Forecasting Inflation with the Hybrid New Keynesian Phillips Curve: A Compact-Scale Global VAR Approach

Medel, Carlos A.

Central Bank of Chile

5 October 2015

Forecasting Inflation with the Hybrid New Keynesian Phillips Curve: A Compact-Scale Global VAR Approach*

Carlos A. Medel†
Financial Stability Area
Central Bank of Chile

October 5, 2015

Abstract

In this article, it is analysed the multihorizon predictive power of the Hybrid New Keynesian Phillips Curve (HNKPC) making use of a compact-scale Global VAR for the headline inflation of six developed countries with different inflationary experiences; covering from 2000.1 until 2014.12. The key element of this article is the use of direct measures of inflation expectations–*Consensus Economics*–embedded in a Global VAR environment, *i.e.* modelling cross-country interactions. The Global VAR point forecast is evaluated using the Mean Squared Forecast Error (MSFE) statistic and statistically compared with several benchmarks. These belong to traditional statistical modelling, such as autoregressions (AR), the exponential smoothing model (ES), and the random walk model (RW). One last economics-based benchmark is the closed economy univariate HNKPC. The results indicate that the Global VAR is a valid forecasting procedure especially for the short-run. The most accurate forecasts, however, are obtained with the AR and especially with the univariate HNKPC. In the long-run, the ES model also appears as a better alternative rather than the RW. The MSPE is obviously affected by the unanticipated effects of the financial crisis started in 2008. So, when considering an evaluation sample just before the crisis, the GVAR also appears as a valid alternative in the long-run. The most robust forecasting devices across countries and horizons result in the univariate HNKPC, giving a role for economic fundamentals when forecasting inflation.

JEL-Codes: *C22*; *C26*; *C53*; *E31*; *E37*; *E47*.

Keywords: *New Keynesian Phillips Curve*; *inflation forecasts*; *out-of-sample comparisons*; *survey data*; *Global VAR*; *time-series models*.

---

*The views and ideas expressed in this paper do not necessarily represent those of the Central Bank of Chile or its authorities. Any errors or omissions are responsibility of the author.

†E-mail: cmedel@bcentral.cl.

"[...] This is in defiance of the fine old saying that a difference is a difference only if it makes a difference."

–*How to Lie with Statistics,* Darrell Huff (1993), p. 58.

1 Introduction

Accurate forecasts have been always considered one of the key inputs for policymakers for both a conjunctural economic assessment and policy analysis. During the last decades, the increase of globalisation and the advent of powerful computing capacity for *big data* have pressured the development of macromodels explicitly including interaction terms of as many countries as possible. Especially after the collapse of Lehman Brothers bank in the US in 2008, the world witnessed how fast a country-level unexpected shock can be transmitted worldwide.[1] Some of these macroeconomic shocks involve disruptive real-economy wealth effects in countries that *a priori* seem isolated from the mainstream world trade, and with a minor role in world's financial market.

For the particular case of the monetary policy, the challenge of model external inflationary pressures has to deal also with the link between past and future domestic inflation rates.[2] This link reflects the traditional inertia exhibited by backward-looking price setters firms and a forward-looking component provided by rational expectations agent's behaviour–following Muth's (1961) traditional argument.[3] One successful proposal is the Hybrid New Keynesian Phillips Curve (HNKPC), introduced by Galí and Gertler (1999), analysed further in Galí, Gertler, and López-Salido (2001, 2005). To sketch its foundations, assume a staggered price-setting scheme *á la* Calvo (Calvo, 1983). Let $1-\theta$ the fraction of firms that change prices at a given period, and $1-\omega$ the fraction of firms that set prices optimally in a forward-looking manner. Hence, the HNKPC consists of a weighted average between past and future values of inflation plus a driving process $\widetilde{y}_t$, leading to the HNKPC baseline equation:

$$\pi_t = \gamma \widetilde{y}_t + \lambda_b \pi_{t-1} + \lambda_f \mathbb{E}_t[\pi^f_{t,t+h}] + \varepsilon_t, \tag{1}$$

where π_t is headline inflation, $\mathbb{E}_t[\pi^f_{t,t+h}] = \widetilde{\pi}_t$ is the inflation expectation at period f measured with a forecast made h-steps-ahead at period t, and $\widetilde{y}_t$ is a real marginal cost measure. $\{\gamma, \lambda_b, \lambda_f, \sigma^2_\varepsilon\}$ are parameters to be estimated, and ε_t is a cost-push shock, $\varepsilon_t \sim iid\mathcal{N}(0, \sigma^2_\varepsilon)$. This specification constitutes a reduced form coming from the optimisation problem of a structural NKPC where:

$$\begin{aligned}
\lambda_b &= \frac{\omega}{\phi}, \\
\lambda_f &= \frac{\beta\theta}{\phi}, \\
\gamma &= \frac{[(1-\omega)(1-\theta)(1-\beta\theta)]}{\phi}, \\
\phi &= \theta + \omega\left[1-\theta(1-\beta)\right],
\end{aligned} \tag{2}$$

where β is a discount factor. Note that Equation (1) results in a convenient specification for forecasting purposes and allowing many price settings.[4] Some forecasting exercises using an expression similar to

[1] A nice summary of this argument is presented in Bloom (2009) and empirically extended in Carriére-Swallow and Medel (2011) and Carriére-Swallow and Céspedes (2013).

[2] Note that this link–*i.e.* models with forward-looking feedback–is not especially circumscribed to the case of inflation. See Elliot, Granger, and Timmermann (2006) and Clements and Hendry (2011) for details on other processes.

[3] It could be argued that the sole inclusion of a forward-looking term in an inflation model turn consumer prices into a variable similar to an asset price, allowing for jumps. This fact found little empirical support. Hence, as Fuhrer (2011) argues, the inertia term primarily stands for a better fit to data.

[4] Some theoretical derivations of the HNKPC can be found in Smets and Wouters (2003, 2005), Christiano, Eichenbaum, and Evans (2005), Erceg and Levin (2003), and Collard and Dellas (2004), among others.

Equation (1) can be found in Nason and Smith (2008) for the US, and Jean-Baptiste (2012) for the UK case, among others reviewed later.

Galí and Monacelli (2005) develop an *open economy* version of the HNKPC which explicitly includes the interaction of a *domestic* country with the rest of the world. This is made through the real exchange rate and certain commodity prices in the output gap measure. The model is based on a richer economic environment but delivering a reduced form specification including domestic inflation and output gap also suitable for forecasting exercises. Nevertheless, there is neither a unique nor widely accepted manner in which a foreign component may be considered in the HNKPC. A comprehensive review of the open economy HNKPC can be found in Corsetti, Dedola, and Leduc (2010).

Despite the wide range of research conducted using the HNKPC with its many versions, some criticism still remains. Rudd and Whelan (2005) and Lindé (2005), for instance, claim that Galí and Gertler (1999) base their findings on a misspecified biased model. This is due to the simultaneous inclusion of the three base variables despite an estimation method especially controlling for simultaneity, *e.g.* the Generalised Method of Moments (GMM). Several solutions have been proposed regarding different estimations methods and specifications, but the debate remains open. One of the Rudd and Whelan's (2005) bottom line argument consists on the use of lagged inflation as a proxy of expected inflation. Hence, the endogeneity leads to biased estimations, as the authors argue. This is a key issue for this article, since there are used *direct measures–i.e.* exogenous–inflation expectations.[5] This article follows closely the Galí and Gertler (1999) and Galí, Gertler, and López-Salido (2005) view of the HNKPC.

Pesaran, Schuermann, and Weiner (2004) and Dees *et al.* (2007) have developed a special structural vector autoregressive (SVAR) modelling technique that captures explicitly the interaction between a domestic and a block of foreign economies. This is the so-called Global VAR (GVAR) methodology; fully described in Chudik and Pesaran (2014) and with supporting material freely available online. The GVAR methodology provides several advantages with respect to other alternatives. This is due to the possibility of estimating a set of equations considering particular econometric features for each country (including different variables with different lag length), to then stack all of them in one SVAR. This can be made using a predetermined exogenous weight scheme for each country that embraces the remaining countries of the VAR. This also alleviates the *curse of dimensionality* of traditional VAR estimations keeping the number of estimated coefficient at a minimum. Once stacked, the model is able to perform traditional econometric exercises such as impulse response functions and forecasting. Pesaran, Schuermann, and Weiner (2004) make use of 25 countries in a cointegration VAR-alike GVAR analysis, whereas Pesaran, Schuermann, and Smith (2009) make use of 33 countries to especifically focus on the forecasting ability of the GVAR technique.

The GVAR also brings a key feature for the NKPC estimation. As analysed in Dees *et al.* (2009) and Chudik and Pesaran (2014), the use of GVAR allows simultaneity for instrumental variables (IV) and the number of potential IV can be large. All these characteristics are certainly desirable in any forecasting device. Note that a previous development prior to the GVAR, it is a particular vector error correction model (VECM) augmented with foreign exogenous variables–referred to as VECX*. The VECX* obviously contains a long-run relationship between variables in levels and it is modelled the short-run adjustment to that long-run equilibrium. Nevertheless, to perform this kind of forecasts, several conditions must be fulfilled in regard of data stationarity. Some interesting forecasting results using the VECX* are presented in Garratt *et al.* (2006) and Assenmacher-Wesche and Pesaran (2008).

In this article, it is analysed the multihorizon predictive power of the HNKPC making use of a compact-scale GVAR for the headline inflation. The GVAR includes six developed countries (five countries plus a

[5] Some articles, such as Agénor and Bayraktar (2010), Mazumder (2010, 2011), Abbas and Sgro (2011), Lawless and Whelan (2011), and Vašíček (2011), supports the Rudd and Whelan's (2005) findings especially from a theoretical point of view.

region) spread across the world and exhibiting different inflationary experiences.[6] The analysed monthly sample covers from 2000.1 until 2014.12 (180 observations), divided in the *estimation* sample (2000.1-2005.12, 72 observations) and the *evaluation* sample (2006.1-2014.12, 108 observations). A special focus is given to the period 2006.1-2008.8 (32 observations; just before the financial crisis) given some atypical projections obtained with the GVAR; hence, evaluating it in *normal* times too. The analysed forecast horizons are h={1,6,12,24} months ahead. The driving process in this case, the marginal cost proxy variable, is the Hodrick-Prescott (HP)-based output gap with a treatment for the *end-of-sample* problem.

The key element of this article is the use of direct measures of inflation expectations embedded in a GVAR environment for inflation forecasting purposes. The expectations are taken from the monthly *Consensus Forecasting* report, being both the sample limiting element and the series defining the dependent variable stationary transformation, *i.e.* annual percentage change of the total Consumer Price Index (CPI). As a fixed-horizon prediction–for December of the current and the next year–a special adjustment is made.

The GVAR point forecast (henceforth referred to as GVAR) is evaluated using the Mean Squared Forecast Error (MSFE) statistic and statistically compared with several benchmarks using the Giacomini and White (2006) procedure. These benchmarks belong to traditional statistical modelling, such as autoregressions, the exponential smoothing model, and the random walk model (henceforth, AR, ES, and RW). One last economics-based benchmark is the univariate HNKPC, referred in the literature as the *closed economy* HNKPC (henceforth, HNKPC).[7]

The results indicate that the GVAR is a valid forecasting procedure especially in the short-run. This is the case for the Euro Zone, Japan, and Switzerland for h=6. Overall cases (countries and horizons), the most accurate forecasts are obtained with the AR and especially with the HNKPC. In the long-run, here corresponding to h={12,24}, the ES model also appears as a better alternative rather than the RW. When forecast errors are depicted in time, it is noticed that, especially at short-run horizons, the MSPE is mainly driven by the unanticipated effects of the financial crisis started in the US in 2008.9. To take this limitation into account, in the shortened evaluation sample the GVAR appears as a valid alternative to the RW also in the long-run for the US, the Euro Zone, Switzerland, and the UK for h=24, and the Euro Zone again at h=12. The most robust forecasting device across countries and horizons is the HNKPC, suggesting that there is a role for economic fundamentals when forecasting inflation.

The rest of the article proceeds as follow. In Section 2, it is reviewed the relevant literature concerning the many topics that confluent in this article. These are statistical versus economics-based inflation forecasts with uni- and multi-variate models, and the more recent GVAR predictions. It is also reviewed the macroeconomics of the NKPC. In Section 3, it is fully described the econometrics methods used for the GVAR and competing benchmarks. It is also defined the in-sample strategy to determine which will be the specifications used for prediction. Also, it is detailed the statistical inference carried out for the out-of-sample results. Finally, it is described the dataset and the building blocks of the output gap measure. In Section 4 there are presented the results divided in estimation diagnostics and forecast accuracy. Finally, Section 5 concludes.

2 Literature review

The quest of accurate inflation forecasts has a long tradition in macroeconometrics and central banking literature. Given that inflation typically presents a high level of persistence, close to a unit root, its

[6] These countries are the US (acting as a reference country), Canada (CAN), the Euro Zone (EUR; henceforth considered as a country), Japan (JPN), Switzerland (SWI), and the United Kingdom (UK).

[7] Nevertheless, it is virtually impossible to support that it is truly closed economy *estimation*. Despite that the *specification* is defined for closed economies, headline inflation contains embedded in virtually all its components, prices set or affected by international markets. This is specially the case of commodity prices in food and energy CPI components. See Neely and Rapach (2011) for an overview of CPI foreign fundamentals.

modelling has concerned many econometric issues with economic implications. There are two broad views of forecasting a macroeconomic variable–particularly visible for the case of inflation–: the atheoretical statistical manner, and the economics-based procedure.[8]

The atheoretical or statistical manner refers to the case when the prediction comes from a model without economic fundamentals, and the appropriate model is obtained purely based on statistical tests' results. Typical procedures included in this category are the autoregressive integrated moving average (ARIMA) family of models (Box and Jenkins, 1970), the RW, and the ES models (Hyndman *et al.*, 2008). There exist some more sophisticated versions including endogenous regime-switching parameters and nonlinearities. Belonging to the latter category it is found the Self-Exciting Threshold AR (SETAR)[9] model used, for instance, in van Ruth (2014) but with a little success (Calhoun and Elliott, 2012). Since the criticism of the many techniques relies on forecasting results, there is a huge part of the literature with proposals not superior to the existing simple time-series benchmarks. These proposals includes time-varying specifications, re-sampling computations, financial instruments-based data, bias-correction estimators, purposely mis-specified models, rule-of-thumb forecasts–*i.e.* the inflation target–and imported stuff from different fields that works at least combined with the existing procedures (Faust and Wright, 2014).

The majority of these models are used primarily as benchmarks, delivering fruitful results in a wide range of countries at any horizon. Some successful applications of these atheoretical models to the inflation forecasting case are Stock and Watson (1999), Atkson and Ohanian (2001), Giacomini and White (2006), Marcellino, Stock, and Watson (2006), Ang, Bekaert, and Wei (2007), and Elliot and Timmermann (2008) among others for the US case. They make use of different AR specifications using either the Akaike or Bayesian Information Criteria (AIC and BIC)–or both–, the RW, plus the IMA(1,1) and the equivalent single ES.

Groen, Kapetanios, and Price (2009) evaluate Bank of England's forecasts comparing with the AR and RW models. Andersson, Karlsson, and Svensson (2007) consider the same scheme to the Riksbank. Canova (2007) compares economics-based models for the G-7 with the RW and time-varying coefficients AR models. A comprehensive related work in which it is analysed a family of time-series forecasting models is Pincheira and Medel (2015). It provides robust evidence on the accuracy of statistical models making use of driftless extended seasonal ARIMA family (labelled DESARIMA). Their results comprise multihorizon forecasts for 11 inflation targeting countries plus the US.

A special case of atheoretical predictions are survey forecasts. They become atheoretical because the anonymity veil imposed to the respondents–and more important, to the manner in which they perform the forecasts–that turn the averaged consensus forecasts into an atheoretical forecast. Same as the ARIMA forecasts, these forecasts provide a limited simulation capacity for different policy scenarios. Nevertheless, several articles have pointed out the particular accuracy that they provide. Aiolfi, Capistrán, and Timmermann (2011) suggest that the combinations of these forecasts with other strategies, deliver substantial precision gains. Same results are found in Ang, Bekaert, and Weil (2007) and Pincheira (2012). This finding is relevant for a forecasting exercise like this since a restricted version of the HNKPC ($\lambda_b + \lambda_f = 1$) already consists of *á la* Granger combined scheme of an AR(1) and a survey forecast (Bates and Granger, 1969). More over, and assuming that the HNKPC already exists, some extra accuracy compared to the purely combination case, will necessarily be originated from the information contained in the output gap.

