policy at the euro area level where the Spanish weight is around 10%. Real wages generally stand out with a lower loading in the common factor and interestingly, with a negative sign for Spain, confirming Messina *et al.* (2009) results on countercyclicality of the Spanish real wages, irrespectively of the deflator used. This real interpretation of the first common factor goes in line with Sargent and Sims (1977), who in a much larger set of series also assess that additional second and third factors share the nominal content of the GDP deflator, nominal wages and money supply variables.

Table 6. Factor loadings

Series	United States	Euro area	Spain
Y	0.75	0.77	0.85
C	0.86	0.89	0.85
I	0.90	0.91	0.77
H	0.85	0.78	0.90
Pr	0.14	0.09	0.10
W	-0.04	0.34	-0.31
I	0.51	0.52	0.26

Figure 10. DFM US identification results according to the cross-correlation matrix



The estimation of the structural DSGE-VAR sheds some light on possible misspecification issues. The posterior mode of the parameter λ for the US, the euro area and Spain is 1.66, 1.31 and 1.47, respectively. Following Consolo *et al.* (2009) $\frac{\lambda}{1+\lambda}$ is defined as the weight attached to the DSGE-generated data. The corresponding weights are 62%, 57% and 60%, confirming overall that the DSGE model restrictions are broadly supported by the data for the three countries. Although the US model is better specified as could be expected, the divergence between the three countries is minor despite their very different structures. The estimation of the Smets and Wouters (2005) model, tailored for a large-closed economy seems to work equivalently for the euro area and even for Spain, a small open economy without independent monetary policy.

3.4 Forecasting results: pre-crisis vs. crisis period

The design of the out-of-sample experimental forecasts follows a recursive procedure. Considering data until 2002Q4, consecutive estimations are performed up to 2016Q4, keeping one to eight steps out-of sample forecasts. The forecasting period is divided into different samples of 18 data points for the one-step ahead forecasts, covering two smooth growth periods [2003Q1-2007Q2] and a recession phase [2007Q3-2011Q4]. The forecasting performance can be dissected through four different dimensions: a time dimension (from one to eight quarters ahead), a contextual dimension (smooth growth period versus crisis period), a country-specific dimension (results for Spain, USA and the Euro area) and a model-specific dimension.

Abstracting first from model-specific aspects, figures 11 to 13 present RMSE results for the United States, the euro area and Spain, respectively. RMSEs are shown for the smooth as well as for the crisis period, signaling the first and third quartile, as well as the median and existing outliers. Several general conclusions appear to be robust across the different countries. First, as expected, the smooth growth period is characterized by smaller RMSEs at all horizons, especially for real variables. Second, the performance of the different models during the crisis worsens throughout the forecast horizon, contrary to the results for the stable period, except for employment. Third, the relative dispersion of the results is bigger in the short-term for the stable period, confirming that bad forecasting performance is generalized during the crisis. Fourth, independently of the country considered, the dynamic factor model clearly outperforms the rest in terms of interest rate forecasts and the mixed DSGE-VAR obtains significantly worse results in predicting inflation and salaries This can be seen in the form of an ”outlier” in the box-plot diagrams. Fifth, the size of the RMSEs is similar across countries for the different variables, with a notable exception, employment, which exhibits abnormally large RMSEs for the Spanish and the European case, confirming possible misspecification problems.

Figure 11. US RMSE analysis, different forecast horizons (1 to 8 steps ahead)

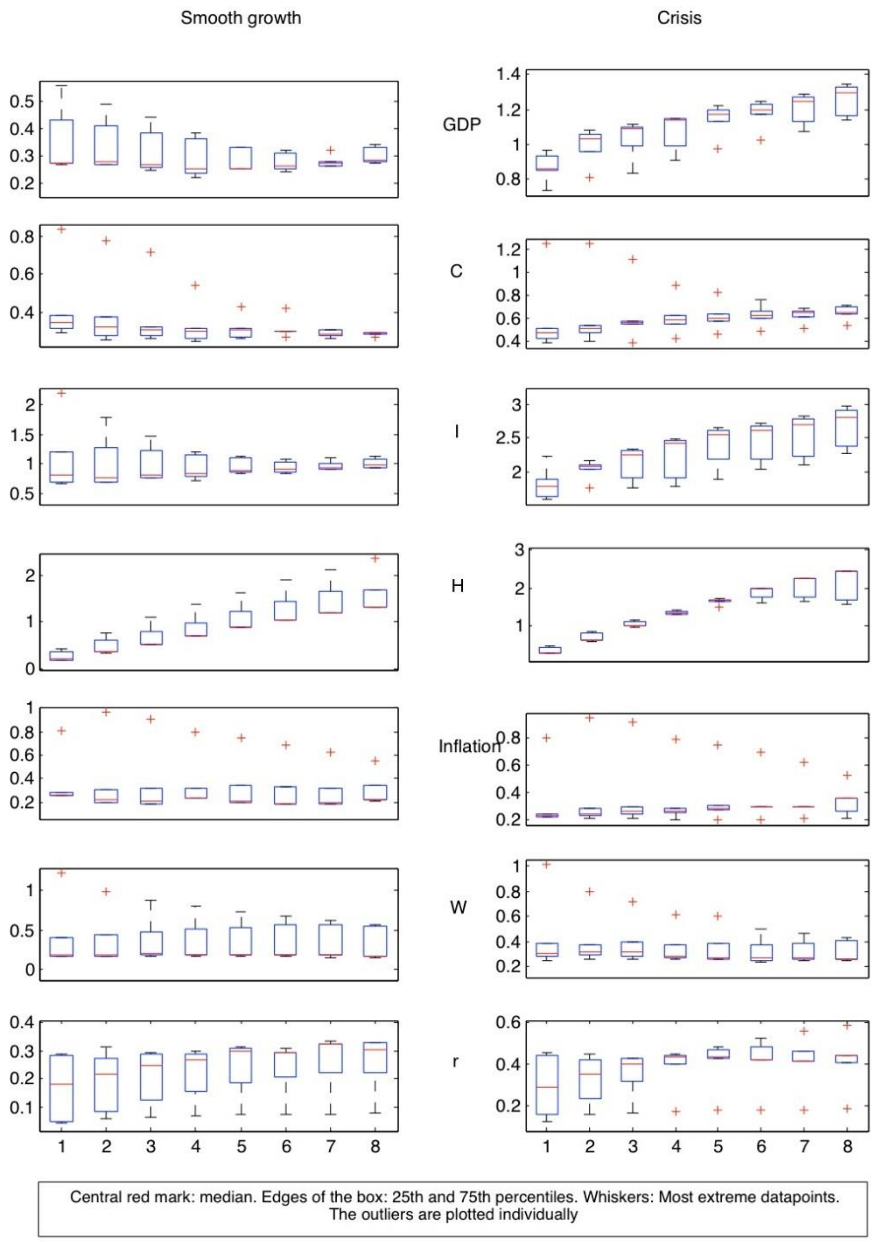
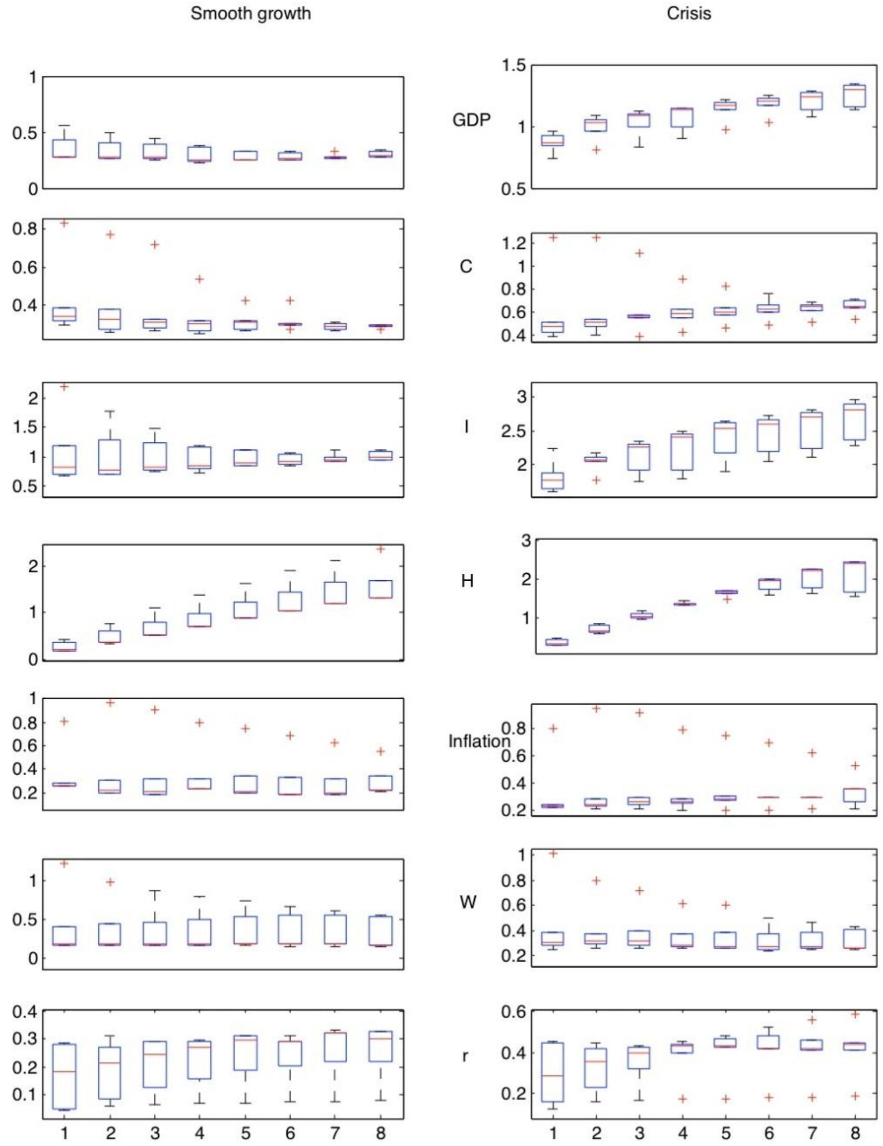
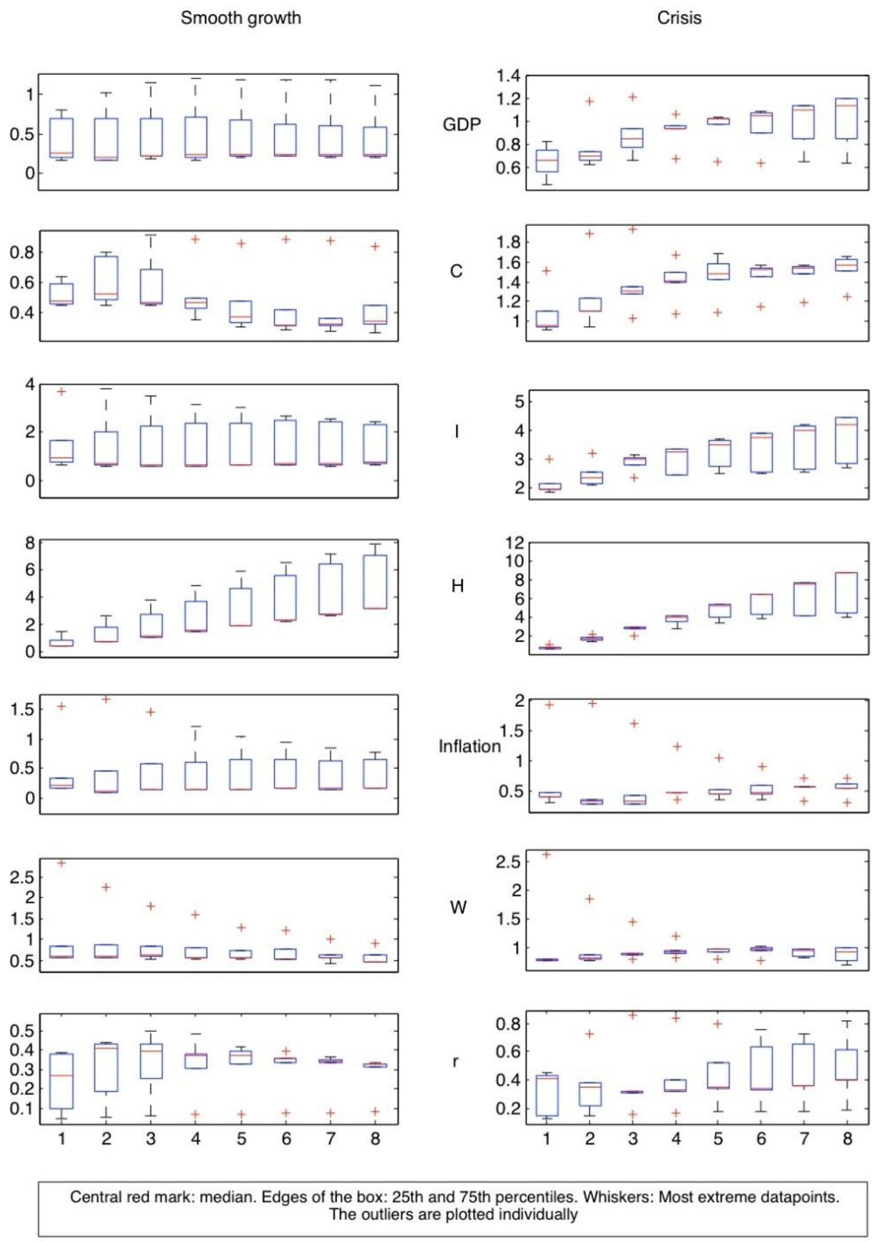