Despite the traditional statistical models, another branch of research analyse some more exotic specifications and estimation methods aiming to forecast inflation too. This is the case of, for instance, the Least Angle Regression (LARS; Efron *et al.*, 2004), Least Absolute Shrinkage and Selection Operator

[8] A recent survey of the many inflation forecasting methods can be found in Faust and Wright (2014).

[9] At this point it is easy to notice the complexities that the many acronyms are referring to. A reference to follow is Granger (1982), whereas especifically for ARMA models is Holan, Lund, and Davis (2010).

(LASSO; Bai and Ng, 2008), elastic net soft-thresholding (Bai and Ng, 2008), artificial neural networks (Choudharya and Haider, 2012), ridge regression (Groen, Paap, and Ravazzolo, 2013), copula methods (Charemza, Díaz, and Malakova, 2015), among others.

When inflation is forecast with economic models, the task is typically made with a Phillips Curve specification. Yet far from the original model of Phillips (1958), the basic foundation still remains. This is a trade off between an activity measure and a price level.[10] The HNKPC, however, includes more economic elements since it is derived from an optimisation problem in the style of modern macroeconomics. It was introduced by Galí and Gertler (1999) and extended in Galí, Gertler, and López-Salido (2001, 2003). Closer literature analysing the existence of the HNKPC can be found in Sbodrone (2002), Smets and Wouters (2003, 2007), Levin *et al.* (2005), and Rabanal and Rubio (2005).[11] Some articles using direct measures of expectations are Paloviita and Mayes (2005) using *Consensus Forecasts* for 11 European countries, Nason and Smith (2008) for the US–using the Survey of Professional Forecasters (SPF)–, Henzel and Wollmershauser (2008)–using CESifo World Economic Survey for Italy–, Paloviita (2009) for the Euro Area, and Medel (2015a) for Chile–using the Central Bank of Chile's SPF.[12]

The majority of the HNKPC estimations concern developed countries. Some examples are Roberts (1997) and Brissimis and Magginas (2008) for the US, Jean-Baptiste (2012) for the UK, McAdam and Willman (2003) for the Euro Area, Jondeau and Le Bihan (2005) for the UK and major Euro Area countries together, and Paloviita and Mayes (2005) for a panel of OECD countries. Some estimations with real-time datasets are available in Gruen, Robinson, and Stone (2002), Robinson, Stone, and van Zyl (2003) for the Australian case, and Paloviitta and Mayes (2005) for OECD countries.

Several articles analyse the out-of-sample behaviour of the HNKPC in different versions. Kichian and Rumler (2014) analyse the case for Canada using an open economy version, defining a marginal cost measure based on certain commodity prices. In the same vein, Rumler and Valderrama (2010) analyse the case of Austria, Balakrishnan and López-Salido (2002), Batini, Jackson, and Nickell (2005), and Posch and Rumler (2015) of the UK, Canova (2007) and Leith and Malley (2007) of G-7 countries, Rumler (2007) of Euro Area countries, and Mihailovic, Rumler, and Scharler (2011) of some OECD countries. The evidence is rather mixed between the HNKPC and time-series benchmarks. However, the simple AR model more than often results in a superior predictive method.

The GVAR methodology was introduced by Pesaran, Schuermann, and Weiner (2004) in search for a flexible procedure able to include key interactions across a big number of countries. The result is a specific SVAR that comes from stack country-level VARs previously defined in two blocks: the domestic and the foreign variables. The foreign variables enter into the domestic equation as weighted averages of the same variables defined for the remaining countries. As the weights are exogenously imposed–*e.g.* fixed known trade weights–it is easy to define first the model in a "compressed" manner, making possible its estimation, to then "decompress" it for further postestimation handling. The extensive form model eliminates any block of variables, treating every variable as part of an ordinary VAR (see Dennis and López, 2004, for a similar intuitive sketch description of the GVAR). Nevertheless, given the mechanics of the GVAR, it avoids the *curse of dimensionality* confronted by VAR models with too many coefficients to be estimated (and exponentially arisen when a new variable is included).

[10] An interesting exercise is conducted in Granger and Jeon (2011) where it is studied how the original Phillips Curve paper could be estimated with the time-series econometrics known 50 years later. This is made using the same original variables and sample, and providing some extensions for robustness.

[11] It is relevant to test the *existence* of the NKPC as some research suggests that must be flat in the (π_t, x_t) plane. See Kuester, Müller, and Stölting (2008) and the references therein for details.

[12] It is worth mentioning that the US economy has richer conclusions on this matter as it has several sources of survey expectations data with a long sample span, as is the case of the SPF of the Federal Reserve Bank of Philadelphia, the Livingstone Survey, the Michigan Survey, the Greenbook, CF, the Congressional Budget Office, and the Real-Time Data Set for Macroeconomists (Croushore and Stark, 2001).

Model's flexibility comes from the fact that it is possible to model a country-level VAR including specific variables and different lag length. This is permitted since the key issue of the GVAR is the stacking step; described in the next section. Notice that this also allows for multi-regional analysis (a group of countries) at the same stage with country level analysis. As a SVAR procedure, it provides the advantage of accommodate non-stationary series, compute cross-country impulse response functions, and forecasting.

Obviously, the GVAR is not the first attempt into explicitly linking world areas and countries, or disentangle the domestic economy in a sectorial manner. Some competing macromodels, comprising a large number of equations are Barrell *et al.* (2001), the National Institute's Global Econometric Model (NiGEM) for China and OECD countries, and several central bank's macromodels.[13] It is worth mentioning that, as all the available macromodels, the GVAR contains a rich documentation (di Mauro and Pesaran, 2013) plus a freely available *Matlab* platform containing a user-ready dataset for 33 countries.

At this point it is easy to notice the many outputs and research questions that just one GVAR estimation could provide. Hence, the potential applications by far outreach the exercises found in the literature. The introduction of the GVAR by Pesaran, Schuermann, and Weiner (2004) also provides an application estimating the effect of economic shocks on firm's conditional loss distributions using 25 countries grouped in 11 regions. For purposes similar to this article, Garratt *et al.* (2006) fully describe a macromodel for just one economy–the UK–but considering sectorial interactions. Certain forecasting equations are developed facilitating results interpretation. The book contains an interesting mixture of the previous VECX* model plus the GVAR, finding that the inclusion of long-run effects improves forecast accuracy, and the model is able to capture complicated short-run dynamics.

Dees *et al.* (2007) provide further development of the GVAR analysing special issues on modelling. For instance, if the GVAR includes the foreign exchange rate with the US economy as reference, a special treatment must be done. It also provides some convenient rules for stability checking such as the *persistence profiles*, of particular interest in long-run VECM-alike estimations. The article also discusses the forecast error variance decomposition, bootstrapped standard deviation for the impulse response function, and it carries out an application for the Fisher's interest rates equation. In the same spirit, Dees *et al.* (2007; DdMPS) provide a useful application of the GVAR when analysing the international linkages of the Euro Area. The authors make use and carefully describe the GVAR mechanics behind the estimated impulse response functions. The article is a clear exhibition of the many GVAR flexibility and available capacities.

Contained in the available GVAR handbook, two chapters are devoted to the particular task of forecasting. In Smith (2013), it is analysed a huge exercise for 134 variables from 26 regions made up of 33 countries, covering about 90% of world GDP. As the scale of the exercise is large and the heterogeneity of the countries is present, there is developed a special forecast accuracy assessment, averaging forecasts errors across horizons and regions. The article follows closely that previously published by Pesaran, Schuermann, and Smith (2009). Assenmacher (2013) describes the second forecasting exercise using the GVAR and follows closely the previous work Assenmacher-Wesche and Geissmann (2012) for the Swiss economy. The authors find considerable prediction gains specially in the short-run, compared to the case of the simple country-specific VAR(1) model.

Another forecasting application of the GVAR can be found in De Waal, Van Eyden, and Gupta (2015) for the South African economy. The authors make use of the baseline setup available online to analyse the accuracy of inflation and output forecasts comparing with some equally rich procedures (VECX*,

[13] As for instance the MONA of Denmark, EAGLE, ECB New Area Wide Model, and EURO-STING of the Euro Area, KITT of New Zealand, Ñ-STING of Spain, MOSES of Sweden, and IMF Global Economy Model and Federal Reserve Board SIGMA model for major world's economies.

Bayesian VAR, and traditional benchmarks). Their results support the GVAR as a good forecasting device in the long-run, being outperformed in the short-run by both the BVAR and the VECX*.[14]

A special attention is devoted to weights' estimation in Gross (2013) article. This development goes further to base GVAR original introduction. A major author's claim is that it is convenient to estimate them within the GVAR ensemble. This is because typically-used trade weights differ from those estimated, allowing for a chance to have biased estimation of the GVAR parameters. The author also argues that weights leading to unbiased estimators may result in a better prediction performance.

The exercise analysed in this article is considered of a compact rather than small scale simply because it includes countries spread in the world and with different inflationary experiences. It is kept at the minimum complexity to evaluate sharply the evolution of the GVAR accuracy. Note that, as Hansen (2009) argues, it is not clear the relationship between in-sample fit and forecast accuracy, but forecasts tend to be worst with overfitted models.[15] So, if the aim is to forecast a particular set of variables using the GVAR, it is preferred to include explanatory variables contributing to capture the variance of inflation series. These are not necessarily coming from countries exhibiting a high GDP level, which tends to show smoother macroeconomic dynamics.

3 Econometric setup

In this section all forecasting models are described: GVAR, HNKPC, AR, RW, and ES. It presents both kind of inflation data, actual and forecast, plus the construction of the output gap measure. As part of the methodological procedures used for out-of-sample statistical inference, it is defined the RMSFE Ratio and the Giacomini and White (2006; GW) testing procedure.

3.1 The Global VAR

For description purposes (following closely Pesaran, Schuermann, and Weiner, 2004), assume that there are i=0,1,...,N+1 countries across the time t=1,...,T, where the country i=0 is the reference country (the US). Now, assume that each country is modelled using k_i domestic and k_i^* foreign variables (hereafter, "*" will refer to foreign variables). In this article, for each country k_i=k_i^*=3, and hence k=6 (accounting: k_i=$\{\pi_{i,t-1}, \widetilde{\pi}_{it}, \widetilde{y}_{it}\}$ and k_i^*=$\{\pi_{i,t-1}^*, \widetilde{\pi}_{it}^*, \widetilde{y}_{it}^*\}$). So, for each country i it is defined the $k_i \times 1$ vector $\mathbf{x}_{it} = [\pi_{i,t-1}; \widetilde{\pi}_{it}; \widetilde{y}_{it}]'$ and the vector of order $k_i^* \times 1$ of foreign variables $\mathbf{x}_{it}^* = [\pi_{i,t-1}^*; \widetilde{\pi}_{it}^*; \widetilde{y}_{it}^*]'$, and hence a GVAR version of the HNKPC is:

$$\mathbf{x}_{it} = \mathbf{a}_{i0} + \mathbf{\Phi}_i \mathbf{x}_{i,t-1} + \mathbf{\Lambda}_{i0} \mathbf{x}_{it}^* + \boldsymbol{\varepsilon}_{it}, \tag{3}$$

where $\mathbf{a}_{i0}$ is a $k_i \times 1$ vector containing constants to be estimated, Φ_i is a $k_i \times k_i$ matrix containing lagged coefficients, Λ_{i0} is a $k_i \times k_i^*$ matrix containing the foreign variables relevant for the country i, and $\boldsymbol{\varepsilon}_{it}$ is $k_i \times 1$ vector of errors. Notice that Equation (3) could include more lags of the foreign variables vector, and it nests the VAR(1) if Λ_{i0}=...=Λ_{ip^*}=0. It is assumed that $\boldsymbol{\varepsilon}_{it} \sim iid(\mathbf{0}, \mathbf{\Sigma}_{ii})$, hence, errors are uncorrelated and with mean equal to 0. Note that $\mathbf{\Sigma}_{ii} = \mathbb{C}\left[\varepsilon_{ilt}, \varepsilon_{ist}\right]$ with $l \neq s$, and $\mathbf{\Sigma}_{ii}$ is nonsingular. This assumption could be easily relaxed for a spillover analysis with a long enough sample, since the elements of the diagonal must be estimated now. However, since $\mathbf{x}_{it}^*$ it is included in the estimation, $\boldsymbol{\varepsilon}_{it}$ already contain some foreign information.

The foreign variables included in $\mathbf{x}_{it}^* = [\pi_{i,t-1}^*; \widetilde{\pi}_{it}^*; \widetilde{y}_{it}^*]'$ constitute a weighted average of the same variable defined for the remaining N countries:

$$\pi_{it}^* = \sum_{j=0}^{N} \omega_{ij}^{\pi} \pi_{jt}, \;\; \widetilde{\pi}_{it}^* = \sum_{j=0}^{N} \omega_{ij}^{\widetilde{\pi}} \widetilde{\pi}_{jt}, \;\; \widetilde{y}_{it}^* = \sum_{j=0}^{N} \omega_{ij}^{\widetilde{y}} \widetilde{y}_{jt}, \tag{4}$$

[14] More evidence of similar economics-based procedures can be found in De Waal, Van Eyden, and Gupta (§2, 2015), and the references therein.

[15] See Medel (2015b) for some calibrated estimations of the effect of overfitting in the quality of the predictions, and Calhoun (2014) for a theoretical background.

where $\{\{\omega_{ij}^{\pi}\},\{\omega_{ij}^{\widetilde{\pi}}\},\{\omega_{ij}^{\widetilde{y}}\}\}_{j=0}^{N}$ is the set of N weights for each of the k_i^* foreign variables relevant for the country i. The simplest weight scheme is the equally-weighted average with $\omega_{ij}^{\pi}=\omega_{ij}^{\widetilde{\pi}}=\omega_{ij}^{\widetilde{y}}=1/N$, $\forall i \neq j$. If Equation (3) includes foreign exchange rate variable (e_t) using the US as a measure unit–being also the reference country i=0–then $e_{0t}^*=\Sigma_{j=1}^{N}\omega_{ij}e_{jt} - e_{it}$. Obviously, as the sequences $\{\omega_{ij}^{x}\}$ are weights, $\Sigma_{j=0}^{N}\omega_{ij}^{x} = 1$.

By now, Equation (3) represents a VARX*(1,1) model, *i.e.* a VAR(1) model including exogenous variables X*. So, the advantage of the GVAR method is that it actually models all the variables contained in the weighted average. Hence, it includes the N+1 variables $\mathbf{x}_{it}$. This is made by stacking all the countries into one equation using the predetermined weights. As the weights are known, it is possible to estimate the equations separately and then continue with the stacking step.

Define the next $(k_i + k_i^*) \times 1$ vector $\mathbf{z}_{it}$:

$$\mathbf{z}_{it} = \begin{bmatrix} \mathbf{x}_{it} \\ \mathbf{x}_{it}^* \end{bmatrix}. \tag{5}$$

Equation (3) could be rewritten as:

$$\mathbf{A}_i\mathbf{z}_{it} = \mathbf{a}_{i0} + \mathbf{B}_i\mathbf{z}_{i,t-1} + \varepsilon_{it}, \tag{6}$$

where $\mathbf{A}_i$ contains contemporaneous restrictions, $\mathbf{A}_i$=$[\mathbf{I}_k, -\mathbf{\Lambda}_{i0}]$, with $\mathsf{rank}(\mathbf{A}_i)$=$k_i$ and $\mathbf{B}_i$=$[\mathbf{\Phi}_i, \mathbf{0}]$. If the foreign variables are included with a lag, then its coefficient matrix $\Lambda_{i,t-1}$, will appear in $\mathbf{B}_i$ as $\mathbf{B}_i$=$[\mathbf{\Phi}_i, \mathbf{\Lambda}_{i,t-1}]$. A global vector $\mathbf{x}_t$ (suppressing the i-index) will be of the shape $\mathbf{x}_t = [\mathbf{x}_{0t}, \mathbf{x}_{1t}, ..., \mathbf{x}_{Nt}]'$, and the order in which the foreign variables enters into $\mathbf{x}_{it}$ and the stacking order is irrelevant. To have a view on the matrices involved, it is suggested to have a look at the $\mathbf{A}_i$ shape for the case considered in this article:

$$\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 & -\gamma_{ii}^{\widetilde{y}^*} & 0 & 0 \\ 0 & 1 & 0 & 0 & -\lambda_{ii}^{\pi^*} & 0 \\ 0 & 0 & 1 & 0 & 0 & -\lambda_{ii}^{\widetilde{\pi}^*} \end{bmatrix}. \tag{7}$$

Now, once that all the $\mathbf{x}_{it}$ vectors are already contained in the $\mathbf{z}_{it}$ vectors, it is easy to notice the following identity:

$$\mathbf{z}_{it} = \mathbf{W}_i\mathbf{x}_t, \tag{8}$$

where $\mathbf{W}_i$ (time-fixed) is a $(k_i + k_i^*) \times k$ matrix containing the known country-level weights. Pesaran, Schuermann, and Weiner (2004) label Equation (8) as the "link", as it links the country-specific model ($\mathbf{z}_{it}$) using all the global variables ($\mathbf{x}_t$). The shape of the $\mathbf{W}_i$ matrix when i=0 is the following:

$$\mathbf{W}_{i=0} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_{01}^{\widetilde{y}} & 0 & 0 & \omega_{02}^{\widetilde{y}} & 0 & 0 & \omega_{03}^{\widetilde{y}} & 0 & 0 & \omega_{04}^{\widetilde{y}} & 0 & 0 & \omega_{05}^{\widetilde{y}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \omega_{01}^{\pi} & 0 & 0 & \omega_{02}^{\pi} & 0 & 0 & \omega_{03}^{\pi} & 0 & 0 & \omega_{04}^{\pi} & 0 & 0 & \omega_{05}^{\pi} & 0 \\ 0 & 0 & 0 & 0 & 0 & \omega_{01}^{\widetilde{\pi}} & 0 & 0 & \omega_{02}^{\widetilde{\pi}} & 0 & 0 & \omega_{03}^{\widetilde{\pi}} & 0 & 0 & \omega_{04}^{\widetilde{\pi}} & 0 & 0 & \omega_{05}^{\widetilde{\pi}} \end{bmatrix},$$

and the 3×3 submatrix of zeros (below the 3×3 identity submatrix) is moving one block (of 3 columns) to the right when the country is changed across i=1,...,5.

Using the link equation in the country-specific model delivers:

$$\mathbf{A}_i\underbrace{\mathbf{W}_i\mathbf{x}_t}_{\mathbf{z}_{it}} = \mathbf{a}_{i0} + \mathbf{B}_i\underbrace{\mathbf{W}_i\mathbf{x}_{i,t-1}}_{\mathbf{z}_{i,t-1}} + \varepsilon_{it}, \tag{10}$$

and $\mathbf{A}_i\mathbf{W}_i$ and $\mathbf{B}_i\mathbf{W}_i$ are both $k_i \times k$ matrices. Stacking these equations yields:

$$\mathbf{G}\mathbf{x}_t = \mathbf{a}_0 + \mathbf{H}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{11}$$

where:

$$\mathbf{a}_0 = \begin{bmatrix} \mathbf{a}_{00} \\ \mathbf{a}_{10} \\ \vdots \\ \mathbf{a}_{N0} \end{bmatrix}, \ \mathbf{G} = \begin{bmatrix} \mathbf{A}_0\mathbf{W}_0 \\ \mathbf{A}_1\mathbf{W}_1 \\ \vdots \\ \mathbf{A}_N\mathbf{W}_N \end{bmatrix}, \ \mathbf{H} = \begin{bmatrix} \mathbf{B}_0\mathbf{W}_0 \\ \mathbf{B}_1\mathbf{W}_1 \\ \vdots \\ \mathbf{B}_N\mathbf{W}_N \end{bmatrix}, \ \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{0t} \\ \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{bmatrix}. \tag{12}$$

As $\mathbf{G}$ is a $k \times k$ matrix and of full rank generally, it is nonsingular allowing the GVAR representation:

$$\mathbf{x}_t = \mathbf{G}^{-1}\mathbf{a}_0 + \mathbf{G}^{-1}\mathbf{H}\mathbf{x}_{t-1} + \mathbf{G}^{-1}\boldsymbol{\varepsilon}_t, \tag{13}$$

which can be solved recursively as a SVAR(1) model. Note that the structure of the model is commanded by the $\mathbf{G}$ matrix, which contains no row-crossed terms. This allows to estimate each country-level equation separately, to then stack all the $\mathbf{A}_i\mathbf{W}_i$ results (numerically) in $\mathbf{G}$. This method provides the advantage of achieving a large number of countries (or regions) and allowing different specifications for each country. Some technical difficulties could arise when $\mathbf{G}$ is nonsingular. However, as Chudik and Pesaran (§6, 2014) suggest, the problem could be alleviated by including more lags of the foreign variables acting as an external unobservable factor.[16]

There are many results obtained from the estimation of Equation (13). For the particular purpose of this article, it is reported the point estimation across the evaluation window of the lagged inflation coefficient, mimicking in a dynamic way the *persistence profile* suggested in Dees *et al.* (2007) and De Waal, Van Eyden, and Gupta (2015). The residuals plots of each GVAR equation in the traditional diagnostics checking way are also shown.

A birds-eye summary of the GVAR derivation is presented in Table 1.

Table 1: GVAR derivation scheme (*)

*Step 1: Country-level VARX**
$\mathbf{x}_{it} = \mathbf{a}_{i0} + \boldsymbol{\Phi}_i\mathbf{x}_{i,t-1} + \boldsymbol{\Lambda}_{i0}\mathbf{x}^*_{it} + \boldsymbol{\varepsilon}_{it}$ $\boldsymbol{\varepsilon}_{it} \sim iid\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{ii})$
Step 2: Single country VAR(1) representation
$\mathbf{z}_{it} = \begin{bmatrix} \mathbf{x}_{it} & \mathbf{x}^*_{it} \end{bmatrix}$ $\mathbf{A}_i\mathbf{z}_{it} = \mathbf{a}_{i0} + \mathbf{B}_i\mathbf{z}_{i,t-1} + \boldsymbol{\varepsilon}_{it}$ $\mathbf{A}_i = (\mathbf{I}_{k_i}, -\boldsymbol{\Lambda}_{i0}),\ \mathbf{B}_i = (\boldsymbol{\Phi}_i, \mathbf{0})$
Step 3: Defining and using link equation
$\mathbf{z}_{it} = \mathbf{W}_i\mathbf{x}_t$ $\mathbf{A}_i\mathbf{W}_i\mathbf{x}_t = \mathbf{a}_{i0} + \mathbf{B}_i\mathbf{W}_i\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_{it}$ $\mathbf{G}\mathbf{x}_t = \mathbf{a}_0 + \mathbf{H}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$
Step 4: Stacking for the GVAR representation
$\mathbf{x}_t = \mathbf{G}^{-1}\mathbf{a}_0 + \mathbf{G}^{-1}\mathbf{H}\mathbf{x}_{t-1} + \mathbf{G}^{-1}\boldsymbol{\varepsilon}_t$ $\mathbf{G} = \begin{bmatrix} \mathbf{A}_0\mathbf{W}_0 & \mathbf{A}_1\mathbf{W}_1 & \ldots & \mathbf{A}_N\mathbf{W}_N \end{bmatrix}'$

(*) Source: Author's elaboration based in Pesaran, Schuermann, and Weiner (2004).

[16] From Equation (13), it is easy to note that the baseline GVAR constitutes a SVAR specification. This is precisely the argument elaborated by Dennis and López (2004) to criticise the policy analysis capacity of the GVAR. In principle, it allows limited dynamics (coming from the restrictions placed in $\mathbf{G}$), not leaving room for a policy instrument. This actually could be a setback for policymakers, but not of major interest when forecasting.

3.2 Benchmark models

3.2.1 Country-level Hybrid New Keynesian Phillips Curve

This subsection follows closely the description of the forecasting exercise for the Chilean inflation reported in Medel (2015a). The baseline specification is the univariate Equation (1). To avoid part of the simultaneity in the variables of the right hand side, Equation (1) is estimated with GMM. However, this method eliminates *methodological* simultaneity only, as the series exhibits a high correlation given their underlying data generating process. I make use of lagged observations of the same variables as IV. Recall that the problem that GMM addresses is the orthogonality condition $\mathbb{E}_t[\mathbf{x}_t'\varepsilon_t]$ that no longer holds. Hence, it is needed to "instrumentalise" the $\mathbf{x}_t'$ matrix with another one, say $\mathbf{m}_t$, containing ℓ IV ($\ell \geq k$) which fulfils:

$$\mathbb{E}_{t-1}[(\pi_t - \gamma\widetilde{y}_t - \lambda_b\pi_{t-1} - \lambda_f\mathbb{E}_t[\pi^f_{t,t+h}]]) \times \mathbf{m}_{t-1}] = 0. \tag{14}$$

In this context, a formal test for IVs' suitability is analysed through the Hansen's J-statistic:

$$J(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}_T) = \frac{1}{T}(\pi_t - \mathbf{x}_t'\widehat{\boldsymbol{\beta}})_t'\mathbf{m}\widehat{\mathbf{w}}_T^{-1}\mathbf{m}'(\pi_t - \mathbf{x}_t'\widehat{\boldsymbol{\beta}}), \tag{15}$$

where $\widehat{\mathbf{w}}_T$ is a $\ell \times \ell$ symmetric and positive-definite *weighting matrix*, as it weight the moments considered in the estimations. Hence, GMM finds the vector of coefficients:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{m}\widehat{\mathbf{w}}_T^{-1}\mathbf{m}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{m}\widehat{\mathbf{w}}_T^{-1}\mathbf{m}'\pi_t, \tag{16}$$

that minimises Equation (15). As $J(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}_T) \sim \chi^2_{\ell-k}$, along with the estimated coefficients it is also reported the p-value that test the null hypothesis: $\mathbb{E}_T[J(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{w}}_T)] = 0$. If p-value $> \alpha\%$, the IV are valid at the $\alpha\%$-level of significance, and the specification qualifies to be the forecasting model.

The estimation of the weighting matrix is made according to Hansen (1982) recommendation–the inverse of covariance matrix, *i.e.* $\widehat{\mathbf{w}}_T = \widehat{\mathbf{s}}^{-1}$, and avoiding potential autocorrelation with the Newey and West (1987) HAC method. The estimation of both covariance matrices–for the two stages: IV and final regression–is set in the same manner. The whitening lag specification is set automatic, to be selected according the BIC choosing in a maximum of 3 lags (following the rule $T^{1/3}$).

All the estimations are made through the GMM estimator to find a particular specification using the estimation sample, and following a *General-to-Specific* (GETS) strategy for the first stage regression. There are many reasons to prefer GMM as estimation method. First, and following Galí, Gertler, and López-Salido (2005), the GMM results are robust to the Non Linear IV GMM (NLIVGMM) estimator, which has been criticised by, for instance, Lindé (2005) and Rudd and Whelan (2005). This is a good reason to keep GMM since NLIVGMM estimation requires more computer time and it is more sensitive to the IV election in a univariate ensemble. However, to perform the forecasting estimations, it is used the Ordinary Least Squares (OLS) estimator following the same methodology used by Jean-Baptiste (2012) for the UK.[17] As emphasised by Cochrane (2001), the election between one (GMM) or another maximum likelihood estimator for univariate cases is a trade-off, and no consensus has been achieved.

3.2.2 Econometric time series models

Stationary autoregressions Alongside the RW, stationary AR models complements the most traditional benchmarks used for forecasting inflation as well as many other macroeconomic time-series (Ghysels, Osborn, and Rodrígues, 2006). The fitted models often includes a MA component (following the Box and Jenkins, 1970, model selection view); and so I refer to the ARIMA(p,1,0) particular case for simplicity. This also is due to the high persistence exhibited by inflation series, whose dynamics is well described by an AR(1) with a near-unity coefficient. More than often the inflation is measured

[17] Empirical results do not deliver substantial parameter differences between GMM and OLS.

with the annual percentage change of the CPI already seasonally adjusted. These transformations reduce the possibility of identifying (additive) seasonality, and MA terms could be neglected with ease as dynamics of the series is less complex. The literature of ARMA modelling applied to the inflation case is incommensurable. Some especially devoted surveys are Stock (2001) and Stock and Watson (2009).

The strategy used in this article consists simply of the estimation of Equation (17) across the different p integers using the estimation sample. In this case, using $p^{\max}$=s=12 (s=annual frequency of the series):

$$\pi_t = \overline{\pi} + \sum_{i=1}^{p\in P} \phi_i \pi_{t-i} + \varepsilon_t, \tag{17}$$

where $\{\overline{\pi},\{\phi_i\}_{i=1}^{p\in P},\sigma_\varepsilon^2\}$ are parameters to be estimated, $\varepsilon_t \sim iid\mathcal{N}(0,\sigma_\varepsilon^2)$, and P=$\{1,...,12\}$. For each "p"-model, it is computed the BIC whereas the forecasting model is that with the minor BIC score (reflecting the better adjustment to the true model given the sample size). The BIC is defined as $\mathsf{BIC}{=}{-2\mathcal{L}} + (1+p)\log(T)$, where $\mathcal{L}$ is the log-likelihood function, T the sample size, and $(1+p)$ is the number of coefficients of the model (accounting: one constant plus p AR coefficients).

Many articles analyse the appropriateness of information criteria for forecasting purposes.[18] Among the most used it is found the BIC, AIC, the Hannan-Quinn, and the Mallows Cp Criterion. However, at least these four are derived under the same Kullback and Leibler (1951) principle of *cross entropy*, delivering the same asymptotic results. In Medel and Salgado (2013), it is conducted a simulation exercise to analyse to what extent the BIC is *superior* to the AIC for estimation and forecasting; both strengths measured accordingly and tested jointly.[19] It is found that what is referred to as *asymptotically* or *long-sample* equivalence occurs with an unlikely-available sample span (around 83 years of monthly data).

The BIC produce more parsimonious (in-sample) results with intermediate sample size compared to the AIC. But, it is still unable to reject the null hypothesis of higher out-of-sample accuracy and parsimony jointly. Moreover, in Medel (2015b) it is found that the overfitting is hazardous for forecasting accuracy only when the number of parameters of the model exceeds at least the annual frequency of the series, *i.e.* when $p > s$. Hence, for the sake of parsimony, it is preferred the AR with BIC.

The estimation of the ϕ_i-coefficient(s) is made with the OLS method. This is in full acknowledgement of the downward bias that OLS provides for $\widehat{\phi}_i$ (see Lovell, 2008). Hence, it is not used any available bias-correction estimation as those of Andrews (1993), Andrews and Chen (1994), Hansen (1999), Kim (2003), among others. This option is left because, as shown in Pincheira and Medel (2012) and Medel and Pincheira (2015), among the competing models to the GVAR it is included the RW, which results in a superior alternative for near-unity series. As the RW is used as a numerary model to compare the RMSFE, it results in a demanding benchmark for the GVAR–recalling the aim of this article.

The random walk forecasts The RW consists of the special AR(1) case where ϕ is not estimated and it is restricted to ϕ=1 instead. This restriction, although simple, entails several methodological as well as economic consequences. The most significant impact is that it turns inflation into a non-stationary variable theoretically without available statistical inference and divergent predictions with the forecasting horizons. Due to this non-stationarity, it sounds unlikely–at least theoretically–to have room for stabilisation policymaking, since past unpredictable shocks do not vanish in time. Note that this argument is raised because inflation exhibits a unit root; hence, with a CPI∼I(2). For forecasting

[18] More details on derivation and comparison between AIC and BIC criteria can be found in Akaike (1974), Shibata (1976), Rissasen (1978), Schwarz (1978), Stone (1979), Lütkepohl (1985), Koehler and Murphree (1988), Zucchini (2000), Kuha (2004), and Weakliem (2004).

[19] The AIC is defined as $\mathsf{AIC}{=}{-2\mathcal{L}} + 2(1+p)$, hence differing with respect to the BIC in the "penalty term", reflecting a trade-off between more parameters and a higher log-likelihood score. It is hence expected that for a sample size of $T \geq 8$ and a given value of $\mathcal{L}$ that $p^{\mathsf{BIC}} \leq p^{\mathsf{AIC}}$.

purposes, it does not comprise a major setback since over-differentiation does not necessarily jeopardise the accuracy (Dickey and Pantula, 1987).

Yet the empirical evidence has been overwhelmingly in favour of the RW. This is due to the benefit of misspecification that more than offset the parameter uncertainty arisen from finite sample estimation. In particular, well-known Atkeson and Ohanian (2001) taunt conclusion in regard of forecasting inflation with three simple versions of the Phillips Curves pointed out: "*The likelihood of accurately predicting a change in the inflation rate from these three forecasts is no better than the likelihood of accurately predicting a change based on a coin flip.*" (Atkeson and Ohanian, 2001, Abstract). More evidence can be found in Canova (2007) for G-7 countries and the references therein.