Figure 12. Euro area RMSE analysis, different forecast horizons (1 to 8 steps ahead)



Central red mark: median. Edges of the box: 25th and 75th percentiles. Whiskers: Most extreme datapoints. The outliers are plotted individually

Figure 13. Spain RMSE analysis, different forecast horizons (1 to 8 steps ahead)



As stated during the methodological section, predictive results from purely econometrical models [AR, VAR models as well as their Bayesian estimated version with statistical priors] are compared to those from structural models

[DSGE], considering the DSGE-VAR as a natural, mixed benchmark. The DSGE-VAR framework presents a good fit to aggregated data and moreover retains the theoretical prescriptions from DSGE models. Figures 14 to 16 provide such a comparison across models, considering the time-dimension and differences between countries. The percentage deviations of the different models with respect to the DSGE-VAR root mean square errors at the different horizons are represented. A negative value implies a gain in forecast accuracy and therefore lower RMSEs than the benchmark DSGE-VAR.

Independently of the country or the period considered, there are accuracy gains from the DSGE-VAR along the time dimension (as in Wieland and Wolters, 2011). It follows that its relative forecasting results turn out to be better in the medium to long-run than in the short-run.

The performance of structural models during both periods is markedly better for the real variables, where they obtain relative predictive gains. Misspecification concerns arise when looking at employment forecasts during the smooth growth period, as its behavior is closer to nominal variables than to GDP, consumption and investment. Nominal variables, in turn, are dominated by the DFM model, which scores better for inflation, wages and interest rates across countries and periods, independently of the forecast horizon.

The relative performance of structural models improves during the crisis period for all countries and all variables. This is particularly striking for Spain, where DSGE fore- casting results are relatively poor during the smooth growth period. This finding could point out to possible misspecification problems of the Smets and Wouters (2005) model for the Spanish economy.

Figure 14. RMSE gap (%) with respect to the DSGE-VAR, US (1 to 8 steps ahead)

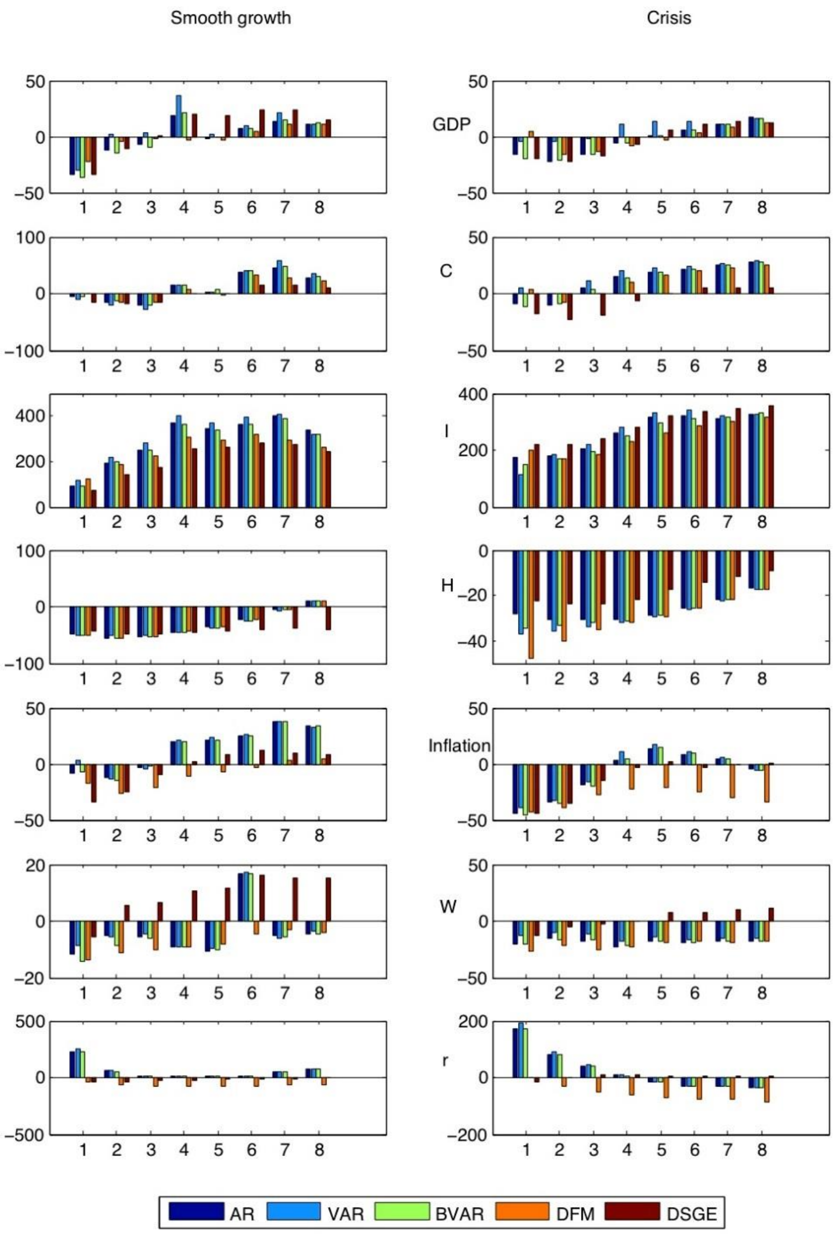


Figure 15. RMSE gap (%) with respect to the DSGE-VAR, euro area (1 to 8 steps ahead)