In this article, it is used a *driftless* RW forecast, following the argument given in Pincheira and Medel (2012) and Medel and Pincheira (2015) that driftless RW-based forecast are unbiased. Iterating forward the AR(1) model we have:

$$\pi_{t+h} = \overline{\pi}\left[\frac{1-\phi^h}{1-\phi}\right] + \phi^h \pi_t + \sum_{i=0}^{h-1} \phi^i \varepsilon_{t+h-i}. \tag{18}$$

If π_t is model with a driftless RW, *i.e.* ϕ=1 and $\overline{\pi}$=0, the optimal forecast becomes π_{t+h}=π_t at any horizon. Hence, the h-step-ahead forecast error $\mathbb{E}\left[\varepsilon^{RW}_{t+h|t}\right]$=$\mathbb{E}\left[\pi_{t+h} - \pi^{RW}_{t+h|t}\right]$ satisfy:

$$\begin{aligned} \mathsf{Bias}_h &\equiv \mathbb{E}\left[\varepsilon^{RW}_{t+h|t}\right], \\ &= \mathbb{E}\left[\overline{\pi}\left[\frac{1-\phi^h}{1-\phi}\right] - (1-\phi^h)\pi_t + \sum_{i=0}^{h-1}\phi^i \varepsilon_{t+h-i}\right], \\ &= \overline{\pi}\left[\frac{1-\phi^h}{1-\phi}\right] - (1-\phi^h)\mathbb{E}\left[\pi_t\right], \\ &= 0, \end{aligned} \tag{19}$$

as $\mathbb{E}\left[\pi_t\right]$=$\overline{\pi}/(1-\phi)$. More details can be found in Medel and Pincheira (2015). The article also reports a simulation exercise confirming the notorious RW capacity even in non-Gaussian environments.

Exponential smoothing forecasts The ES corresponds *per sé* to a forecasting model. The version used in this article corresponds to the *single* ES, but there are available more specifications such as the *double* ES and the Holt-Winters model (see Hyndman *et al.*, 2008). The prediction for h-steps ahead is the same independently of the horizon:

$$\pi_{t+h|t} = \alpha \pi_{t-1} + (1-\alpha)\pi_{t-1+h|t-1}, \tag{20}$$

with $0 < \alpha \leq 1$. Note that if α=1, the ES coincide with the RW model. The model has been also used for forecasting purposes in Corberán-Vallet, Bermúdez, and Vercher (2011), Kolassa (2011), He, Shen, Tong (2012), and Pincheira and Medel (2015) with relative success for the same reasons of the RW.

3.3 Data

This subsection statistically described the dataset used in this article. There are two kinds of data: inflation time series and the output gap which is constructed using the Industrial Production (IP) index. The source of actual headline inflation and the IP of all countries is the *OECD Database*, whereas for inflation expectations is the monthly *Consensus Forecasts* (CF) report elaborated by *Consensus Economics*. In Appendix A, it is presented a more detailed summary of the data in its original format.

The whole sample span comprises from 2000.1 to 2014.12 (180 observations). For in-sample modelling diagnostic checking, the first six years of observations (2000.1-2005.12) are used, and the remaining part for evaluation purposes (108 observations; 2006.1-2014.12). As abovementioned, it is analysed the predictive ability of all the models with a *shortened evaluation sample* (2006.1-2008.8, 32 observations) for an analysis on model's behaviour prior to the crisis.

3.3.1 Inflation time series

As the six considered countries are developed, it is expected a similar dynamics during the sample. Note that the commodity prices boom of 2006-7 and the financial crisis of 2008-9 are included in the evaluation sample, making the task of forecasting more demanding. This is explicitly considered in this article using the *shortened evaluation sample*. This has to be considered when comparing with previous studies using a sample with smoother series.

The descriptive statistics of the series are presented in Table 1 for three samples. Actual inflation is transformed using the annual percentage change of the CPI. This is made to fit the specification used by the expectation series. CF survey is entirely reported for the same transformation (for inflation variable); even if CPI-basket re-definitions will be undertaken. The expectation series are also the limiting variable for the sample span, starting in 2000. Inflation and IP are available in a useful quality since 1960s (assuming a backward reconstruction for the Euro Area).

Notice that for the full sample, it is presented the Augmented Dickey-Fuller (ADF) testing for stationarity. According to the ADF test, the inflation series are stationary at 10% of confidence. As the GVAR makes use of a weighting scheme, in this article it is used those coming from the first principal component. These weights are obtained with the full sample, but do not change dramatically with the estimation sample. This is worth mentioning since a reliable forecasting exercise has to make use of the information conditional in the period in which is available.

For robustness, the exercise was re-do with an equally-weighted scheme delivering similar results. The factor loading reported includes the estimation with all the countries. Nevertheless, for each country-level estimation the weights are re-scaled in a leave-one-out manner to add to unity.

From Table 1 it is noticeable that the mean of the series are similar between countries, close to 2%, except for Japan and Switzerland. For the evaluation sample, and due to the major disruptions in 2008-9, inflation has decreased, except for Japan. Consequently, the standard deviation has slightly increased for all the countries also.

In Figure 1 there are presented the time series plot for both the level and the annual percentage change series. There are three salient features. The most obvious is the different dynamics in the CPI level of Japan, which seems already stationary. As abovementioned, the use of a stationary transformation of another already stationary series may not have an important deal for forecast accuracy nor out-of-sample inference (Dickey and Pantula, 1987).

A second feature is that for the six countries the dynamics on both types of series during the crisis of 2008-9 show a similar hump-shaped pattern (in level) and, consequently, a V-shape pattern in the annual change.

One last final distinctive feature is that the dynamics of inflation during the estimation sample, except for the case of Canada, is similar between countries and with little variation in the annual change. This is a relevant ingredient to take into consideration for the model evaluation, *i.e.* the ability to capture out-of-sample forecast with a variance higher to that of the estimation sample. This fact–forecasting with breaks–*per sé* represents a natural robustness check for any modelling strategy. This break also leads later to a careful analysis of the forecasting errors across time; at least to broadly compare which model made the best crisis tracking. It is found, and discussed later, that model's ranking changes in favour of the AR and the HNKPC when considering the crisis.

Table 2: Descriptive statistics of actual inflation series (*)

	US	*CAN*	*EUR*	*JPN*	*SWI*	*UK*	*US*	*CAN*	*EUR*	*JPN*	*SWI*	*UK*
	Actual Inflation (π_t)						Consensus Forecasts (weighted) ($\widetilde{\pi}_t$)					
Full sample: 2000.1-2014.12 (180 observations)												
Mean	2.382	2.014	2.002	-0.032	0.637	2.207	2.094	1.885	1.561	0.077	0.795	2.688
Median	2.343	2.084	2.125	-0.200	0.541	1.987	2.137	1.917	1.563	-0.049	0.875	2.676
Max.	5.654	4.745	4.114	3.707	3.083	5.355	3.431	2.695	2.421	2.334	1.908	4.193
Min.	-2.081	-0.905	-0.608	-2.524	-1.196	0.494	-0.448	0.575	0.307	-1.068	-0.218	0.574
Std. dev.	1.286	0.940	0.857	1.087	0.875	1.079	0.618	0.349	0.411	0.817	0.424	0.624
Skewness	-0.581	-0.185	-0.618	1.278	0.344	0.764	-1.291	-0.825	-0.400	1.219	-0.074	-0.394
Kurtosis	4.146	3.891	3.760	5.462	3.095	3.051	7.056	5.185	3.024	4.118	2.982	5.013
JB-Stat.	19.967	6.977	15.807	94.436	3.620	17.523	173.39	56.198	4.809	53.986	0.168	35.048
p-value	0.000	0.031	0.000	0.000	0.164	0.000	0.000	0.000	0.090	0.000	0.919	0.000
ADF-Stat.	-4.059	-4.068	-3.023	-2.652	-2.600	-2.533	-4.944	-4.845	-2.748	-2.608	-3.291	-3.556
p-value	0.002	0.001	0.035	0.085	0.095	0.109	0.000	0.000	0.068	0.093	0.071	0.008
FLoading	24.08%	20.67%	16.90%	5.38%	23.74%	9.23%	-	-	-	-	-	-
Estimation sample: 2000.1-2005.12 (72 observations)												
Mean	2.695	2.391	2.223	-0.480	0.964	1.340	2.213	1.968	1.398	-0.349	0.929	2.429
Median	2.836	2.369	2.162	-0.441	0.962	1.314	2.188	1.924	1.453	-0.294	0.959	2.422
Max.	4.707	4.745	3.206	0.797	1.887	2.448	3.061	2.695	1.822	0.228	1.297	2.835
Min.	1.103	0.680	1.620	-1.567	-0.106	0.494	1.570	1.247	0.307	-1.068	0.528	2.011
Std. dev.	0.826	0.858	0.287	0.436	0.479	0.450	0.320	0.299	0.244	0.322	0.191	0.183
Skewness	-0.108	0.434	0.691	0.080	-0.065	0.441	0.442	0.234	-1.594	-0.206	-0.409	0.146
Kurtosis	2.322	3.513	3.800	3.287	2.257	2.841	3.467	3.553	7.370	2.003	2.408	2.418
JB-Stat.	1.521	3.052	7.648	0.323	1.707	2.407	3.003	1.577	87.782	3.490	3.057	1.269
p-value	0.467	0.217	0.022	0.851	0.426	0.300	0.223	0.454	0.000	0.175	0.217	0.530
Evaluation sample: 2006.1-2014.12 (108 observations)												
Mean	2.173	1.763	1.854	0.267	0.418	2.785	2.014	1.830	1.670	0.361	0.705	2.860
Median	2.046	1.892	1.924	0.000	0.201	2.700	2.002	1.902	1.761	0.250	0.711	2.907
Max.	5.654	3.604	4.114	3.707	3.083	5.355	3.431	2.486	2.421	2.334	1.908	4.193
Min.	-2.081	-0.905	-0.608	-2.524	-1.196	0.539	-0.448	0.575	0.581	-1.023	-0.218	0.574
Std. dev.	1.484	0.911	1.058	1.275	1.005	0.985	0.745	0.370	0.462	0.918	0.506	0.745
Skewness	-0.331	-0.509	-0.174	0.717	0.880	0.510	-0.998	-1.086	-0.813	0.662	0.391	-1.032
Kurtosis	3.422	3.516	2.469	3.742	3.301	2.926	5.001	4.900	2.866	2.744	2.489	4.746
JB-Stat.	2.772	5.856	1.812	11.728	14.360	4.710	35.94	37.474	11.979	8.184	3.921	32.888
p-value	0.250	0.054	0.404	0.003	0.001	0.095	0.000	0.000	0.003	0.017	0.141	0.000

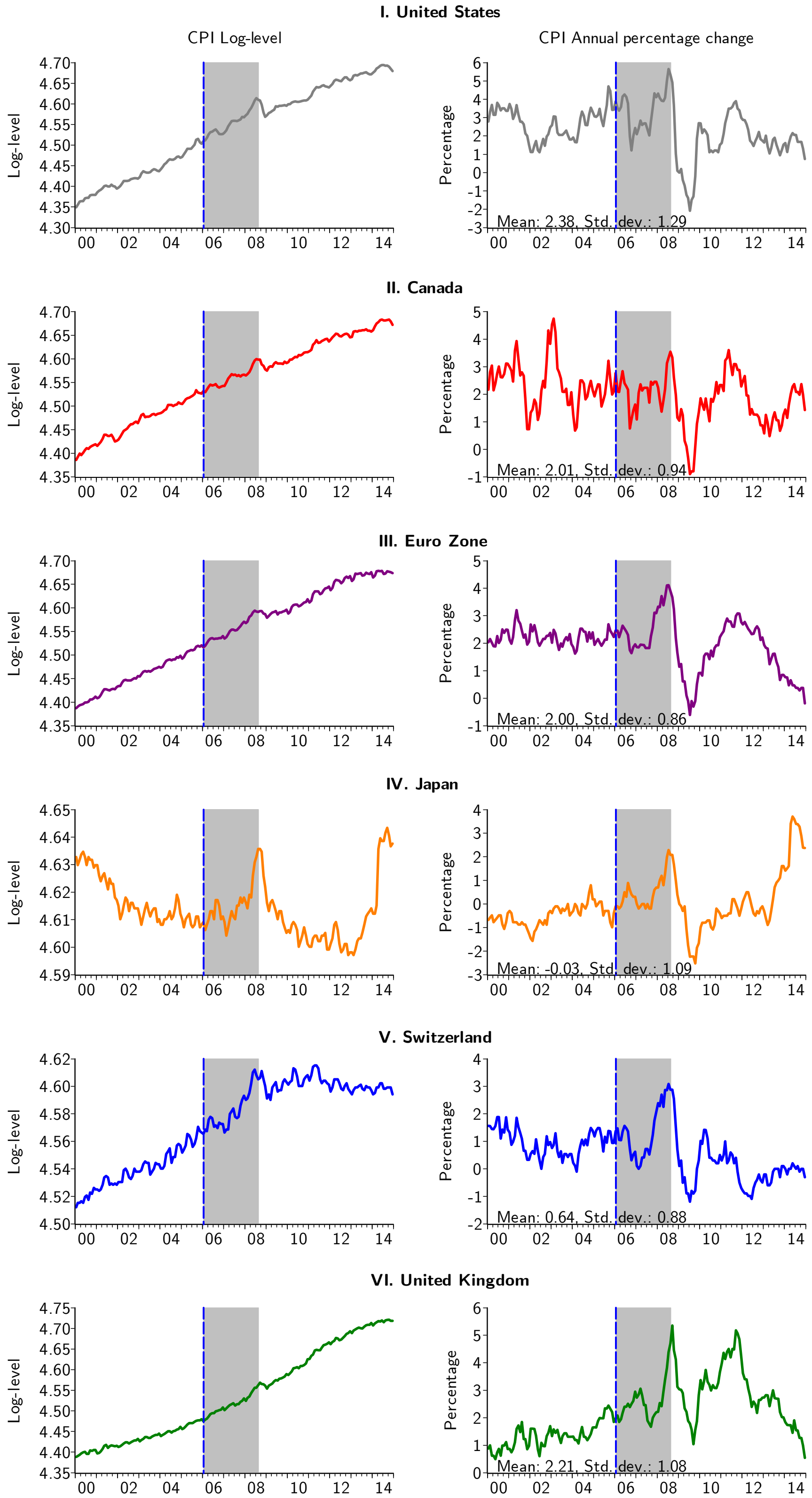
(*) "JB-Stat." stand for Jarque-Bera test statistic (NH: Data are random). "ADF-Stat." stand for Augmented Dickey-Fuller test statistic (NH: Series has a unit root). ADF equations for π_t includes a constant with 1 lag (US, CAN, SWI), 2 lags (UK), 7 lags (EUR), and 8 lags (JPN). ADF equations for $\widetilde{\pi}_t$ includes a constant and 12 lags (EUR, JPN, SWI, UK) and a trend with 3 lags (US, CAN). Source: Author's elaboration.

The CF expectations are reported monthly, providing the point forecast of 15-20 agencies and private consultants for several variables at two fixed horizons: December of the current and the next year. The names of the respondents are explicitly revealed along with their forecasts, making possible a one-by-one accuracy analysis. Given this specific richness of the survey, several articles make use of CF for testing economic/statistic hypothesis.[20]

However, as the estimation is made with constant frequency using recursive estimation, it is needed to adjust the series to have a unique rolling-event forecast. The approach used in this article is to create

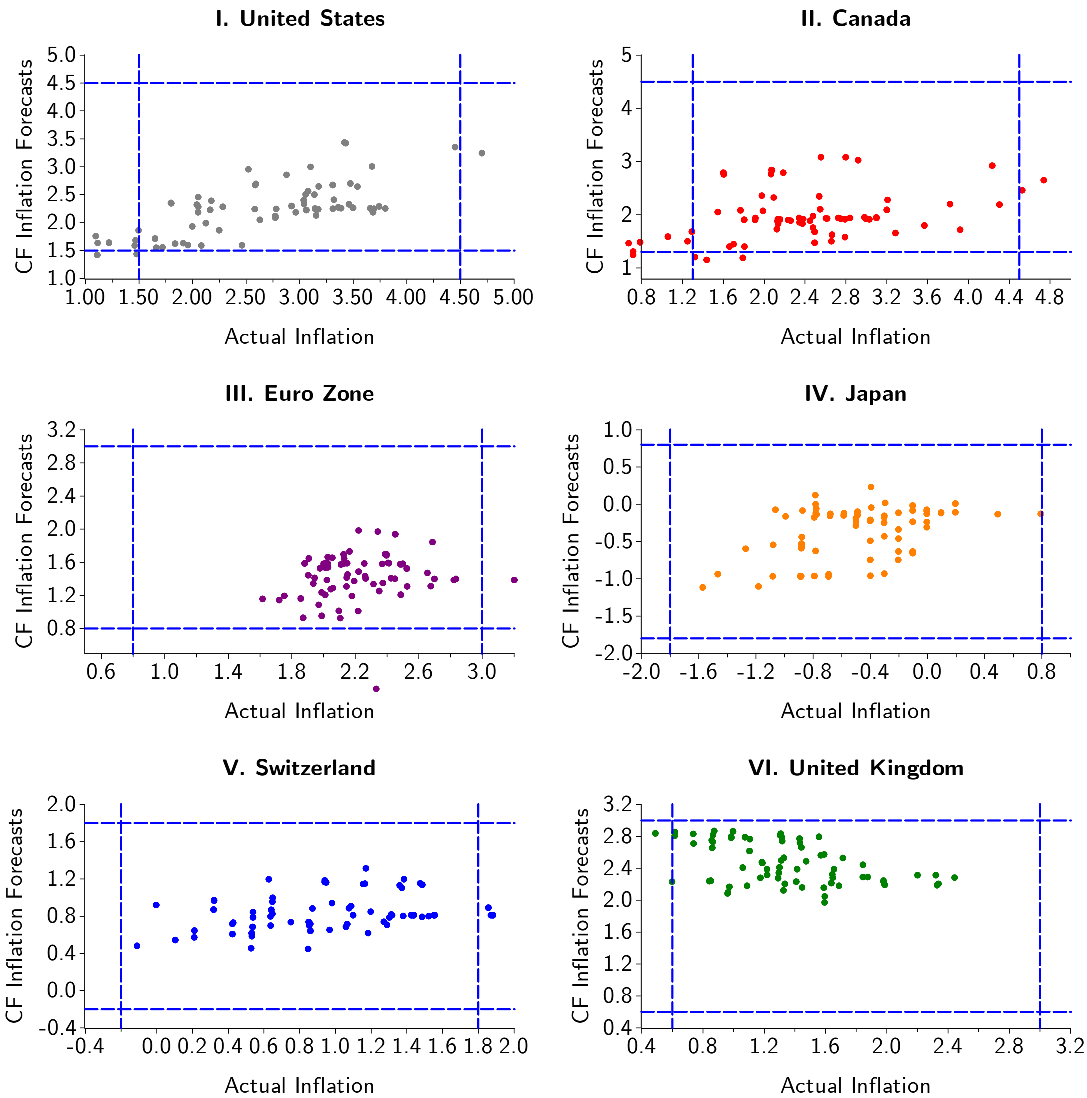
[20] Some examples are Loungani (2001), Ager, Kappler, and Osterloh (2009), and Pincheira and Alvarez (2009).

Figure 1: Actual inflation time series (*)



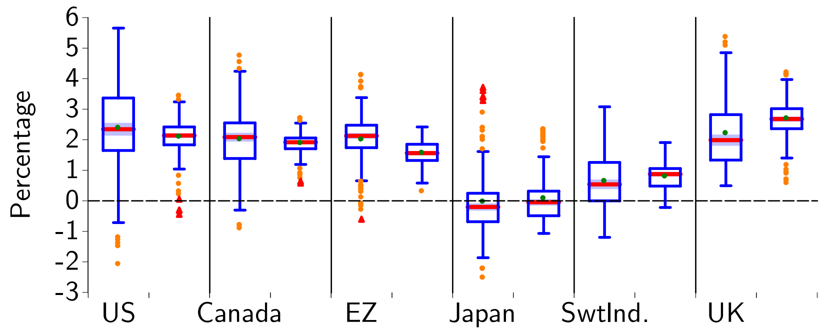
(*) Vertical line = evaluation sample start point. Shaded area = shortened evaluation sample. Source: Author's elaboration based on OECD database.

Figure 2: Scatter plot of CF inflation forecasts for December of current year (*)



(*) Source: Author's elaboration based on *Consensus Economics* data.

Figure 3: Descriptive statistics of actual inflation and weighted inflation forecasts (*)



(*) Source: Author's elaboration.

one series with a weighting scheme of the two forecasts in order to accommodate better the information to the targeted rolling-horizon. Hence, the CF forecast series for each month are weighted according to:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Current Dec	92%	83%	75%	67%	58%	50%	42%	33%	25%	17%	8%	0%
Next Dec	8%	17%	25%	33%	42%	50%	58%	67%	75%	83%	92%	100%

In Figure 2, it is presented the scatter plot between actual inflation and the CF for December of the current year. The result, despite that CF is already accurate for the fixed-horizon, is no longer useful in a rolling-event scheme since the majority of the observations lie outside the 45° line.

The last six columns of Table 1 show the descriptive statistics of the weighted CF series. In this case, and judging by points estimations (mean and median) the accuracy is notably improved across the sample. A more suitable way to visualise this is presented in the boxplots of Figure 3. In Figure 3 there are presented six pairs of boxplots, each pair showing first the actual and then the CF (weighted) statistics using the full sample. Note that the CF weighted series fulfils three desirable features in a forecast series: the mean (green dot) is close to the mean of the actual series, the volatility (proxied with the width of the blue box) is smaller than that of the actual series, and finally, CF exhibits less outliers (orange and red dots) than the target variable.

3.3.2 Output gap building blocks

One of the major drawbacks when estimating the NKPC is the impossibility to accurately measure the excess of demand–*i.e.* marginal costs. As the HNKPC and the GVAR make use of this measure, it is more challenging to have a stable series as new observations are added. The typical alternative to marginal cost variable is the output gap ($\widetilde{y}_t$)–*i.e.* the difference between the current and potential output.[21] As the estimations are made with monthly data, it is used the IP index as a proxy of the quarterly GDP. Table 3 presents the descriptive statistics of these series for all countries and for two sample spans: the estimation and the evaluation sample, for the annual percentage change (Δ^{12}) of the level series.

Note that the transformation achieves stationarity according to the ADF test. Given that the transformation lose the level information, the dynamics of IP for developing countries must not differ considerably. This is precisely the case described in Table 3, whereas some remarkable features are found in regard of the recent financial crisis. For all countries the mean has declined except for Switzerland. For Japan, the mean even reaches a negative value. Again excepting from Switzerland, the volatility has increased considerably, making more demanding a stable output gap variable. As expected, the range between the minimum and maximum values has considerably widen for the last part of the sample, and the minimum values achieves two up to three times that previously observed.

Basically, instability in the output gap arise with the "end-of-sample" problem of filtering, especially when the Hodrick-Prescott (HP) procedure is used to obtain the potential output: an unobservable component.[22] To alleviate this setback, I follow the approach proposed by Bobbitt and Otto (1990), Kaiser and Maravall (1999), and more recently re-launched by Mise, Kim, and Newbold (2005). This consists of adding forecast observations to level series prior to perform any filtering procedure. Hence, the method applied to obtain the output gap follows the steps of Figure 4. Note that the seasonal adjustment is made with X12-ARIMA in its default mode, and the filtering method is HP (λ=129,600).[23]

[21] Note that I focus on *output* gap instead of *unemployment* gap following the recommendations of Staiger, Stock, and Watson (1997a, 1997b).

[22] See Orphanides (2001), Orphanides and van Norden (2002, 2005) and Garratt *et al.* (2008) for a discussion on this matter.

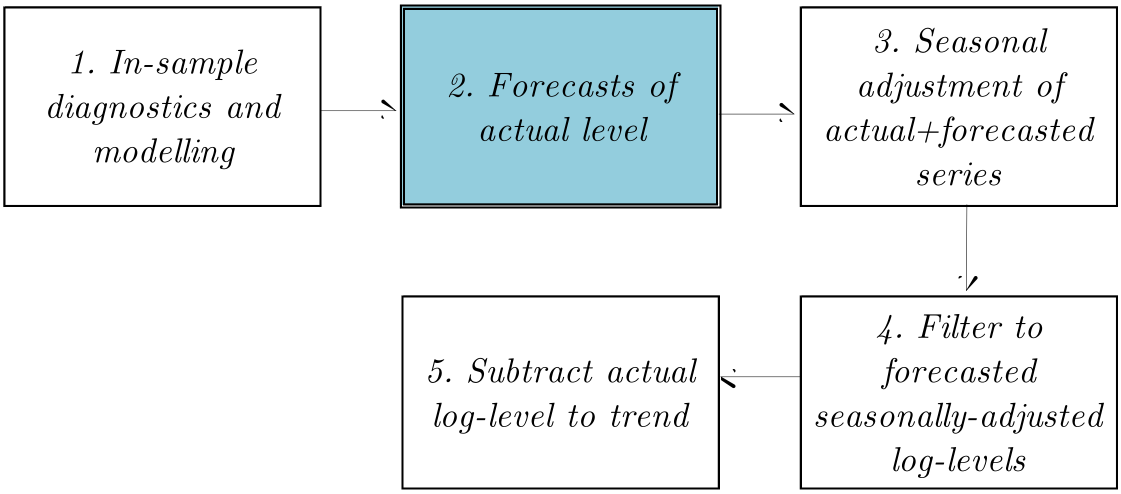
[23] Note that the X12-ARIMA seasonal adjustment method is based on a battery of moving average filters. This is to identify and decompose the series into a trend-cycle plus a remainder which contains the seasonal component plus an irregular component (Findley *et al.*, 1998). This procedure also starts with several diagnostics tests (contained in the *RegARIMA* module) to eliminate statistically undesirable anomalies. The next step is forecast to filter the forecasted series. As these predictions are subject to error, the output gap measure includes instability due to this *methodological* setback. These distortions could be substantial as reported in Cobb and Medel (2010) and Medel and Pedersen (2010).

Table 3: Descriptive statistics of Industrial Production series (*)

	US	*CAN*	*EUR*	*JPN*	*SWI*	*UK*	*US*	*CAN*	*EUR*	*JPN*	*SWI*	*UK*
	Industrial Production (y_t)											
	Estimation sample						*Evaluation sample*					
Mean	1.282	1.669	1.494	1.389	0.852	-0.340	1.120	0.082	0.026	-0.174	2.370	-1.062
Median	2.180	1.440	1.290	2.736	0.833	-0.549	2.672	0.465	1.449	1.905	2.339	-0.183
Max.	5.475	11.368	7.288	7.794	6.625	3.693	8.550	8.602	9.333	27.320	6.801	5.131
Min.	-5.686	-8.824	-3.748	-12.762	-4.603	-4.775	-15.074	-14.832	-21.671	-33.333	-2.687	-11.142
Std. dev.	2.883	4.229	2.456	5.267	2.455	1.825	5.402	5.064	6.673	10.734	1.861	3.702
Skewness	-0.916	0.051	0.127	-1.260	0.101	0.114	-1.775	-1.030	-1.444	-0.718	0.020	-1.110
Kurtosis	2.948	3.430	2.649	3.661	2.603	2.868	5.311	4.000	4.889	4.812	2.691	3.964
JB-Stat.	10.081	0.585	0.564	20.374	0.594	0.208	80.727	23.600	53.571	24.052	0.438	26.356
p-value	0.006	0.746	0.754	0.000	0.743	0.901	0.000	0.000	0.000	0.000	0.803	0.000
ADF-Stat.	-3.620	-3.868	-2.631	-2.902	-2.630	-3.697	-	-	-	-	-	-
p-value	0.031	0.015	0.089	0.047	0.089	0.005	-	-	-	-	-	-

(*) "JB-Stat." stand for Jarque-Bera test statistic (NH: Data are random). "ADF-Stat." stand for Augmented Dickey-Fuller test statistic (NH: Series has a unit root). ADF equation includes a constant and 12 lags (EUR, JPN, SWI, UK), and a trend and 3 lags (US, CAN), using the full sample. Source: Author's elaboration.

Figure 4: Output gap building blocks (*)



(*) Source: Author's elaboration.

The ARMA forecasting model for IP corresponds to $\Delta^{12}y_t = c + \phi\Delta^{12}y_{t-1} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_{12}\varepsilon_{t-12} + \theta_1\theta_{12}\varepsilon_{t-13}$, with $\varepsilon_t \sim iid\mathcal{N}(0, \sigma_\varepsilon^2)$. This is the so-called *airline model* (Box and Jenkins, 1970) which has proved to be a model that fit macroeconomic data with substantial success (Ghysels, Osborn, and Rodrígues, 2006). The in-sample estimations are presented in Table 4, which also reveals robust results across countries, and a correct specification according to the Durbin-Watson statistic, defined as DW$=\Sigma_{t=2}^{T}(\varepsilon_t - \varepsilon_{t-1})^2/\Sigma_{t=1}^{T}\varepsilon_t^2 \approx 2(1-\rho_\varepsilon)$, where ρ_ε is errors' autocorrelation.

Several articles use output gap as a proxy of marginal costs, differing often on the way how to obtain de-trended output (whether based on HP or other filtering device. See Pollock, 2014, for a review of some filtering techniques available in macroeconometrics). The economic rationale behind this measure is striking; it considers the distance between the current state of the economy and the counterfactual that may be obtained if all factors were employed in the absence of shocks. Some examples using output gap are Rudebusch and Svensson (1999), Stock and Watson (1999), Galí, Gertler, and López-Salido (2005), Lindé (2005), Paloviita and Mayes (2005), Rudd and Whelan (2005), Canova (2007), Dees *et al.* (2009), Nunes (2010), and Jean-Baptiste (2012), among others. Moreover, Batini, Jackson, and Nickell (2005) use output gap alongside the labour share on the basis of an endogenously determined price mark-up.

Stock and Watson (1999) suggests that especially when the aim is to forecast, the output gap measure provides a convenient alternative since relies basically on a univariate ensemble. Also, some of the major problems associated with output gap–instead of using marginal cost–are rather an empirical issue. The forecasts provided by the models of Table 4 tackle part of the "end-of-sample" problem.

Table 4: In-sample diagnostics of IP forecasting models (*)

	US	*CAN*	*EUR*	*JPN*	*SWI*	*UK*
	Dependent variable: $\Delta^{12} y_t$					
	Estimation sample					
ϕ	0.056	-0.166	-0.320	-0.381	-0.578	-0.376
	[0.658]	[0.054]	[0.001]	[0.000]	[0.000]	[0.000]
θ_1	-0.808	-0.933	-0.852	-0.900	-0.869	-0.811
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
θ_{12}	0.552	0.465	0.539	0.455	0.534	0.552
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Constant	0.106	0.055	0.133	0.101	0.051	-0.020
	[0.026]	[0.384]	[0.000]	[0.066]	[0.275]	[0.676]
$\overline{R}^2$	0.154	0.299	0.273	0.286	0.465	0.217
S.E. Reg.	0.516	0.868	0.703	1.075	1.389	0.961
DW Stat.	1.950	2.017	2.167	1.731	2.219	2.042

(*) Equation: $\Delta^{12} y_t = c + \phi \Delta^{12} y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_{12} \varepsilon_{t-12}$ $+ \theta_1 \theta_{12} \varepsilon_{t-13}$ with $\varepsilon_t \sim iid\mathcal{N}(0, \sigma^2_\varepsilon)$. Coefficient p-value in $[\cdot]$.
"DW stat." stand for the Durbin-Watson statistic.
Source: Author's elaboration.

3.4 Out-of-sample assessment

The statistical measure used to evaluate the accuracy of point forecast is the RMSFE:

$$\mathsf{RMSFE}_h = \left[\frac{1}{T} \sum_{t=1}^{T} (\pi_{t+h|t+h} - \pi_{t+h|t})^2 \right]^{\frac{1}{2}}, \tag{21}$$

where $\pi_{t+h|t}$ is the h-step-ahead forecast of $\pi_{t+h|t+h}$ made at period t. Note that this statistic is computed given a forecasting horizon h, and hence, the difference $T - t$ is variable depending on h–*i.e.* $T = T(h)$. To make a more plausible comparison with the RW, the analysed statistic corresponds to the RMSFE Ratio defined as:

$$\mathsf{RMSFE\ Ratio} = \frac{\mathsf{RMSFE}_h^{\mathcal{M}}}{\mathsf{RMSFE}_h^{RW}}, \tag{22}$$

where $\mathcal{M}$=$\{\mathsf{GVAR,HNKPC,AR,ES}\}$. Hence, as the RW acts as a pivot, values greater than unity imply a worse performance of the competing model. Figures below unity represent a "predictive gain" of (1-RMSFE Ratio)% compared to the RW.

Note that this evaluation is specifically made by "country×variable" forecast elements (the identifier is unique). Nevertheless, from the GVAR it is possible to evaluate the predictive accuracy of all the variables comprising a single country, a region, or a set of variables (where the "country×variable" elements are no longer unique). Obviously, several pooling techniques could be used to evaluate all or sets of forecasts coming from the GVAR. This issue is analysed in Granger and Jeon (2007) suggesting different ways on how to analyse accuracy. A feature remarked by the authors, is that considering the risk involved in the prediction, *e.g.* standard deviation of the forecasting errors, must be considered. This element is considered later in this article. Note that the pooling evaluation techniques hide particular patterns of the GVAR forecasting ability. For instance, the GVAR could be statistically better than other competing model due to a particularly well job at forecasting GDP, the oil price, exchange rate, or any other variable different from inflation. Hence, all the conclusions arisen at least from these articles, must be read conditional to the case of inflation (and a compact-scale estimation). Indeed, the results presented in, for instance, Ericsson and Reisman (2012) entail a high degree of heterogeneity across countries and variables.