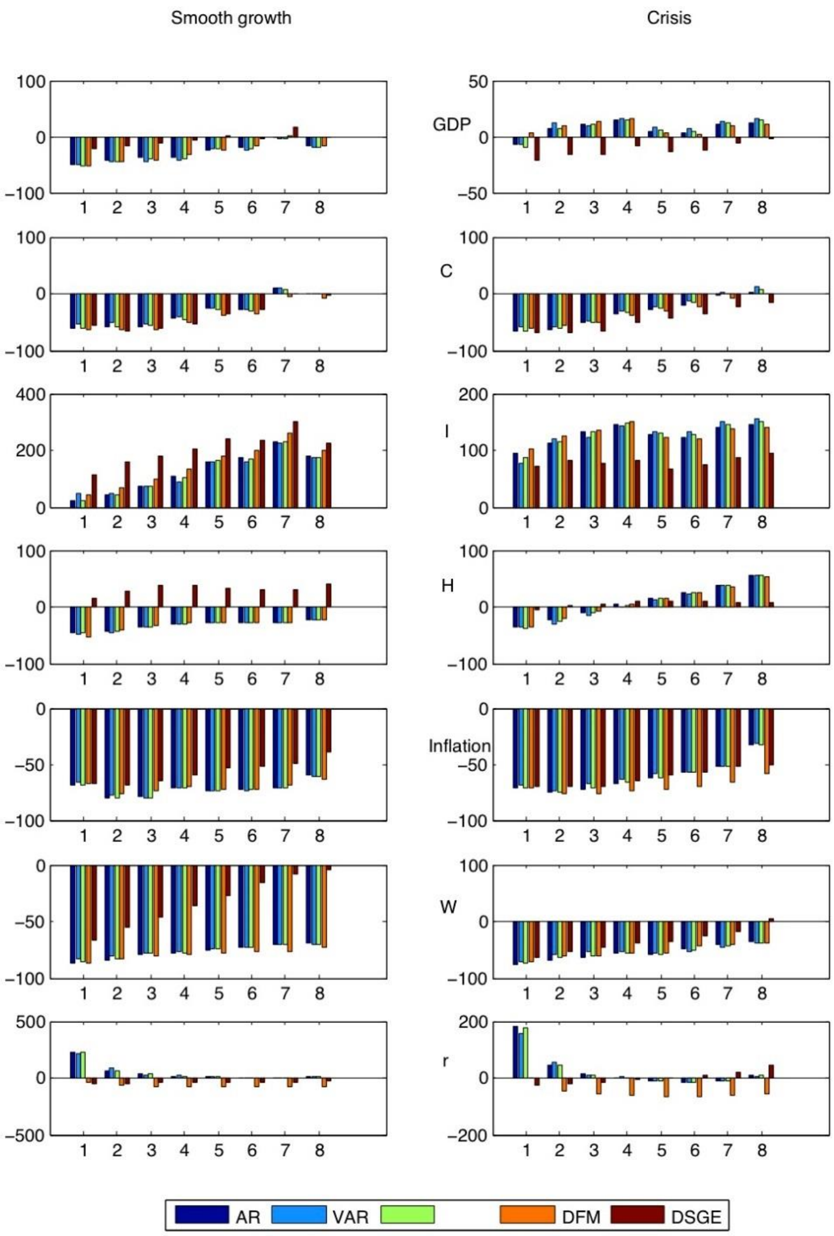
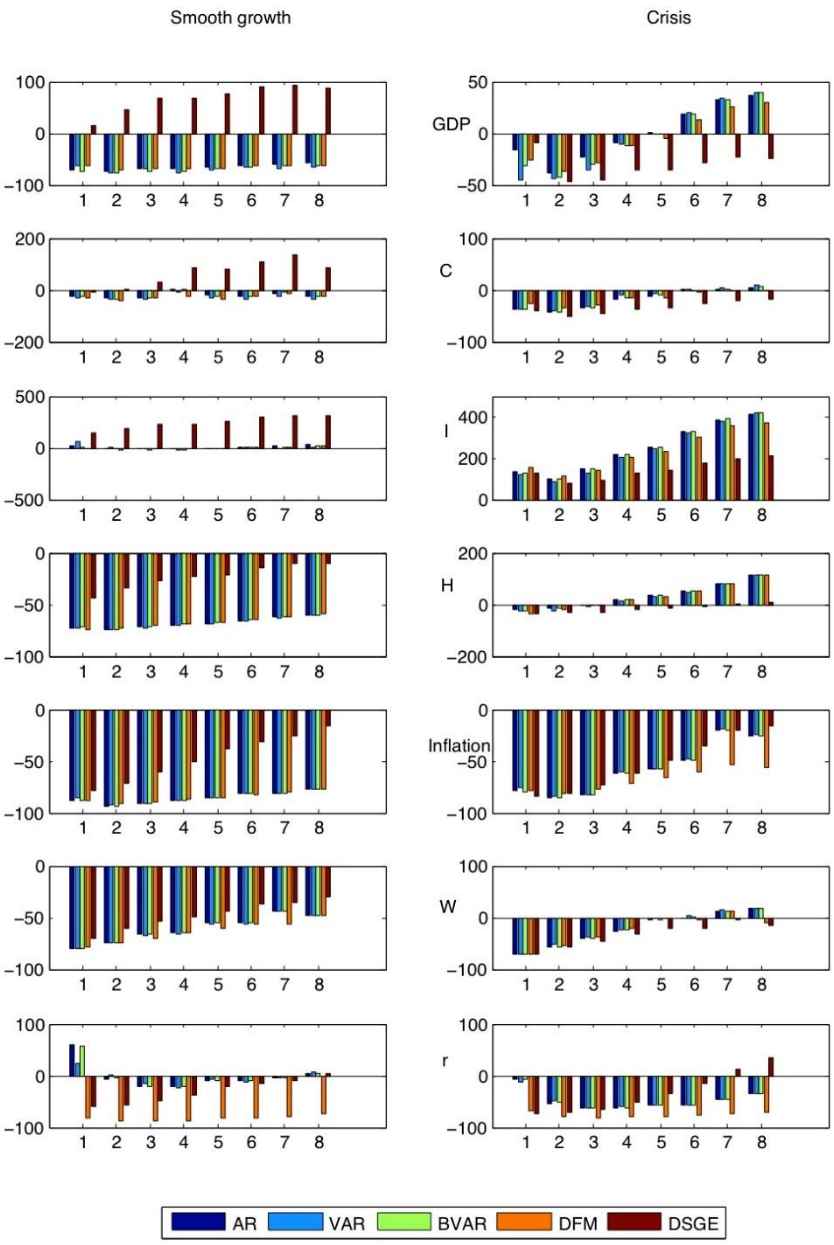


Figure 16. RMSE gap (%) with respect to the DSGE-VAR, Spain (1 to 8 steps ahead)



All in all, non-structural models perform relatively well during the smooth growth period at all forecast horizons, for all variables. This results would clearly point to the benefits of theoretical foundations for forecast accuracy in a period of higher volatility.

There seems to be a trade-off between considering stationary series in the non-structural models (with an extra difference in wages, inflation, interest rate and employment) and sticking to theoretically relevant concepts. The former strengthens the performance in the smooth growth periods and the short to medium term, while the latter is particularly relevant while facing turbulent times and high volatility in the economic cycle.

3.5 Post-crisis period: back to normal?

The post-crisis period [2012Q1-2016Q4] can be looked upon to check whether the conclusions of the pre-crisis years hold again. The updated results for the US case can be seen in Figures 17 and 18.

In terms of the absolute error size, the pre-crisis period conclusions hold. The latest years are also characterized by smaller and non-increasing rmse values, again except for employment forecasts. The relative dispersion has also shrunk, and is comparable to the previous smooth period. Second, when looking at the specific models, the benefits of the DSGE-VAR benchmark appear again to be present in the medium term. The performance of the dynamic factor model outperforms the rest not only for the nominal variables, it now presents very competitive results for the GDP forecast at all horizons.

Figure 17. Spain RMSE analysis, different forecast horizons (1 to 8 steps ahead)

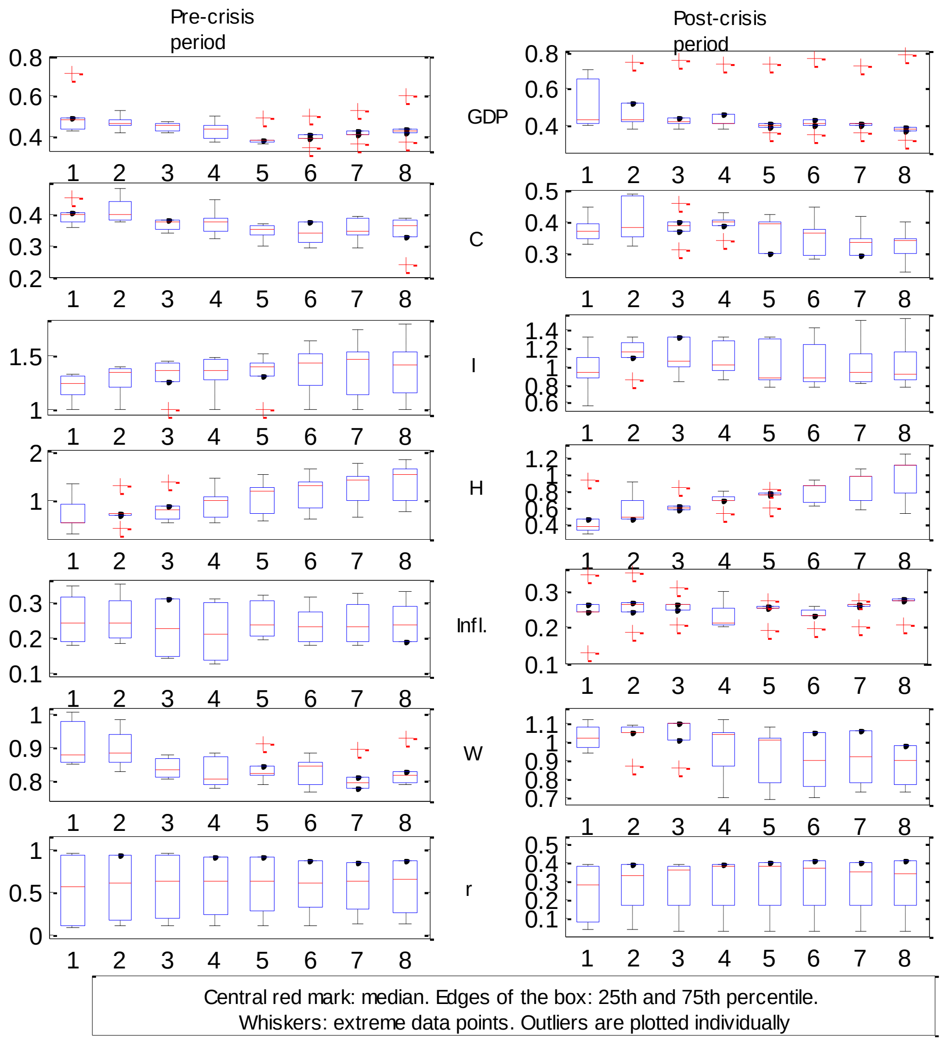
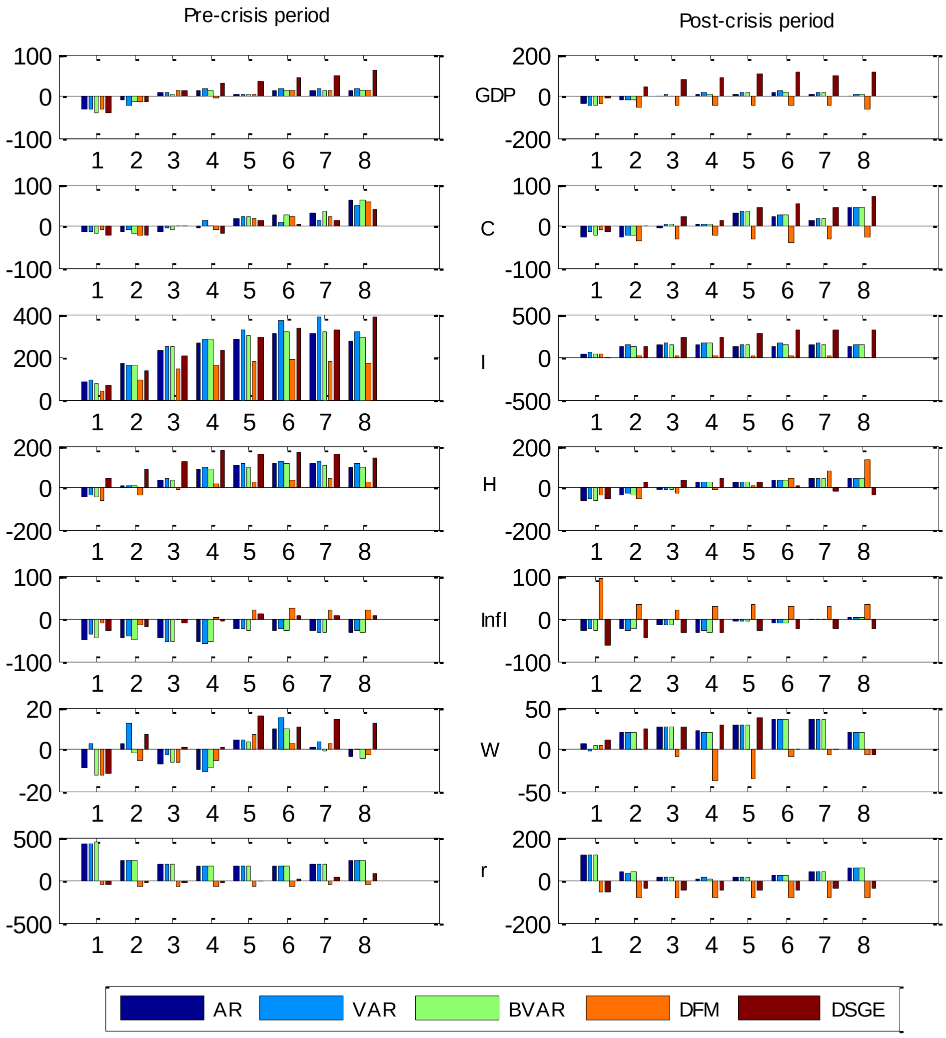