Ericsson and Reisman (2012) provide an insightful treatment on the GVAR evaluation. The authors make use of the impulse indicator saturation (Hendry, Johansen, and Santos, 2008) to evaluate the forecasting ability of the GVAR, given the narrow relation between *parameter constancy* and forecasting ability. This last relationship is especially emphasised by Ericsson and Reisman (2012), and raised as one of the features to be improved by the GVAR. The results of this article share this view in the sense that it is found a low degree of parameter constancy across the sample, resulting in a limited predictive ability. This drawback, as Ericsson and Reisman (2012) argue, is due to an incomplete model selection criterion embedded in the GVAR identification, often providing non-robust results.

To investigate to what extent the predictive gains are statistically significant, I make use of the unconditional t-type test of Giacomini and White (2006) providing the advantage of comparing *forecasting methods* instead of *forecasting models*. As the null hypothesis (NH) is defined as *the competing model has a superior predictive ability compared to the* RW, there is used a one-side t-type GW statistic accordingly.

Formally, it is tested the NH: $\mathbb{E}_t(d_h) \leq 0$, against the alternative AH: $\mathbb{E}_t(d_h) > 0$, where:

$$\mathsf{d}_h = (\pi_{t+h|t+h} - \pi^{RW}_{t+h|t})^2 - (\pi_{t+h|t+h} - \pi^{\mathcal{M}}_{t+h|t})^2, \tag{23}$$

using the Newey and West (1987) HAC estimator of the standard deviation of d_h. The NH is rejected if the subsequent t-statistic is greater than $t_{\alpha\%}$; corresponding to the tabulated value of a normal distribution with probability $\alpha\%$.

4 Results

4.1 In-sample results

This section analyses both kinds of results: in- and out-of-sample. For forecasting purposes. it is used a recursive sample scheme. The in-sample estimations comprise just the estimation sample. The estimations were made with an *ad-hoc* program in *Eviews 8* making use of the `VARForecast` add-in.

4.1.1 Global VAR diagnostics

The GVAR comprise the estimation of 3 models for 6 countries; hence, 18 equations. The used lag length criterion is the BIC, delivering one lag for all countries. Note that, as Ericsson and Reisman (2012) argues, the GVAR could be too permissive in the way how to identify the model leading to unstable parameters, which is analysed later.

Table 4 reports some goodness-of-fit statistics of the GVAR for reference, *e.g.* to be compared with further estimations. As abovementioned, the log-likelihood score is one of the inputs for the AIC and BIC, differing remarkably in this case because the penalty term.

It is needed to say that when the objective of an economic model is to forecast, the diagnostic checking must be done accordingly. This implies giving less relevance to the particular estimated parameter size or even significance as the idea is not to test the economic theory behind the specification (see Kostenko and Hyndman, 2008, for a discussion on this matter). However, the model must fulfil certain stability conditions to deliver computable (in their moments) forecasts. A short-cut for model's suitability are their residuals behaviour.

In Figure 5, all the residual series by countries and variables are presented. Each panel also provides the $\overline{R}^2$ and the standard error of model's residual. There are some remarkable facts to analyse. First, all models regressions presents well-behaved residuals, easily associable to a white noise behaviour and with a few outliers.

Second, the adjustment according to the $\overline{R}^2$ statistic shows a good explanatory power of the GVAR. For actual inflation, except for the Euro Area, the $\overline{R}^2$ ranges from 0.68 to 0.80. For the output gap, same good fit is noticeable for five countries–excluding Switzerland–with $\overline{R}^2$ values reaching 0.93. For CF series it is found the best adjustment according to this measure, with an average $\overline{R}^2$ of 0.86.

Third, regarding the adjustment of the CF series, two facts unadverted by the forecasters are found. These occur in 2001 for the US, Canada, Japan, Switzerland, and the UK, and in 2003 for the Euro Zone. Nevertheless, and due to the interaction terms contained in the GVAR, these errors do not provide further disruptions in the fit of the model.

Inspired in the persistence profile analysis introduced in Dees *et al.* (2007), in Figure 6 it is presented a dynamic version of the *persistence profile*. This is simply the computation of the first inflation lag coefficient across the evaluation sample (recalling that the estimation is made recursively). Despite its significance, the deep interest put into a nonlinear evolution of the parameter: if λ_b exceeds or not $|1|$.

The most harmless results are observed for Japan, Switzerland, and the UK. In the case of Canada, it is adverted a shrinkage in parameter size starting in 2007 and finishing in late-2010, to then exhibit a hump shape for a period less than two years. This reveals a sort of hysteresis in the Canadian inflation dynamics coincident with the commodity prices boom in 2007-8 which deserves further research. For the Euro Zone, it is noticeable a hump-shaped coefficient dynamics during both the commodity prices boom and the financial crisis. This case as well as the US case are the only cases reporting λ_b greater than unity.

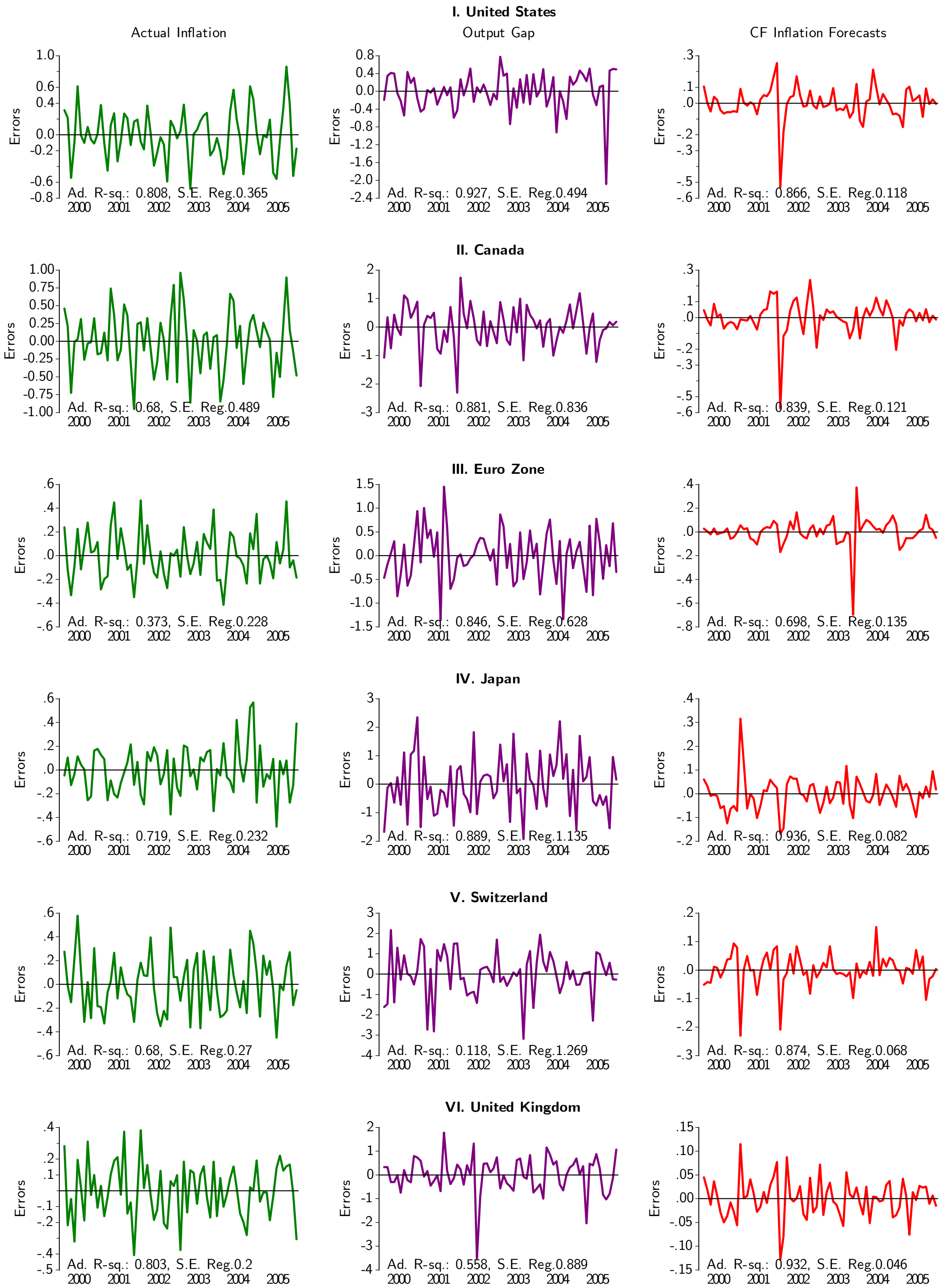
A figure like that of the Euro Zone is, to some extent, shared with the results of the AR diagnostics, achieving the maximum number of permitted lags. This is common when facing turbulences that make the autocorrelation function more complicated.

For the US case, it is noticeable some relatively stable estimates until 2013, to then jump above unity, achieving a peak or 1.25. Note that this is not indicative that this equation represents a unit root nor explosive behaviour, since the autocorrelation function of π_t is influenced by the (stationary) driving process (the output gap; see Fuhrer, 2011, for details on inflation persistence measures).

Overall, the estimated GVAR presents appropriate characteristics for forecasting purposes, and when comparable, are similar to those of Dees *et al.* (2009). Obviously, it is desirable more stable estimations if some other computations are required, especially the forecast error variance decomposition.

A complementary diagnostic check could be the impulse-response functions. Nevertheless, its interpretation must be done considering in the light of macroeconomic theory and not necessarily going through forecast accuracy. A case where not necessarily more NKPC-related economic theory embedded in the econometric setup redounds in forecast accuracy using a NKPC is Posch and Rumler (2015).

Figure 5: GVAR residuals time series. Estimation sample (*)



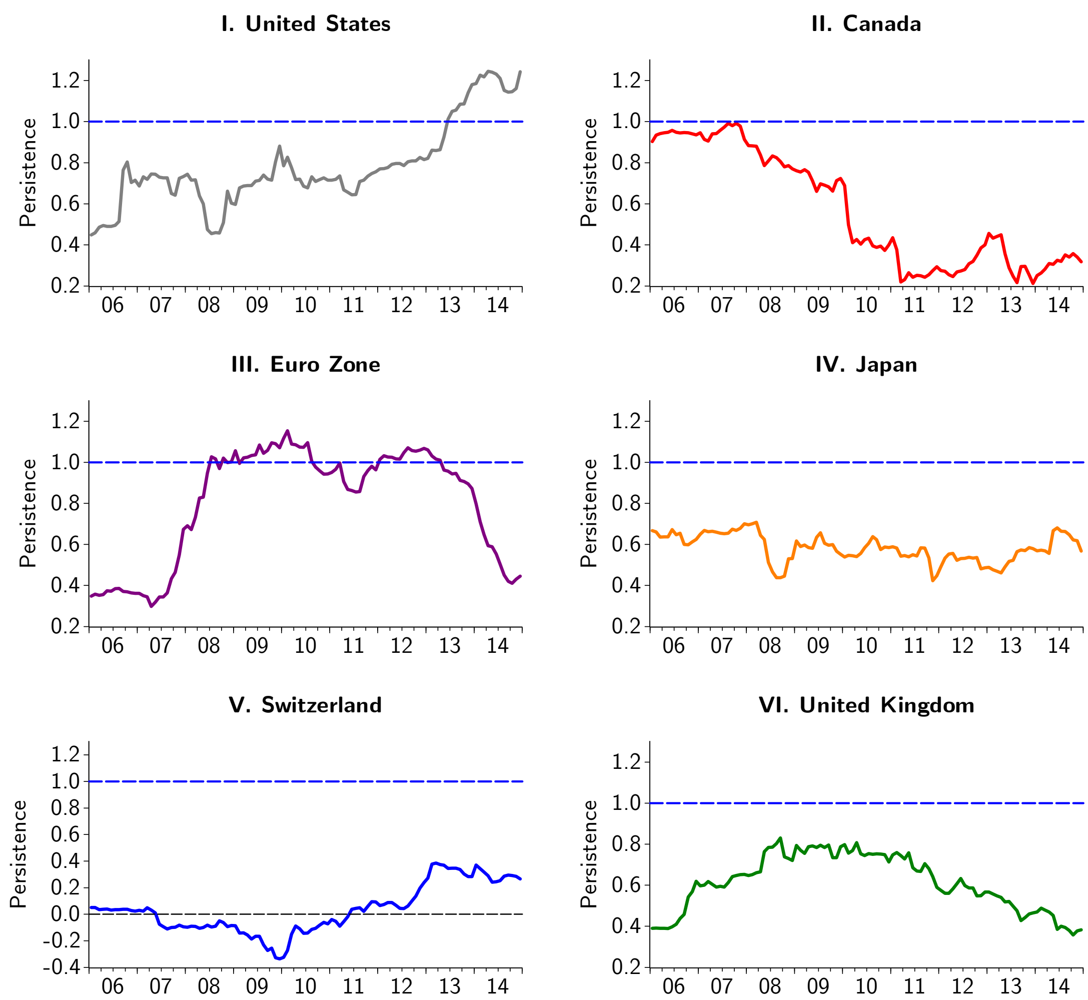
(*) Source: Author's elaboration.

Table 5: GVAR in-sample diagnostics (*)

Estimation sample	
Determinant resid covariance (dof adj.)	9.30×10^{-25}
Determinant resid covariace	3.42×10^{-27}
Log likelihood	350.0017
Akaike Information Criterion	-0.2254
Schwarz Information Criterion	10.6737

(*) Source: Author's elaboration.

Figure 6: GVAR dynamic persistence profile. Evaluation sample (*)



(*) Source: Author's elaboration.

4.1.2 HNKPC diagnostics

The in-sample results for the HNKPC entail the estimation of two sequential regressions: first stage for the instrumentalised variables, and the second stage using the results of the first step. It is with the second step regression that the inference and forecast are made.

The coefficient estimates are presented in Table 6. Note that these results are presented for the same specification across the countries. The only difference comes in the first stage regression, using different IV-sets that are reported in the lower panel. As IV, it is always used a constant and the second inflation lag. The results indicate that the coefficients exhibit the expected size and sign, except for the output gap for Switzerland.

Table 6: GMM estimates of the HNKPC (*)

	US	*CAN*	*EUR*	*JPN*	*SWI*	*UK*
	Dependent variable: π_t					
	Estimation sample					
π_{t-1}	0.905	0.436	0.750	0.906	0.762	0.852
	[0.000]	[0.013]	[0.071]	[0.000]	[0.000]	[0.000]
$\widetilde{\pi}_t$	-0.641	0.841	-1.370	-0.299	0.808	-0.478
	[0.058]	[0.061]	[0.058]	[0.064]	[0.060]	[0.031]
$\widetilde{y}_t$	0.082	0.132	0.150	0.033	-0.240	0.065
	[0.009]	[0.000]	[0.032]	[0.019]	[0.028]	[0.073]
$Constant$	1.555	-0.312	2.483	-0.133	-0.509	1.387
	[0.003]	[0.625]	[0.002]	[0.005]	[0.114]	[0.017]
$\overline{R}^2$	0.699	0.602	-0.760	0.678	0.314	0.712
S.E. Reg	0.458	0.553	0.390	0.249	0.392	0.238
DW Stat.	1.167	0.975	0.932	1.823	1.635	1.913
J-Stat.	6.730	0.448	0.382	2.455	1.304	0.243
p-value	0.080	0.503	0.536	0.117	0.253	0.621
	Instrumental variables list (lags)					
$Constant$	✓	✓	✓	✓	✓	✓
π_{t-p}	(2), (4)	(2)	(2), (3)	(2)	(2)	(2), (3)
$\widetilde{\pi}_{t-p}$	(4)	(1), (4)	(8)	(2)	(3)	(1)
$\widetilde{y}_{t-p}$	(1), (6), (12)	(4)	(4)	(2), (9)	(2), (5)	(7)

(*) Equation: $\pi_t = c + \lambda_b\pi_{t-1} + \lambda_f\widetilde{\pi}_t + \gamma\widetilde{y}_t + \varepsilon_t$, with $\varepsilon_t \sim iid\mathcal{N}(0, \sigma_\varepsilon^2)$.
See notes in Table 4. Source: Author's elaboration.

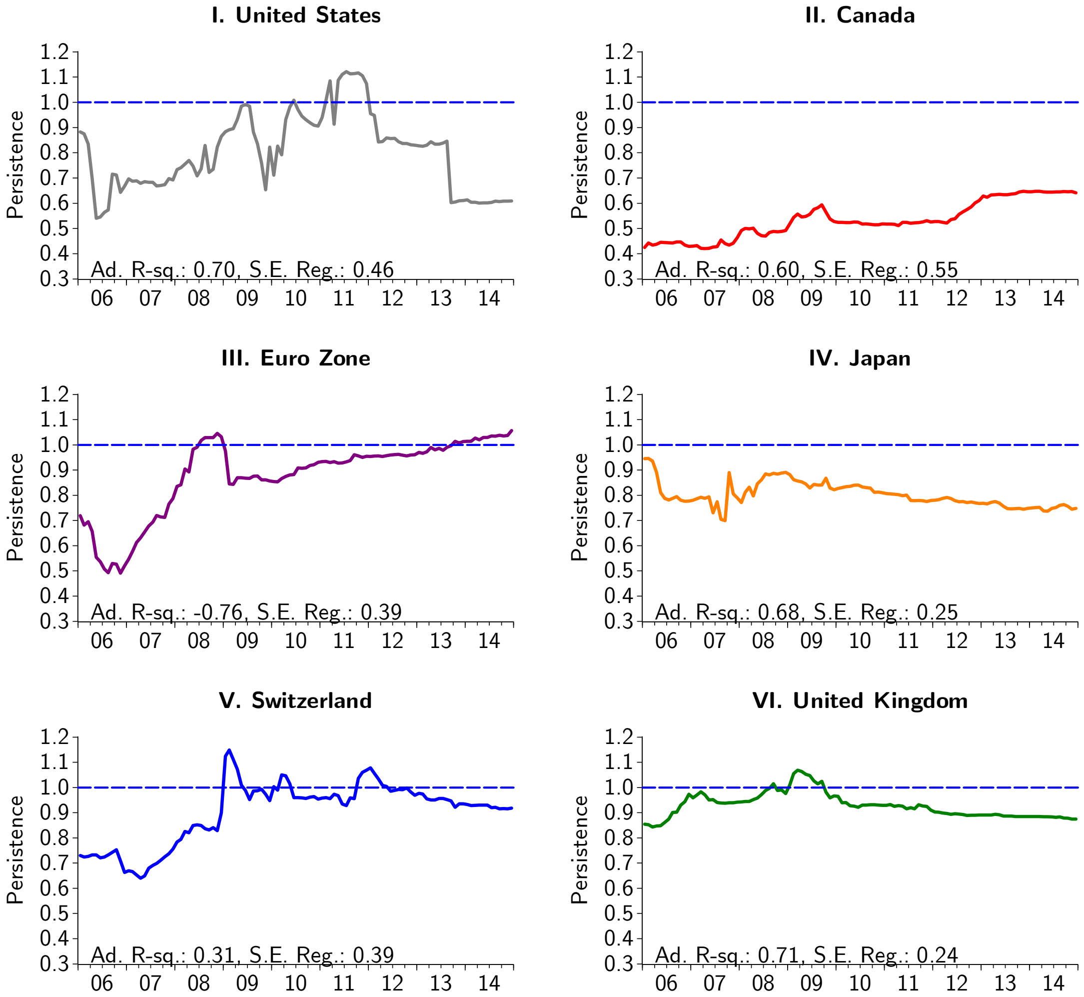
The J-statistic p-value suggests that the IV are valid at the 10% of confidence (8% for the US). The DW statistic reveals that the models are correctly specified for the US, Japan, Switzerland, and the UK, whereas for Canada and the Euro Zone the statistic is barely below unity. This means that while the model could be used assuming unbiasedness, there is room for controlling error autocorrelation using, for instance, MA terms.

In Figure 7, it is presented the dynamic *persistence profile*, similar as Figure 6 for the GVAR. In this case, it is observed a higher degree of constancy in the estimates. Nevertheless, there are two salient features that may affect forecast accuracy: except for Canada, the λ_b parameters are close to unity, and the US (source of the last big inflation shock) reduces its persistence to 0.60.

Same as with the GVAR–which actually is a NKPC–a lagged inflation coefficient above unity does not necessarily means non-stationarity (Fuhrer, 2011). But, from a predictive point of view, the model reduce its stabilisation capacity since shocks do not vanish and persists for a longer period of time. With respect to the US lagged inflation coefficient shrinkage, it could provide some limited information to the system. However, this informational flow must be considered in companion with the inflation information provided by the expectations. Non-reported results (available upon request) confirm this view, with an estimation that fulfil $\lambda_b + \lambda_f = 1$.

Overall, the HNKPC show estimates that are according to economic theory and gives inflation a high importance to lagged values typically improving in-sample fit.

Figure 7: HNKPC persistence profile. Evaluation sample (*)



(*) Source: Author's elaboration.

4.1.3 Autoregressive model diagnostics

The stationary AR model, despite its easy handling, provides an automatic stabilisation behaviour whenever $\left|\widehat{\phi}\right|<1$; *i.e.* already a stationary model. This gives the benefit of adaptiveness across the sample without dramatic changes in estimated parameters.

The estimation results using OLS with the Newey-West HAC correction for the standard deviation are presented in Table 7. The specifications obtained with a GETS strategy delivers p=1 for the Euro Zone, Japan, Switzerland, and the UK, p=2 for Canada, and p=3 for the US. Considering the six cases, it is obtained an average $\widehat{\phi}$=0.80, characterising well the inflationary persistence.

In these cases, and as is explicitly taken into consideration error's autocorrelation, the DW statistic suggests well specified models. The less explanatory power of the AR model is found for the Euro Zone ($\overline{R}^2$=0.342) and for Switzerland ($\overline{R}^2$=0.587). For the remaining countries, the $\overline{R}^2$ ranges from 0.70 to 0.82.

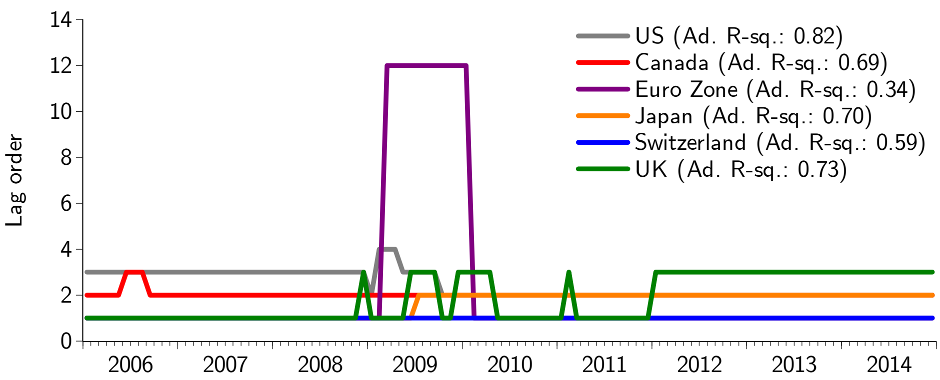
For a dynamic overview of the model's behaviour across the evaluation sample, a tracking plot of the chosen AR order for the six countries is presented in Figure 8. Some minor shift in the vicinity of p=3 for all countries are observed, and notably different for the Euro Zone, achieving the maximum number of permitted lags (p=12). Despite that this deviation is observed just for the crisis period (2008-9), it could jeopardise forecast accuracy according to the findings of Hansen (2009) and Medel (2015b). Hence, further research aiming to produce more demanding benchmarks could be a shrinkage or LASSO-based estimation of the AR model. Nevertheless, and considering all potential setbacks, these models provide valuable efficiency in terms of information usage and precision with minor complexity.

Table 7: AR models diagnostics (*)

	US	*CAN*	*EUR*	*JPN*	*SWI*	*UK*
	Dependent variable: π_t					
	Estimation sample					
ϕ_1	1.216	1.064	0.591	0.841	0.761	0.858
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
ϕ_2	-0.715	-0.302	-	-	-	-
	[0.000]	[0.000]	-	-	-	-
ϕ_3	0.426	-	-	-	-	-
	[0.000]	-	-	-	-	-
Constant	2.756	2.359	2.229	-0.841	0.928	1.429
	[0.000]	[0.000]	[0.000]	[0.006]	[0.000]	[0.000]
$\overline{R}^2$	0.824	0.692	0.342	0.704	0.587	0.731
S.E. Reg.	0.348	0.481	0.233	0.237	0.306	0.233
DW Stat.	1.874	1.852	1.855	1.843	2.130	2.008

(*) Equation: $\pi_t = c + \phi_1\pi_{t-1} + \phi_2\pi_{t-2} + \phi_3\pi_{t-3} + \varepsilon_t$, with $\varepsilon_t \sim iid\mathcal{N}(0, \sigma_\varepsilon^2)$. See notes in Table 3. Source: Author's elaboration.

Figure 8: AR chosen lag length profile. Evaluation sample (*)



(*) Source: Author's elaboration.

4.2 Out-of-sample results

As mentioned above, the out-of-sample results comprise the RMSFE Ratio to ease a comparison with the RW, for two samples: the evaluation and the shortened evaluation sample. The results using the whole evaluation sample are presented in Table 8, obviously considering the simple average of forecasting errors for a given horizon. Shaded cells are used to signal figures below unity–when the competing model outperforms the RW.

The results show that, overall, the HNKPC exhibit the best performance remarkably for the US and Canada, and with some exceptions at certain horizons for Japan (exception: h=24), Switzerland (h=24), and the Euro Zone (h=1). For the case of the UK, the HNKPC is never superior to the RW. The GW-test results indicate significant differences, except for Japan (all horizons), h=1 for the US and Canada, and h=6 for Canada.

The second best alternative is the AR, proving that it is a good statistical option as found in Posch and Rumler (2015). The evidence show that outperform the RW for the US (in all horizons), Canada (except for h=1), Japan (h=6), and Switzerland (h=1). It is never superior to the RW for the Euro Zone and for the UK. When the AR is superior to the RW, it is for a considerable predictive gain in all countries, except for Japan. The GW-test results indicate that the difference is statistically significant just for the US at h=1.

Particularly in the long-run, the ES show higher precision than the RW for all countries at h=24, for US, Switzerland, and the UK at h=12, and for Canada at h={6,12} also. Nevertheless, only for the Euro Zone at h=24 the superiority is statistically significant.

The GVAR outperforms the RW for Switzerland and the Euro Zone, both at h=6 and not statistically significant. Despite these results, it seems still a valid option when predicting at short-run for Japan, Switzerland, en the Euro Zone, showing RMSFE Ratio figures no greater than 17%. Note that these estimations contain several outliers in the case of the GVAR, which is analysed next.

When considering the shortened evaluation sample (2006.1-2008.8), the AR forecast show a good performance but was outperformed in the US at h=24, in Japan was never superior than the RW, and for the Euro Zone and Switzerland was superior just at h=12.

The ES show no major differences across the both samples, which is due to its close behaviour to the RW. The HNKPC undoubtedly has experienced an improvement in its accuracy after the financial crisis, which constitutes useful evidence in regard of the use of the HNKPC models.

The GVAR has also showed a better performance before the financial crisis, especially at the long-run. In particular, for the US at h=24 show a predictive gain of 26%, plus a 31% for Switzerland, 44% for the UK, and a small 3.1% for the Euro Zone at the same horizon. Also for the Euro Zone, the GVAR at h=12 showed a 22% of predictive gain before the crisis.

Overall, it is possible to conclude that the GVAR loses its predictive ability especially observed for the long-run previous to the crisis. The HNKPC makes a more accurate inflationary tracking during and after the crisis. Finally, the AR model has proved to have a robust behaviour when predicting inflation under different statistical scenarios (including breaks).

To have an appraisal of the accuracy across the estimation sample, in Figures 9-14 it is presented the forecast error for a given horizon for all the countries. Note that there are some outliers observations not depicted for the GVAR forecast errors, which are omitted to ease a visual inspection. The RMSFE, and consequently the RMSFE Ratio, obviously includes these observations; indeed motivating an in-depth analysis. The criterion to define an observation as an outlier is just accommodative, not following–but likely delivering the same results–an particular statistical procedure.

In Figure 9, it is observed for the US that the financial crisis, commonly dated start point in September 2008, directly affects the accuracy of all the models, but particularly for the GVAR at h={6,12,24}. An outlier for the GVAR at h=24 is noticeable. Note that since h=12 that the GVAR seems unanchored to the targeted variable. The HNKPC errors are significant at h={12,24} during the crisis, outperforming remaining models.

In Figure 10, for Canada, it is observed some similar results than the US, but with a much greater variance at h=12. In this case, when the actual inflation already exhibits a higher variance than the US, the GVAR produces some out bounded errors since h=6. Note that at h=24 the behaviour of the

GVAR follows closely those of remaining models except for some particular observations. These error peaks, however, could provide valuable information for other purposes, such as turning point detection, spillover effects, and events probability estimations.

Figure 11 shows the case of the Euro Zone, also with an outlier of the GVAR at h=24, but at h={6,12,24} there are clear peaks of the AR model. Same as previous cases, the HNKPC exhibits the best inflation tracking during the crisis. Note also that the RW is often closer to the zero line (=zero forecast error) than some competing models. Consequently, the ES also seems reasonable for h={6,12,24}.

For the Japanese case in Figure 12, it is noticeable well-behaved errors, with almost all models describing the same shape. Different error dynamics of the GVAR are observed at h=6, predicting better the inflation during the crisis. Nevertheless, at h={12,24} the results are not in favour of the GVAR.

The case of Switzerland is presented in Figure 13, with one GVAR outlier at h=24, and follow closely the Japanese case. This is a better behaviour for h=6, but spoiled out during the crisis for h={12,24}. Particularly at h=12, when excluding a few atypical errors observations of the GVAR, the model becomes indistinguishably different from the candidates.

Finally, in Figure 14, it is presented the case of the UK. It is easy to notice that both the HNKPC and GVAR show some atypical observations. There is one particular outlier of the GVAR omitted at h=24. Note that, according to Tables 8-9, this is one of the most difficult cases to beat the RW. It is also observable that the GVAR tends to underestimate the inflation rate after the crisis.

Overall, it is confirmed the assumption that the GVAR is negatively affected by its performance during the financial crisis. It is also observed that the HNKPC and the AR are the models that provide a better cast of the inflation dynamics during the crisis. In Table 10, it is presented a summary of just the best forecasting models using the whole evaluation sample. These results are also divided in "Atheoretical" and "Economics" models, clearly showing a better performance of the "Economics" models in the short-run and "Atheoretical" in the long-run.

Table 8: RMSFE Ratio estimates (*)

	AR	ES	GVAR	HNKPC	AR	ES	GVAR	HNKPC
	Evaluation sample							
	US				*JPN*			
h=1	0.817*	1.949*	1.212*	0.947	0.993	1.832*	1.062	0.887*
h=6	0.902	1.034	1.264	0.503*	1.105	1.080	0.978	0.649*
h=12	0.556	0.922	1.589	0.441*	0.955	1.010*	1.291	0.845*
h=24	0.522	0.935	9.469	0.724*	0.869	0.968	2.762	1.143*
	CAN				*SWI*			
h=1	1.001	1.366*	1.512*	0.942	1.049	1.671*	1.133	0.990
h=6	0.785	0.992	1.756*	0.594*	0.836	1.104*	0.981	0.990
h=12	0.510	0.944	3.061*	0.524*	0.564	0.952*	1.472	0.965
h=24	0.426	0.893	2.437*	0.678*	0.769	0.898*	9.465	1.276
	EUR				*UK*			
h=1	1.181*	1.834*	1.171	1.035	1.087*	1.757*	1.353*	1.064
h=6	1.310	1.167*	0.872	0.684	1.145*	1.057	1.648*	1.453*
h=12	1.232	1.022*	1.263	0.388*	1.311	0.932	2.647*	2.344
h=24	1.476	0.955*	4.521	0.301*	1.691	0.979	7.123*	4.237

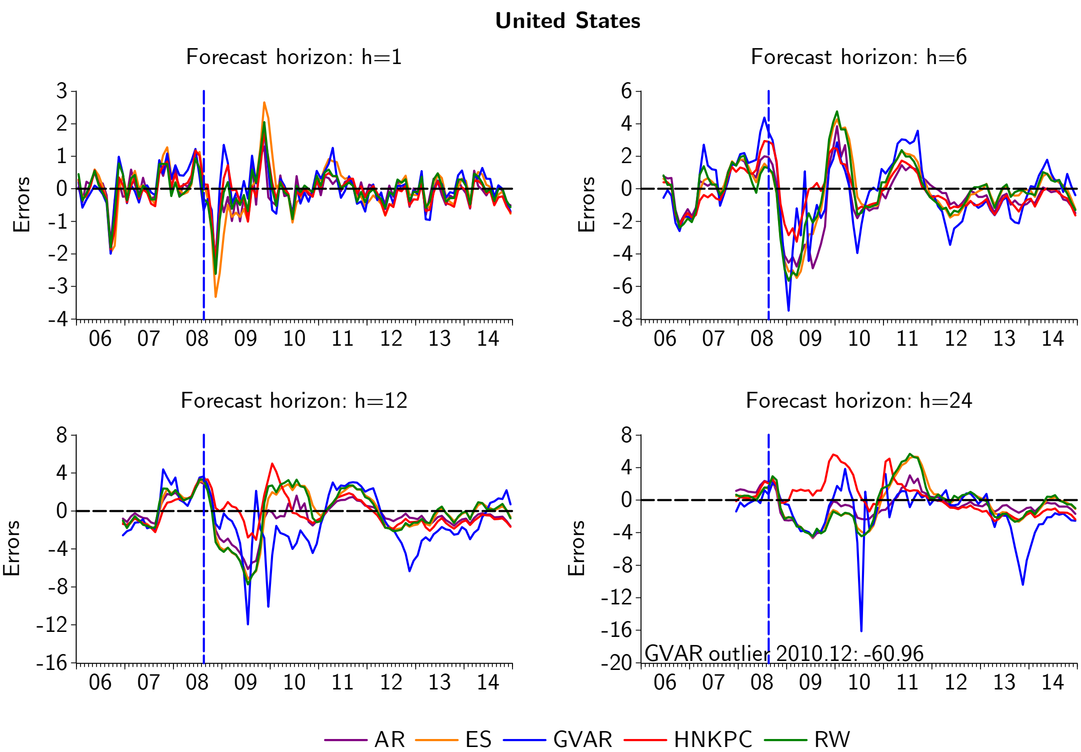
(*) Shaded cells = figures below unity. GW test results: (***) p<1%, (**) p<5%, (*) p<10%. Source: Author's elaboration.