Figure 18. RMSE gap (%) with respect to the DSGE-VAR, US (1 to 8 steps ahead)



3.6 Concluding remarks

This chapter conducted a comparative analysis of the out-of-sample forecasting performance of structural and non-structural models with quarterly data covering the period 1980Q1 to 2010Q4 for seven macroeconomic aggregates: Gross Domestic Product (GDP), private consumption, private investment, employment or total hours worked, the GDP deflator, real wages and the nominal interest rate. The forecasting performance was assessed using a recursive procedure through four different dimensions: a time dimension (from one to eight quarters ahead), a contextual dimension (smooth growth period [2003Q1-2007Q2] and recession phase [2006Q3-2010Q4]), a country-specific dimension (results for Spain, USA and the Euro area) and a model-specific dimension (comparison against traditional benchmarks such as VARs and BVARs).

All in all, there is supporting evidence for forecasting accuracy gains from structural models in the medium to long-run, while non-structural models perform generally better in the short-run. The benefits of structural models increase at all time horizons when considering disruptive times, with behavioral restrictions leading to higher parsimony. Indeed, during the ”Great Moderation” preceding the financial crisis, all models seemed to predict reasonably well at the different forecast horizons but this regularity was broken with the onset of the crisis as there is evidence of significant increases in the relative performance of DSGE models for all countries and all variables. It would thus seem that forecasters should beware of too stable periods as the performance of the different models might not reflect their accuracy in terms of capturing the underlying economic developments but rather the regularity of the latter.

Moreover, Bayesian restrictions seem successful in shrinking the parameter space and providing better forecasting results. The results for the structural model, in turn, are robust across the different countries. Although the Smets and Wouters (2005) was initially tailored for the US economy (large and relatively closed), its forecasting gains with respect to non-structural models also apply to Spain and the euro area, despite their very different structures. It is thus natural at this stage to wonder whether misspecification issues have a sizable impact on the forecasting performance of these models. The latest theoretical refinements might not prove worth the effort when the goal is not better knowledge of the transmission channels of the different shocks but simply forecasting. A complementary line of research would mimic these results in a non-linearised environment, to check for the influence of the log-linearisation process in watering out modeling refinements.

3.7 References

Adolfson, M., Laséen, S., Lindé, J. and Villani, M. (2007). "Evaluating an estimated New Keynesian small open economy model," *Journal of Economic Dynamics and Control* vol. 32(8), 2690-2721.

Amisano, G. and Serati, M. (2002). "BVAR models and forecasting: a quarterly model for the EMU-11," *Statistica* vol. 62(1), 51–70.

Baeurle, G. (2008). "Priors from DSGE models for dynamic factor analysis," Diskussionsschriften n. 03, Universitaet Bern, Departement Volkswirtschaft.

Boivin, J., and Giannoni, M. (2006). "DSGE models in a data-rich environment," NBER Working Papers n. 12772, National Bureau of Economic Research.

Box, G., and Jenkins, G. (1970). "Time series analysis: Forecasting and control," Rev. ed., San Francisco, Holden-Day, c1976.

Canova, F. (2007). "Methods for applied macroeconomic research," Princeton University Press.

Christoffel, K., Coenen, G. and Warne, A. (2010). "Forecasting with DSGE models," Working Paper Series n. 1185, European Central Bank.

Consolo, A., Favero, C.A. and Paccagnini, A. (2009). "On the statistical identification of DSGE models," *Journal of Econometrics* vol. 150(1), 99 – 115.

Del Negro, M. and Schorfheide, F. (2003). "Take your model bowling: forecasting with general equilibrium models," *Economic Review*, vol. 88(4), 35–50.

Del Negro, M. and Schorfheide, F. (2004). "Priors from general equilibrium models for VARs," *International Economic Review*, vol. 45(2), 643–673.

Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. (2004). "On the fit and forecasting performance of New Keynesian models," Working Paper Series n. 491, European Central Bank.

Del Negro, M.G., Gupta, A., Li, P. and Moszkowski, E. (2017). "The FRBNY DSGE Model Forecast—February 2017," Federal Reserve Bank of New York.

Doan, T., Litterman, R. and Sims, C. (1984). "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews* vol. 3(1), 1–100.

Edge, R.M., Kiley, M.T and Laforte, J.P. (2009). "A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a DSGE model," *Journal of Applied Econometrics*, vol. 25(4), 720-754.

Fagan, G., Henry, J. and Mestre, R. (2001). "An area-wide model (AWM) for the euro area," Working Paper Series n.42, European Central Bank.

Fernández-Villaverde, J. and Rubio-Ramírez, J. (2005). "Estimating dynamic equilibrium economies: linear versus nonlinear likelihood," *Journal of Applied Econometrics,* vol. 20(7), 891–910.

Fernández-Villaverde, J. and Rubio-Ramírez, J. (2007). "Estimating macroeconomic models: A likelihood approach," *Review of Economic Studies*, vol.74(4), 1059–1087.

Forni, M., Giannone, D., Lippi, M. and Reichlin, L. (2009). "Opening the black box: Structural factor models with large cross sections," *Econometric Theory*, vol. 25(05), 1319–1347.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005a). "The generalized dynamic factor model: One-sided estimation and forecasting," *Journal of the American Statistical Association*, vol. 100(471), 830–840.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005b). "The generalized dynamic factor model: identification and estimation," *Review of Economics and Statistics*, vol. 82(4), 540–554.

Gali, J. and Gertler, M. (2007). "Macroeconomic modeling for monetary policy evaluation," *Journal of Economic Perspectives*, vol. 21(4), 25-46.

Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,". *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6(6), 721–741.

Geweke, J. (1977). "The dynamic factor analysis of economic time series," in Latent Variables in Socio-Economic Models, D. Aigner and A. Goldberger (eds.).

Geweke, J. (2005). "Contemporary Bayesian econometrics and statistics," Wiley series in probability and statistics, Wiley-Interscience, Hoboken, NJ.

Giannone, D., Monti, F. and Reichlin, L. (2009). "Incorporating conjunctural analysis in structural models," in The Science and Practice of Monetary Policy Today, Wieland, Volker (eds.), Springer.

Giannone, D., Reichlin, L. and Sala, L. (2006). "VARs, common factors and the empirical validation of equilibrium business cycle models," *Journal of Econometrics*, vol. 132(1), 257–279.

Goodfriend, M. (2004). "Monetary policy in the New Neoclassical synthesis: A primer," *Economic Quarterly*, Federal Reserve Bank of Richmond, issue Sum, 21-45.

Greenberg, E. (2007). "Introduction to Bayesian Econometrics," Cambridge University Press (eds.).

Hansen, L. and Sargent, T. (1980). "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, vol. 2(1), 7–46.

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov Chains and their applications," *Biometrika*, vol. 57(1), 97–109.

Hodge, A., Robinson, T. and Stuart, R. (2008). "A small BVAR-DSGE model for forecasting the Australian economy," RBA Research Discussion Papers n. 04, Reserve Bank of Australia.

Ingram, B. and Whiteman, C. (1994). "Supplanting the 'Minnesota' prior: Forecasting macroeconomic time series using real business cycle model priors," *Journal of Monetary Economics*, vol. 34(3), 497–510.

Kinal, T. and Ratner, J. (1986). "A VAR forecasting model of a regional economy: its construction and comparative accuracy," *International Regional Science Review*, vol. 10(2), 113–26.

Kolasa, M., Rubaszek, M. and Skrzypczynski, P. (2009). "Putting the New Keynesian DSGE model to the real-time forecasting test," Working Paper Series n. 1110, European Central Bank.

Kydland, F. and Prescott, E. (1982). "Time to build and aggregate fluctuations," *Econometrica* vol. 50(6), 1345–70.

Litterman, R. (1986). "Forecasting with Bayesian Vector Autoregressions-five years of experience," *Journal of Business & Economic Statistics*, vol. 4(1), 25–38.

Lucas, R. (1976). "Econometric policy evaluation: A critique," *Carnegie-Rochester Conference Series on Public Policy*, vol. 1(1), 19–46.

Lütkepohl, H. (2005). "New introduction to multiple time series analysis," Springer (eds.).

Messina, J., Strozzi, C. and Turunen, J. (2009). "Real wages over the business cycle: OECD evidence from the time and frequency domains," *Journal of Economic Dynamics and Control*, vol. 33(6), 1183–1200.

Metropolis, N., Rosenbluth, A.N., Rosenbluth, M.N., Teller, A.H. and Teller E. (1953). "Equation of state calculation by fast computing machines," *Journal of Chemical Physics* 21(6), 1087-1092.

Monti, F. (2008). "Forecast with judgment and models," Research Series n. 12-2, National Bank of Belgium.

Peña, D. and Box, G.E.P. (1987). "Identifying a simplifying structure in time series," *Journal of the American Statistical Association*, vol. 82(399), 836–843.

Peña, D. and Poncela, P. (2004). "Forecasting with nonstationary dynamic factor models," *Journal of Econometrics*, vol. 119(2), 291–321.

Pichler, P. (2007). "Forecasting with estimated dynamic stochastic general equilibrium models: The role of nonlinearities," *B.E. Journal of Macroeconomics*, vol. 8(1), 1-35.

Rubaszek, M. and Skrzypczynski, P. (2008). "On the forecasting performance of a small-scale DSGE model," *International Journal of Forecasting*, vol. 24(3), 498–512.

Sargent, T., and Sims, C. (1977). "Business cycle modeling without pretending to have too much a priori economic theory," Working Papers n. 55, Federal Reserve Bank of Minneapolis.

Schorfheide, F., Sill, K., and Kryshko, M. (2010). "DSGE model-based forecasting of non-modelled variables," *International Journal of Forecasting*, vol. 26(2), 348– 373.

Sims, C. (1980). "Macroeconomics and reality," *Econometrica*, vol. 48(1), 1–48.

Sims, C. (1992). "A nine-variable probabilistic macroeconomic forecasting model," Cowles Foundation Discussion Papers n. 1034, Cowles Foundation for Research in Economics, Yale University.

Sims, C. (2002a). "The role of models and probabilities in the monetary policy process," *Brookings Papers on Economic Activity*, vol. 2002(2), 1-40.

Sims, C. (2002b). "Solving linear rational expectations models," *Computational Economics* 20(1-2), 1–20.

Smets, F., and Wouters, R. (2003). "An estimated dynamic stochastic general equilibrium model of the euro area,". *Journal of the European Economic Association*, vol. 1(5), 1123–1175.

Smets, F., and Wouters, R. (2004). "Forecasting with a Bayesian DSGE model: An application to the euro area," *Journal of Common Market Studies*, vol. 42(4), 841–867.

Smets, F., and Wouters, R. (2005). "Comparing shocks and frictions in us and euro area business cycles: a Bayesian DSGE approach," *Journal of Applied Econometrics*, vol. 20(2), 161–183.

Stock, J., and Watson, M. (1996). "Evidence on structural instability in macroeconomic time series relations," *Journal of Business & Economic Statistics*, vol. 14(1), 11–30.

Stock, J., and Watson, M. (2002). "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, vol. 20(2), 147–62.

Waggoner, D., and Zha, T. (2012). "Confronting model misspecification in macroeconomics," *Journal of Econometrics*, vol. 171(2), 167-184.

Wang, M. (2009). “Comparing the DSGE model with the factor model: an out-of-sample forecasting experiment,” *Journal of Forecasting*, 28(2), 167-182.

Wieland, V. and Wolters, M.H. (2011). “The diversity of forecasts from macroeconomic models of the US economy,” *Economic Theory*, vol. 47(2–3), 247–292.

Zarnowitz, V., and Braun, P. (1993). “Twenty-two years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and comparisons of forecasting performance,” Chicago: University of Chicago Press.

4 International Transmission of Shocks

4.1 Introduction

Since the onset of the Economic and Monetary Union, its Member States (MS) have experienced an enhanced interdependence, with growing trade and financial bilateral bonds. These two channels have transmitted country-specific shocks, generating cross-border spillover effects. Moreover, the common institutional framework further reinforced these conduits via the contagion of market participants' perception as well as consumers' and firms' sentiment.

The relative importance of each of these channels (trade, financial and contagion) changes over time, along with the business cycle. Trade flows are considered more stable in nature, while financial flows have proven to be more volatile, especially throughout the crisis. As can be seen in Figure 19 for Germany, bilateral trade flows with its main partners, remained rather steady between 2007 and 2015. Financial flows, defined as cross-border banking exposures coming from Bank of International Settlements (BIS international consolidated banking statistics as in Eickmeier and NG, 2011), have experienced higher volatility. Figure 20 shows that a clear and generalized retreat of German banks in terms of their counterpart countries' total exposures, except for Greece, where the overall capital flight more than compensated for the reduced German exposure, leaving the German share even larger. Added to this, contagion via agents´ perception has also played a varying role throughout the European sovereign debt crisis. The fragmentation between peripheral and core EU bond markets can be seen in Figure 21, which shows a dramatic drop in the correlation between the long-term yields of Spanish bonds with Germany or France, contrary to the rather high co-movement between the Spanish and the EU peripheral markets.[18]

Moreover, from a cross-sectional perspective, the bilateral linkages between countries might appear strikingly different according to each one of these channels, at any point in time. For example, the influence of Germany on other MS via trade exchanges appears the highest for Austria, Switzerland and the Netherlands, as shown in Figure 19. On the other hand, bilateral financial exposures of are more diversified and the corresponding German share in them is significantly smaller when compared to the trade channel.

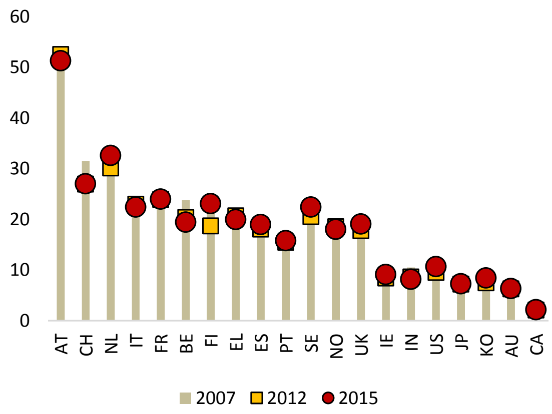
The relative weight attached to each one of these channels will therefore have critical implications in the assessment of the existing bilateral interlinkages and potential spillovers from idiosyncratic shocks. The optimal weighting scheme for the different channels is, however, difficult to grasp empirically and the literature generally opts for simplifying assumptions, such as (i) focusing on one channel, generally trade flows for data availability reasons, omitting the others, as in Pesaran *et al*. (2004); (ii) splitting variables in trade vs. financially interconnected concepts, artificially restraining the complexity of the bilateral relationship among countries (see Eckmeier and Ng, 2011); and (iii) estimating the interlinkages between the countries as a separate parameter, without reference to trade, financial or confidence channels and thus with no possible economic interpretation, as in Gross (2013).

This chapter opts for an intermediate approach, calibrating the relative weights of the different transmission channels (trade, financial and confidence) in a Global VAR framework, according to the short-term GDP forecast accuracy of the model. Once the relative weights of the channels are calibrated, they are used together with the actual bilateral flows to construct a weighted indicator reflecting the potential systemic spillover capacity of a country.

---

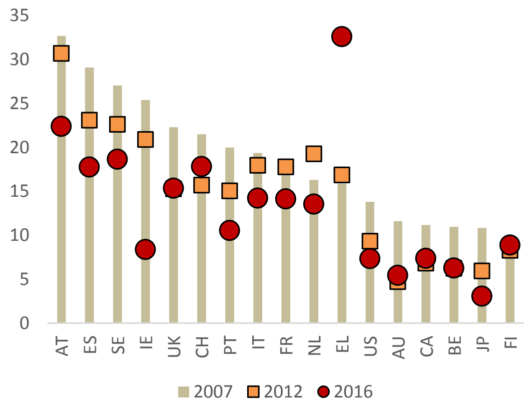
[18] The correlation is obtained by 20 quarters trailing averages of long-term (10 years) sovereign yields.

Figure 19. German trade flows (exports and imports) as a share of its counterparty total trade flows, %



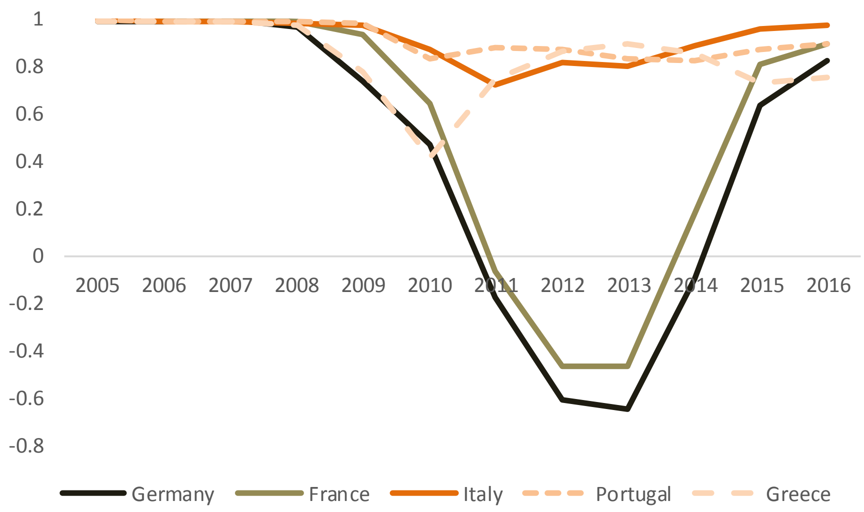
Source of data: IMF Direction of Trade Statistics.

Figure 20. financial bilateral exposures as a share of its counterparty total external exposures, %



Source of data: BIS international banking statistics.

Figure 21. Correlation between the 5-year trailing average of long-term sovereign yields, Spain vs. selected countries



Source of data: IMF, International Financial Statistics.

The chapter is structured as follows; first, section 2 develops the GVAR theoretical framework, then section 3 specifies its set-up and section 4 builds on the specified GVAR to construct a bilateral Index of Systemic Spillovers (BSS). Finally, section 5 concludes.

4.2 GVAR theoretical framework

GVAR models combine individual country Vector Autoregressive models with exogenous variables (VARX) that include domestic as well as country-specific foreign variables (see Pesaran 2015 for a complete introduction to GVAR models). The latter are constructed as a weighted average of domestic variables, according to relative weights (trade, financial or any other transmission pattern could be used) and summarize information on common unobserved factors that drive the domestic economy (e.g. technological progress, which affects all countries but is not observed).

Once estimated, the individual models are stacked into a global vector autoregressive model, which can illustrate the transmission of shocks between different economies through 3 different channels:

i. The contemporaneous impact elasticities of foreign variables on their domestic counterparts.[19]
ii. Weak dependence of shocks between countries via a non-diagonal VCV matrix of the errors of the aggregated model.
iii. Common observed global exogenous variables having an impact on all the specific country models, such as oil price shocks.

4.2.1 Individual country models

VARX(p_i, q_i) structure of the $i = 0,1,\dots,N$ N countries, including a constant (a_{io}) and a linear trend (a_{i1}t) term;

$$x_{it} = a_{io} + a_{i1}\mathrm{t} + \Phi_{i1}x_{it-1} + \cdots + \Phi_{ip_i}x_{it-p_i} + \Lambda_{i0}x_{it}^{*} + \Lambda_{i1}x_{it-1}^{*} + \cdots + \Lambda_{iq_i}x_{it-q_i}^{*} + u_{it} \quad (1)$$

With $x_{it}\colon \mathrm{k}_i \times 1$ being the country-specific vector of domestic variables and $x_{it}^{*}\colon k_i^{*} \times 1$ representing foreign variables, defined as weighted averages of domestic ones in all countries.

[19] A consistent estimation of the contemporaneous effects of foreign variables requires dealing with serial correlation in the residuals of the error correction equations on top of assuming weak exogeneity of the foreign variables.

$$x_{it}^{*} = \sum_{j=0}^{N} w_{ij} x_{jt}, w_{ii} = 0$$

The lag order of domestic and foreign variables (p_i, q_i) is selected according to the usual Akaike or Bayesian information criteria. In practice, it must be noted that a small number of lags is generally enough to capture the underlying dynamics in the data (generally 1 to 2 lags in a multiequational system with quarterly frequency). The lag order of the GVAR, denoted by p, is computed as the maximum value from the individual lags. To keep the model parsimonious while dealing with serial correlation, the lag order selection for the GVAR is restricted so that $p_i \leq q_i, \forall i$ and $q_i \leq p - 1, \forall i$.

The $p_i = q_i = 2$ version of the individual VARX models depicted in (1) corresponds to

$$x_{it} = a_{io} + a_{i1}\mathrm{t} + \Phi_{i1} x_{it-1} + \Phi_{i2} x_{it-2} + \Lambda_{i0} x_{it}^{*} + \Lambda_{i1} x_{it-1}^{*} + \Lambda_{i2} x_{it-2}^{*} + u_{it} \qquad (1')$$

4.2.2 GVAR model

The aggregate GVAR is constructed considering all the variables as endogenous to the global system. First, departing from (1′), the individual models can be written in terms of $z_{it} = (x'_{it}, x^{*\prime}_{it})'$;

$$A_{i0} z_{it} = a_{io} + a_{i1}\mathrm{t} + A_{i1} z_{it-1} + A_{i2} z_{it-2} + u_{it} \qquad (2)$$

with $A_{i0} = \left(I_{k_i}, -\Lambda_{i0}\right)$, $A_{i1} = (\Phi_{i1}, \Lambda_{i1})$ and $A_{i2} = (\Phi_{i2}, \Lambda_{i2})$ and zero mean, serially uncorrelated individual country disturbances $u_{it}: \mathrm{k}_i \times 1$, which can be aggregated in a joint vector u, $u = (u_{0t}, u_{1t}, u_{2t}, \ldots, u_{Nt})'$ defining a non-singular Variance Covariance matrix (VCV), Σ.

$$\Sigma = E(uu') = \begin{pmatrix} V(u_0) & Cov(u_0, u_1) & \ldots & Cov(u_0, u_N) \\ Cov(u_1, u_0) & V(u_1) & \ldots & Cov(u_1, u_N) \\ \vdots & \vdots & & \vdots \\ Cov(u_N, u_0) & Cov(u_N, u_1) & \ldots & V(u_N) \end{pmatrix}$$

Second, building on the identity that links the vector z_{it} with the vector x_t that contains all the endogenous variables, *via* the weight matrices W_i, $z_{it} = W_i x_t$ with $x_t = (x'_{0t}, x'_{1t}, \ldots, x'_{Nt})': k \times 1$ and W_i a $(k_i \times k_i{}^{*}) \times k$, a matrix of known constants defined according to the country-specific weights, with $k = \sum_{i=0}^{N} k_i$.

Third, replacing z_{it} back into (2) and stacking the model for x_t:

$$G_0 x_t = a_o + a_1 \mathrm{t} + G_1 x_{t-1} + G_2 x_{t-2} + u_t$$

$$G_0 = \begin{pmatrix} A_{00}W_0 \\ A_{10}W_1 \\ \vdots \\ A_{N0}W_N \end{pmatrix},\ G_1 = \begin{pmatrix} A_{01}W_0 \\ A_{11}W_1 \\ \vdots \\ A_{N1}W_N \end{pmatrix},\ G_2 = \begin{pmatrix} A_{02}W_0 \\ A_{12}W_1 \\ \vdots \\ A_{N2}W_N \end{pmatrix},\ a_0 = \begin{pmatrix} a_{00} \\ a_{10} \\ \vdots \\ a_{N0} \end{pmatrix},\ a_1 = \begin{pmatrix} a_{01} \\ a_{11} \\ \vdots \\ a_{N1} \end{pmatrix},\ u_t = \begin{pmatrix} u_{0t} \\ u_{1t} \\ \vdots \\ u_{Nt} \end{pmatrix}$$

If G_0 is known and non-singular, it can be inverted to yield a model that can be solved recursively without any restrictions on the VCV, $\Sigma_\varepsilon = \mathrm{E}(\varepsilon_t \varepsilon_t')$,

$$x_t = b_o + b_1 \mathrm{t} + F_1 x_{t-1} + F_2 x_{t-2} + \varepsilon_t \qquad (3)$$

with $F_1 = G_0^{-1} G_1$, $F_2 = G_0^{-1} G_2$, $b_0 = G_0^{-1} a_0$, $b_1 = G_0^{-1} a_1$, $\varepsilon_t = G_0^{-1} u_t$

The aggregate model cannot be estimated directly due to the existing high number of parameters and the lack of degrees of freedom in the estimation process. To circumvent the "course of dimensionality" issue, the weights defining the foreign variables are considered as given[20] and individual models are estimated independently. The latter allows for matrix G_0 to be known and non-singular and thus for a recursive solution of the joint GVAR.

However, the estimation of the individual VARX models requires the number of countries (N) to be sufficiently large and 3 additional conditions to be met, as stated in Pesaran *et al*. (2004);

i. Smallness: the weights $w_{ij}, j = 0,1,\ldots,N$ satisfy the granularity conditions to minimise the impact of one country-dominance and ensure that the weights of country i with respect to all partners sum up to one, $\sum_{j=0}^{N} w_{ij} = 1$.
ii. Dynamic stability: All eigenvalues of matrix F_1 lie on or inside the unit circle.

---

[20] Generally fixed weights are used to avoid an excessive degree of randomness in the inference results. However, to avoid a strong dependence of the results on the chosen period to set the weights, time averages are used in practice (either 3 or 5 year averages).

iii. Weak exogeneity of foreign variables: the estimation of the country-specific VARX models is conditioned on x_{it}^{*} variables by treating them as I(1) weakly exogenous or long-run forcing. Akin of the small open economy assumption in standard macroeconomic models, this assumption implies no long-run feed-back from domestic variables to their foreign counterpart without necessarily ruling out short-run interactions and long-term impact of foreign variables on their domestic counterparts.

Ideally, once conditioning the estimation on the foreign variables, the remaining cross-sectional correlation should stay weak, so that $Cov(x_{it}^{*}, u_{it}) \rightarrow 0, with\ N \rightarrow \infty$. The presence of foreign variables can thus be interpreted as a proxy for common unobserved shocks as it reduces the degree of correlation of the remaining shocks across the different countries. Weak exogeneity of foreign variables is indeed compatible with weak dependence across the individual country disturbances, so that the VCV can be defined as non-diagonal, with the non-zero non-diagonal elements represent spillover effects or weak [21] residual dependencies between the different individual models.

4.2.3 Estimating the country-specific VARX models

The estimation of the individual models uses reduced rank regression techniques on the error correction form of equation (1')

$$\Delta x_{it} = c_{io} - \alpha_i \beta_i' [z_{it-1} - \gamma_i (t-1)] + \Lambda_{i0} x_{it}^{*} + \Gamma_i \Delta z_{it-1} + u_{it} \tag{4}$$

Where the vector z_{it} is defined as $z_{it} = (x_{it}', x_{it}^{*\prime})'$ and the trend term is restricted to lie within the cointegrating space $\alpha_i \beta_i' [z_{it-1} - \gamma_i (t-1)]$ to avoid

[21] Strong dependencies are taken care by the presence of foreign variables in domestic models. Weak residual dependency implies that quadratic elements will converge to 0 as N goes to ∞.

quadratic trends in x_t.[22] Moreover, both the loading matrix $\alpha_i\colon \mathrm{k}_i \times r_i$ and the matrix of cointegrating vectors $\beta_i\colon (\mathrm{k}_i + k_i^*) \times r_i$ are of full column rank, r_i.

In order to allow for the possibility of cointegration within x_{it} and between x_{it} and x_{it}^* and consequently across x_{it} and x_{jt} for $i \neq j$, β_i can be partitioned as $\beta_i = (\beta_{ix}', \beta_{ix^*}')'$ allowing for the r_i cointegration terms to be redefined as $\beta_i'(z_{it} - \gamma_i t) = \beta_{ix}' x_{it} + \beta_{ix^*}' x_{it}^* - (\beta_i' \gamma_i)t$.

In practice, as stated above, estimation is conditional on x_{it}^*, using reduced rank regression to obtain the number of cointegration relationships, the speed of adjustment coefficients (α_i) and the cointegrating vectors (β_i). In a final step, the remaining parameters are estimated conditional on the estimated $\hat{\beta}_i$ via OLS.

4.3 GVAR specification and estimation

4.3.1 Data selection and properties

The selection of the modelled countries and the variables included in the dataset draws extensively from the original work in Dees *et al*. (2007). The GVAR set-up contains 17 countries (see table 7) representing more than 70 per cent of world output and including eight euro area Members. Countries are all modelled individually, although the dynamic analysis of their Impulse Response Functions (IRFs) can pool together regions to make the interpretation of the economic shocks more meaningful, for example in a euro area rebalancing context.

---

[22] This corresponds to Case IV (unrestricted intercepts and restricted trend coefficients) in Johannsen terms. A situation without co-trending restrictions -Case III (unrestricted intercepts and no trend coefficients)- could also be defined, implying a long-run multiplier matrix $\Pi_i = \alpha_i {\beta_i}'$ that is not rank defficient.

Table 7: List of countries included in the GVAR

OECD				
European Union				Rest
Eurozone				
Austria	Sweden	Australia	Brazil	
Belgium	United Kingdom	Canada	China	
Finland		Japan		
France		Switzerland		
Germany		United States		
Italy				
Netherlands				
Spain				

The dataset includes as domestic variables part of the original selection used in Dees *et al*. (2007) –i.e. real output (y_{it}), the rate of inflation (Δp_{it}), the real bilateral exchange rate against the dollar (e_{it}), short-term (ρ_{it}^{S}) interest rates- and it is augmented, when available, with real household consumption expenditure (c_{it}), real gross fixed capital formation (i_{it}), catering for the aggregate demand side, and finally relative asset prices (ap_{it}) and private sector credit flows to GDP (pc_{it}), in order to capture potential spillovers in the housing market and cross-border lending.[23] All variables are seasonally adjusted.

The country-specific foreign variables are built using three sets of weights, in line with our original motivation: (*i*) traditional trade weights assessing the importance of bilateral trade flows; (*ii*) financial weights, represented by cross-border banking exposures coming from BIS international banking

[23] The sources of the original variables can be found in Dees *et al*. (2007) contribution and are also detailed in the GVAR toolbox manual. Consumption and Investment data comes the IMF International Financial Statistics database. The relative asset price index is taken from the BIS and includes information on equities as well as residential and commercial real estate covering all countries but Austria, Brazil and China. Finally, credit data is obtained from the database developed in Dembiermont *et al*. (2013). Data on Consumption, Investment and Credit is missing for both Brazil and China.

statistics;[24] and (*iii*) contagion weights, proxied by the 5-year trailing correlation of the long-term sovereign yields.[25]

The estimation of the country-specific models requires both the domestic and foreign variables to be integrated as this would allow for the existence of cointegrating long-run relationships. Standard Augmented Dickey-Fuller tests as well as the Park and Fuller (1995) variant (generally reported as having superior statistical power, see Leybourne *et al*., 2005) signal that all variables are integrated except for a few cases (e.g. domestic inflation in Australia and Switzerland). Moreover, there is mixed evidence on the order of integration of outstanding private debt, as it happens to be integrated of order 2 for a majority of countries.

Based on these results, the selected variables enter the country models in levels except for the price index and credit to GDP, whose first difference is used:

$$y_{it} = \ln(GDP_{it}/CPI_{it})\,;\ \Delta p_{it} = p_{it} - p_{it-1}, p_{it} = \ln(CPI_{it})\,;\ c_{it} = \ln(C_{it}/CPI_{it})$$

$$i_{it} = \ln(I_{it}/CPI_{it})\,;\ e_{it} = \ln(E_{it})\,;\ \rho_{it}^{S} = 0.25 * \ln\left(1 + \frac{R_{it}^{S}}{100}\right);$$

$$ap_{it} = \ln(ap_{it}/CPI_{it})\,, \qquad \Delta pc_{it} = pc_{it} - pc_{it-1}, pc_{it} = \ln(PC_{it}/GDP_{it})$$

4.3.2 Testing of the individual, country-specific models

The individual models are estimated over the period 1980Q1 to 2016Q4 by first determining the optimal lag for domestic and foreign variables according to the Akaike Information Criterion (AIC) and then determining the cointegration space using Johansen's trace statistic. As seen in Table 8, the estimation yields 43 cointegrating vectors and therefore 83 stochastic trends or unit roots (completing the 126 endogenous variables). Evidence of dynamic stability is found in the roots or eigenvalues of the GVAR, which all lie on or within the unit circle.

---

[24] As in Sun *et al*. (2013). The series cover the period 2005-2016. No data is available for Brazil and China and Finland only reports data from 2010 onwards.

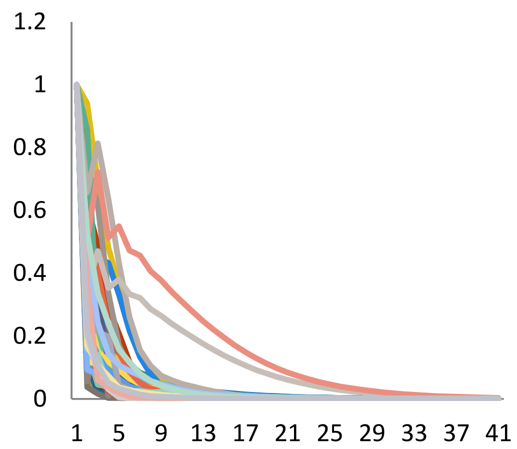
[25] See Garrat *et al*. (2016) for an alternative approach capturing confidence effects via surveys data.

A visual inspection of the persistence profile of the different cointegrating vectors is generally a good indicator of the performance of the model. Persistence profiles represent the effect of system or variable-specific shocks on the cointegrating vectors and should tend to 0 as the time horizon grows to infinity. As can be seen in Figure 22, all the persistence profiles converge to zero after a relatively short time, thus confirming the stability diagnosis. The weak exogeneity assumption of foreign variables is then tested along auxiliary first-difference regressions of x_{it}^{*} on the country-specific error correction terms. The assumption is violated in only 3 cases out of the 132 existing foreign variables.

Table 8. Order selection and cointegration space of individual models

	VARX (p, q)		Cointegrating relations
	p	q	
AUSTRALIA	1	1	3
AUSTRIA	1	1	2
BELGIUM	1	1	3
BRAZIL	2	1	2
CANADA	2	1	3
CHINA	2	1	3
FINLAND	2	1	2
FRANCE	2	1	3
GERMANY	2	1	3
ITALY	2	1	3
JAPAN	2	1	3
NETHERLANDS	2	1	3
SPAIN	2	1	3
SWEDEN	2	1	2
SWITZERLAND	2	1	2
UNITED KINGDOM	2	1	2
UNITED STATES	2	1	1

Figure 22. Persistence Profiles of the cointegrating vectors



Source of data: author's estimations.

4.4 Impact of the crisis on the Bilateral Systemic Spillover Potential

4.4.1 Optimal weighting of the different transmission channels

The estimation of the cross-country spillover effects from domestic economic shocks hinges on the specification of the transmission mechanisms. Moreover, the strength of the linkages between countries changes according to the selected transmission channels, which are themselves time-varying.

As reflected above, in a GVAR context the country-specific foreign variables are built by weighting the relevant domestic counterparts in all countries. For the weighting structure to reflect the optimal bilateral linkages between peers, the three transmission channels -trade, financial and contagion flows- are combined and weighted via the minimization of the short-term GDP forecast error of the model.

More specifically, the final set of bilateral links used to construct the foreign variables are defined as a weighted average of the three types of flows:

$$w_{ij} = \lambda_{trade} * trade_{it} + \lambda_{financial} * financial_{it} + \lambda_{contagion} * contagion_{it}$$

The optimal λ's for every period are defined as those minimizing the log-determinant of the matrix of summed cross-products of 1 to 4 steps-ahead out-of-sample forecasts errors. The log-determinant matrix for the s-step-ahead forecast is defined as follows:

$LD_s = log(|E_s|)$, with $E_S = \sum_{t=1}^{T}(\varepsilon_t^s * \varepsilon_t^{s\prime})$ being the summed cross-product of the forecast errors.

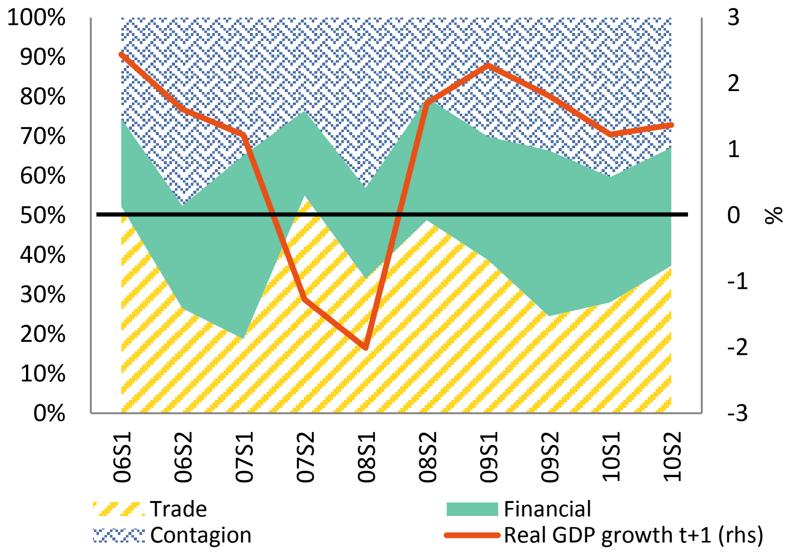
Their calibration for the 2006-2011 period is obtained by conducting a grid search throughout all existing combinations, with the sole restrictions that the λ's lie between 0 and 1 (non-inclusive) and they add up to 1; i.e. $\lambda_i \epsilon (0,1)$ and $\sum_{i=1}^{3} \lambda_i = 1$.[26] In order to avoid excessive volatility of the optimal λ's, the different semesters are considered as rolling windows for the forecast exercise and results are averaged over the different semesters and results are averaged over the first 5 best models (out of a total of 171 combinations) is kept.

Figure 23 shows the results for the crisis estimation period, altogether with the GDP growth rate of year t+1, representing the realized value that corresponds to the short-term out-of-sample forecast horizon at time t. It is important to notice that within the sample period, all three channels appear to be relevant when attempting at forecasting GDP. Their relative importance varies, however, alongside the economic cycle. Contagion flows appear to

[26] A jump parameter of 0.05 was considered when performing the grid search.

have been more relevant in times of economic distress, when economies become more autarchic and increase their reliance on internal savings.

Figure 23. Optimal weighting of the transmission channels and GDP growth



Source of data: author's estimations.

Note: Real GDP growth rate encompasses all 17 countries in the model.

4.4.2 Bilateral Systemic Spillover (BSS) Index

The calibrated structure of the different transmission channels enables the construction of optimal w_{ij} bilateral weights, linking countries via trade, financial and contagion flows. Following the seminal work by Pesaran *et al*. (2004), the indices or weighting matrices are built in two steps; (*i*) normalisation of the bilateral flows data, dividing them by the total amount of trade, financial and contagion flows, respectively; and (*ii*) construction of with w_{ij} bilateral links as a weighted average of the normalized flows, using the calibrated λ's, as defined in the previous section.

These indices represent thus the relative influence of a specific country j on a country i at a specific point in time, i.e. the potential for spillover effects from country j to country i. By construction, the aggregate influence of the rest of the world, defined as the sum of the bilateral links for country i with respect to the remaining j countries, will add up to one;

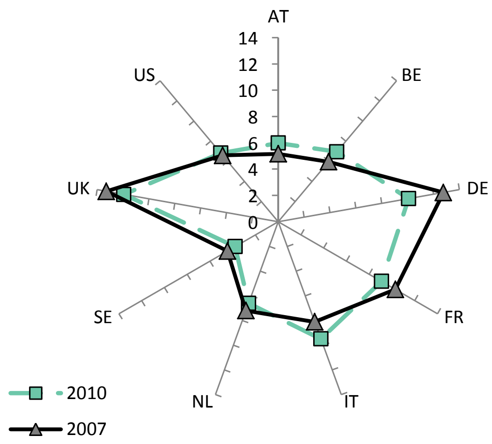
$W_i = \sum_{j=1}^{16} w_{ij} = 1\,, i \neq j$.

Figure 21 represents the optimal bilateral weights w_{ij} between Spain and some selected countries in 2007 and 2010.[27] These Inward Systemic Spillover Potential (ISSP) index reflects the relative loss of influence from Germany, France and the United Kingdom on Spain, which became relatively more dependent on the fate of the Italian, Belgian and Austrian economy. In absolute terms, however, the United Kingdom, Germany and France remain the three largest economic partners of the Spanish economy.

Inverting the optimal weight matrix allows us to gauge the influence of country i on selected j countries, i.e. w_{ji} represents the Outward Systemic Spillover Potential of country i. The example for Spain is very telling and can be seen in Figures 24 and 25. Throughout the crisis, Spain has become less influential and its systemic importance has shrunk considerably, except for Italy. These results are in line with a reduced exposure of euro area and UK banks to the Spanish economy and an increased link between the Italian and the Spanish sovereign bond markets. These results are particularly relevant in the context of the euro area rebalancing, as they show light on the reallocation of systemic relevance amongst core countries.

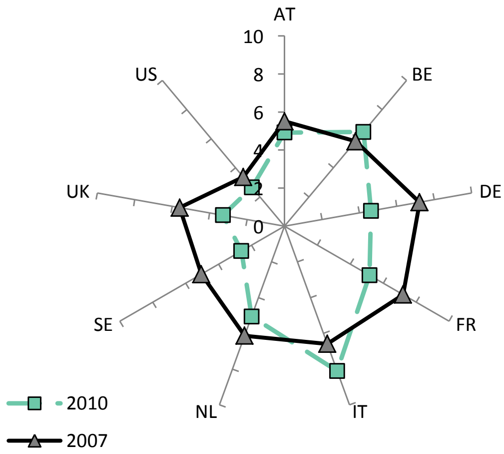
[27] The indicators could not be calculated thereafter due to data availability issues from the BIS asset price series.

Figure 24. Inward Systemic Spillover Potential (ISSP), selected countries towards Spain, %



Source of data: European Commission.

Figure 25. Outward Systemic Spillover Potential (OSSP), Spain towards selected countries, %



Source of data: European Commission.

4.5 Conclusion

The generation of cross-country spillover effects from idiosyncratic shocks ultimately depends on the relative strength of existing transmission channels. The bilateral links between two countries are not homogeneous across channels and thus concentrating on a specific conduit will ultimately yield a biased picture of the potential for spillovers between two economies.

To weight the relative strength of the existing interlinkages between countries, the transmission of shocks is circumscribed to three channels; namely trade flows, banking exposures and contagion via agents' perception (reflected in the co-movement of sovereign yields). Then the weighting scheme is obtained within a Global VAR framework, by minimizing the short-term GDP forecast error of the model. The three channels are found to be relevant in maximizing the forecasting accuracy of the model. Moreover, their relative strength is directly linked to the business cycle, as the crisis brought an increase in the relevance of contagion through bond markets.

Once the optimal relative weights of the channels are calibrated, they are used together with the actual bilateral flows to construct a weighted indicator reflecting the potential for spillovers between countries. Depending on the reference country, this indicator yields the inward (which countries are relatively more important for a specific economy, and to what extent) as well as the outward (which countries are more dependent on the evolution of a selected economy) spillover potential for a country.

These results shed some light on the reallocation of systemic relevance amongst countries and can be useful when calibrating processes such as the on-going rebalancing within the euro area. They provide, however, a static picture of the connection between two economies and have to be complemented with a dynamic analysis to reflect the strength of existing spillovers, which will also ultimately depend on the strength of the shocks at stake. In this vein, a natural extension of our results lies in using the dynamic properties of the GVAR model, taking our transmission channel matrix as a starting point and then deriving the dynamic interaction between countries through impulse response or forecast variance decomposition analysis, as in Greenwood-Nimmo *et al*. (2013).

4.6 References

Dées, S., di Mauro, F., Pesaran, M. H. and Smith, L.V. (2007) "Exploring the international linkages of the euro area: a Global VAR analysis," *Journal of Applied Econometrics*, vol. 22(1), 1-38.

Dembiermont, C., Drehman, M. and Muksakunratana, S. (2013). "How much does the private sector really borrow - a new database for total credit to the private non-financial sector," *BIS Quarterly Review*, March 2013.

Eickmeier, S. and Ng, T. (2011). "How do credit supply shocks propagate internationally? A GVAR approach," *European Economic Review*, vol. 74, 128–145.

Garrat, A., Lee, K. and Shields, K. (2016). "Forecasting global recessions in a GVAR model of actual and expected output," *International Journal of Forecasting*. Vol. 32(2), 374-390.

Greenwood-Nimmo, M.J., Nguyen, V.H. and Shin, Y. (2013). "Using global VAR models for scenario-based forecasting and policy analysis," in F. di Mauro and M.H. Pesaran (Eds.) *The GVAR Handbook: Structure and Applications of a Macro Model of the Global Economy for Policy Analysis*, pp. 97-113.

Gross, M. (2013). "Estimating GVAR matrices," ECB Working Paper Series, n. 1523, March 2013.

Leybourne, S., Kim, T.H. and Newbold, P. (2005). "Examination of some more powerful modifications of the Dickey-Fuller test," *Journal of Time Series Analysis*, vol. 26(3), 355-369.

Park, H. and Fuller, W. (1995). "Alternative estimators and unit root tests for the autoregressive process," J*ournal of Time Series Analysis*, vol. 16(4), 415-429.

Pesaran, M.H. (2015). "Time Series and panel data econometrics," Oxford University Press (eds.).

Pesaran, M.H., Shuermann, T. and Weiner, S.M. (2004), "Modeling regional interdependencies using a global error-correcting macroeconometric model," *Journal of Business and Economic Statistics*, vol. 22(2), 129-162.

Sun, Y., Heinz, F.F. and Ho, G. (2013). "Cross-country linkages in Europe: A Global VAR analysis," IMF Working Papers n. 194, International Monetary Fund.

5 SUMMARY AND FUTURE RESEARCH

This dissertation analyzes topics of special relevance for policy makers in the aftermath of the financial crisis.

Chapter 2 presents estimates of the slack of the economy in pseudo-real time according to the accumulation of macroeconomic imbalances. The analysis presents a novel approach putting the focus on the specification of the estimated model rather than on the prior selection of the methodology itself. Ideally, an agreeable method should achieve three necessary conditions: economic soundness, statistical goodness and transparency. On top of this, a sufficient condition is given by the *smell test*, often implemented by policymakers. In practice, fulfilling these conditions can prove to be challenging. Multivariate methods, coupled with Kalman filtering are generally considered amongst those reaching an acceptable level of compromise between these dimensions and thus are selected as a starting point, allowing for a combination of an economically-sound specification with a well-tested and flexible econometric procedure. The method fulfils the necessary criteria and allows for enough flexibility to get a country-specific approximation to the sufficient (smell test) criteria as it could accommodate specific cycles (financial, external, investment, fiscal, etc.) by incorporating additional variables related to the cycle. This somewhat eclectic approach is illustrated with its application to the Spanish economy, by selecting the best model amongst bivariate combinations of GDP and 52 accompanying variables.

Some conclusions can be drawn at this stage. First, there are some technical aspects that are important to be taken care of before jumping into the selection of the variables specification, such as: (i) modeling of GDP; (ii) cyclical prior of the accompanying series; (iii) transformation of the series (nominal vs. real, ratios vs. logs, etc.). Second, there is no clear algorithm for the selection of the variables to be included in the final specification.

Should it be an incrementalistic approach or rather a *brute force* consideration of all the alternative combinations? Third, this chapter has opted for the definition of necessary vs. sufficient conditions, although other combinations or weighting of the criteria might be possible.

Finally, future extensions of this work include an attempt at answering some of these open questions and providing a full assessment of the methodology in more complex data environments as well as technical improvements adding to the existing selection criteria, for example by estimating the contribution of the observables to the estimation of the output gap.

Chapter 3 provides a forecasting exercise comparing the out-of-sample forecasting performance of structural and non-structural models with quarterly data covering the last 36 years for seven macroeconomic aggregates: Gross Domestic Product (GDP), private consumption, private investment, employment or total hours worked, the GDP deflator, real wages and the nominal interest rate. The forecasting performance is assessed using a recursive procedure through four different dimensions: a time dimension (from one to eight quarters ahead), a contextual dimension (smooth growth periods and recession phase), a country-specific dimension (results for Spain, USA and the euro area) and a model-specific dimension (comparison against traditional benchmarks such as VARs and BVARs). All in all, there is supporting evidence for forecasting accuracy gains from structural models in the medium to long-run, while non-structural models perform generally better in the short-run. The benefits of structural models increase at all time horizons when considering disruptive times, with behavioral restrictions leading to higher parsimony. Indeed, during the ”Great Moderation” preceding the financial crisis, all models seemed to predict reasonably well at the different forecast horizons but this regularity was broken with the onset of the crisis as there is evidence of increases in the relative performance of DSGE models for all countries and all variables. It would thus seem that forecasters should beware of too stable periods as the performance of the different models might not reflect their accuracy in terms of capturing the underlying economic developments but rather the regularity of the latter. Moreover, Bayesian restrictions seem successful in shrinking the parameter space and providing better forecasting results. The results for the structural model, in turn, are robust across the different countries. Although the Smets and Wouters (2005) was initially tailored for the US economy (large and relatively closed), its forecasting gains with respect to non-structural models also apply to Spain and the euro area, despite their very different structures.

A complementary line of research would mimic these results in a non-linearised environment, to check for the influence of the log-linearisation process in watering out modeling refinements. It is also natural at this stage to wonder whether misspecification issues have a sizable impact on the forecasting performance of the DSGE models. The latest theoretical refinements might not prove worth the effort when the goal is not better knowledge of the transmission channels of the different shocks but simply forecasting.

Chapter 4 deals with the international transmission of shocks and more specifically, aiming at the calibration of the mian transmission channels. To calibrate the relative strength of the existing interlinkages between countries, the analysis first circumscribes the transmission of shocks to three channels; namely trade flows, banking exposures and contagion via agents' perception (reflected in the co-movement of sovereign yields). Then the weighting scheme is obtained within a Global VAR framework, by minimizing the short-term GDP forecast error of the model. The three channels are found to be relevant in maximizing the forecasting accuracy of the GVAR model. Moreover, their relative strength is directly linked to the business cycle, as the crisis brought an increase in the relevance of contagion through the sovereign bond markets. Once the optimal relative weights of the channels are calibrated, they are used together with the actual bilateral flows to construct a weighted indicator reflecting the potential for spillovers between countries. Depending on the reference country, this indicator yields the inward (which countries are relatively more important for a specific economy, and to what extent) as well as the outward (which countries are more dependent on the evolution of a selected economy) spillover potential for a country. These results shed some light on the reallocation of systemic relevance amongst countries and can be useful when calibrating processes such as the on-going rebalancing within the euro area.

The results provide, however, a static picture of the connection between two economies and must be complemented with a dynamic analysis to reflect the strength of existing spillovers, which will also ultimately depend on the strength of the shocks at stake. In this vein, a natural extension of our results lies in using the dynamic properties of the GVAR model, taking our transmission channel matrix as a starting point and then deriving the dynamic interaction between countries through impulse response or forecast variance decomposition analysis, as in Greenwood-Nimmo *et al*. (2013).