Table 9: RMSFE Ratio estimates (*)

	AR	ES	GVAR	HNKPC	AR	ES	GVAR	HNKPC
	Shortened evaluation sample							
	US				*JPN*			
h=1	0.819	1.506*	1.339*	1.168	1.159*	1.538*	1.259*	1.091
h=6	0.938	0.936	2.293*	1.175	1.369*	1.140*	1.893	1.036
h=12	0.682*	0.924	1.631*	0.826	1.161	0.930*	1.311*	0.760*
h=24	3.152*	0.960	0.744	1.826	1.935*	1.232*	1.082	0.226*
	CAN				*SWI*			
h=1	0.956	1.141	1.529*	0.892	1.225*	1.514*	1.093	1.311*
h=6	0.628*	0.790*	2.991*	0.668	1.076	1.179*	1.856*	1.458*
h=12	0.429*	0.729*	1.987	0.794	0.590*	0.970	1.159	1.158
h=24	0.855	0.803*	1.926*	1.496	1.559*	1.040	0.696*	4.695*
	EUR				*UK*			
h=1	1.501*	1.597*	2.018*	1.562*	1.123*	2.085*	1.263	1.181
h=6	1.255	1.147*	1.675*	1.345	1.230*	1.047	1.138	1.460*
h=12	0.630*	0.974*	0.785*	0.641*	1.221*	0.813*	1.203	1.289*
h=24	1.267*	0.986	0.969	0.989	2.131*	1.098	0.565*	2.004*

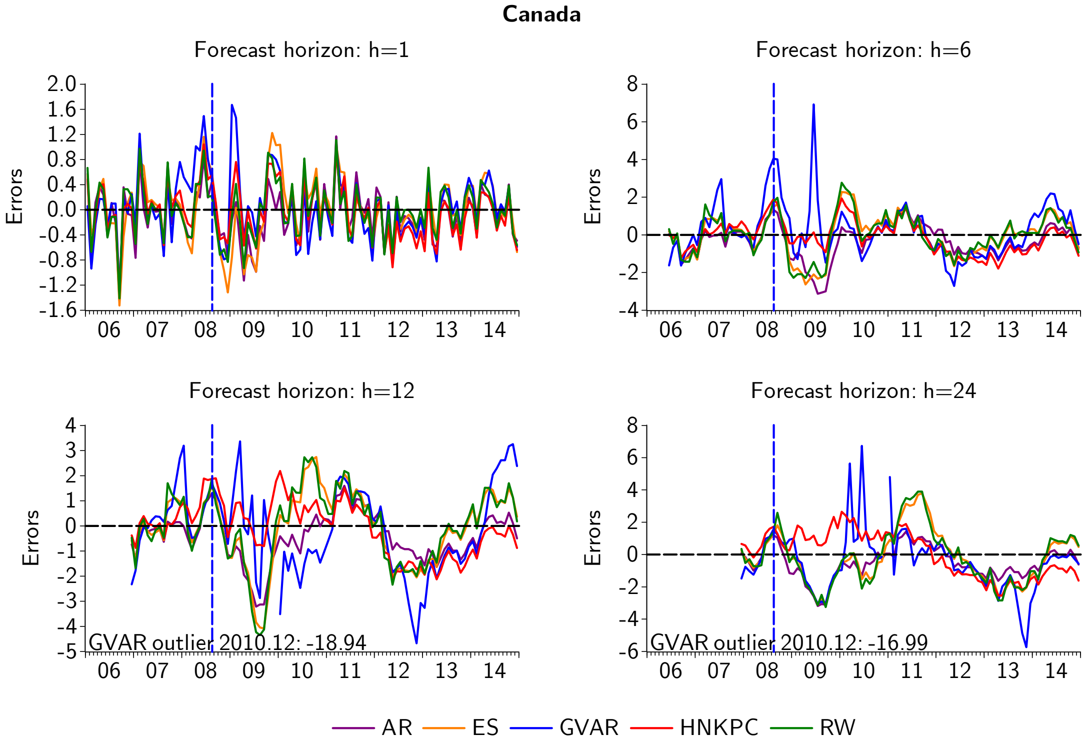
(*) Shaded cells = figures below unity. GW test results: (***) p<1%, (**) p<5%, (*) p<10%. Source: Author's elaboration.

Figure 9: United States. Multihorizon forecasting errors across time (*)



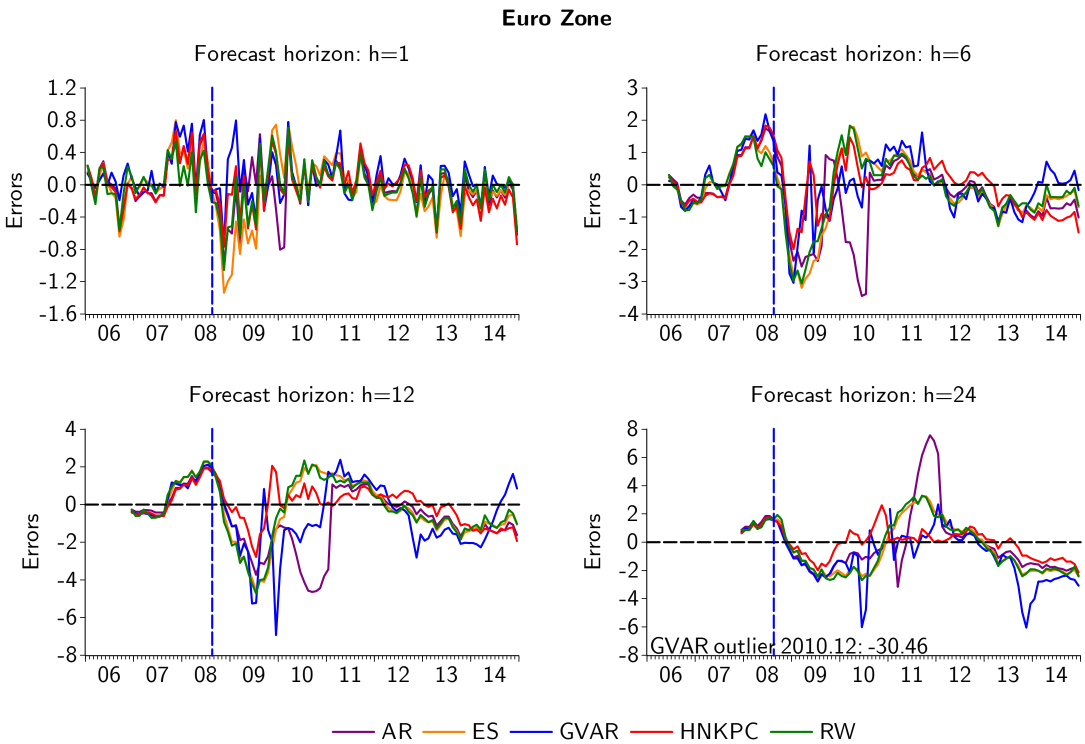
(*) Vertical line = end of shortened evaluation sample. Source: Author's elaboration.

Figure 10: Canada. Multihorizon forecasting errors across time (*)



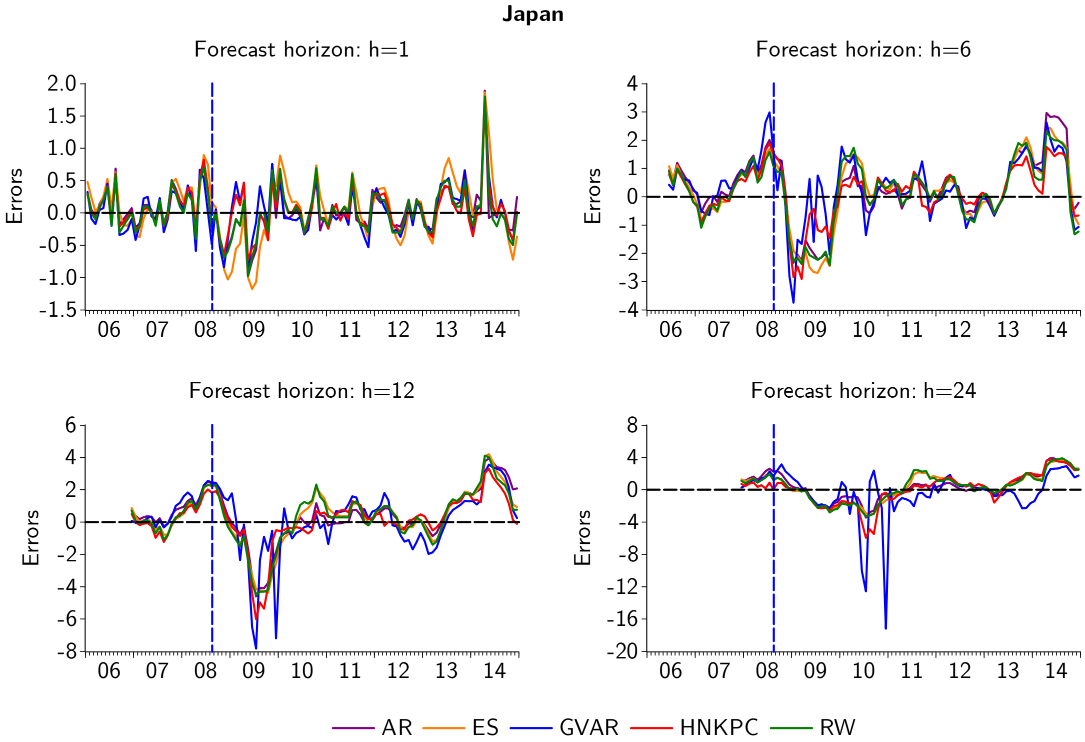
(*) Vertical line = end of shortened evaluation sample. Source: Author's elaboration.

Figure 11: Euro Zone. Multihorizon forecasting errors across time (*)



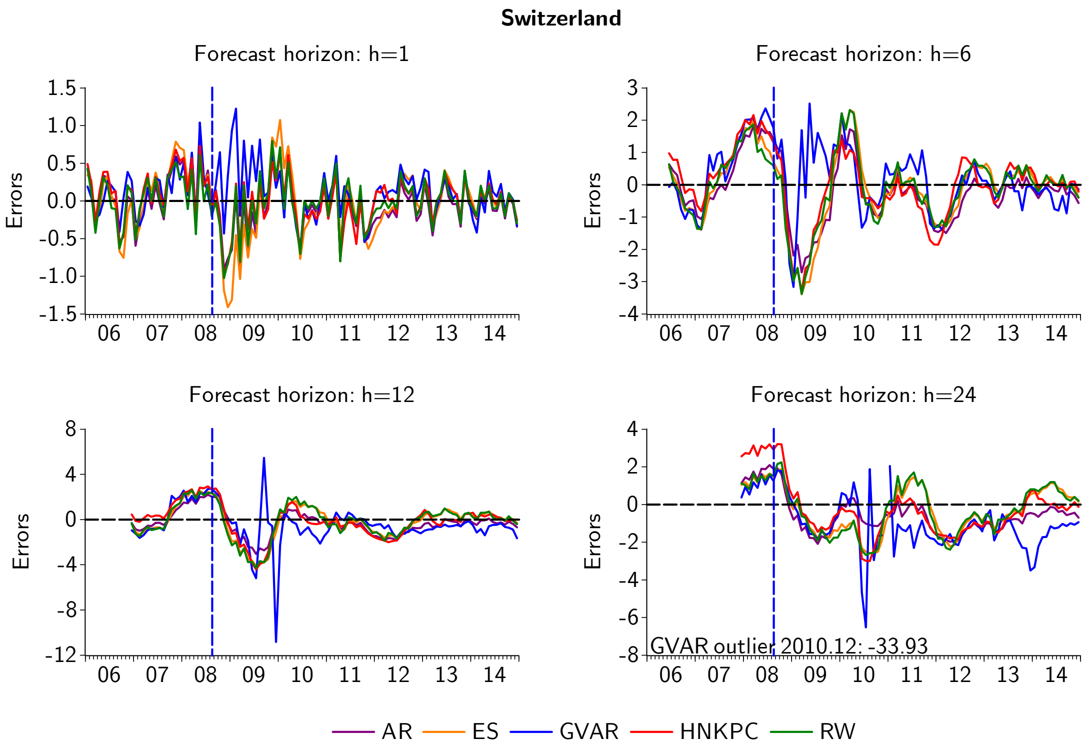
(*) Vertical line = end of shortened evaluation sample. Source: Author's elaboration.

Figure 12: Japan. Multihorizon forecasting errors across time (*)



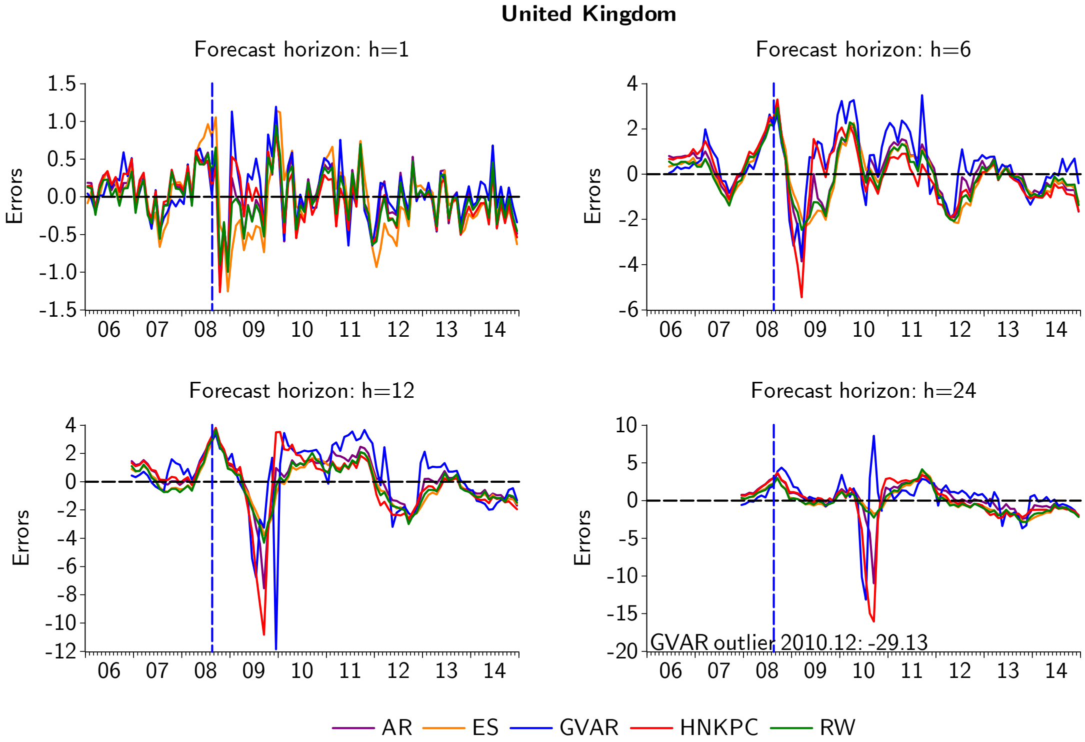
(*) Vertical line = end of shortened evaluation sample. Source: Author's elaboration.

Figure 13: Switzerland. Multihorizon forecasting errors across time (*)



(*) Vertical line = end of shortened evaluation sample. Source: Author's elaboration.

Figure 14: United Kingdom. Multihorizon forecasting errors across time (*)



(*) Vertical line = end of shortened evaluation sample. Source: Author's elaboration.

Table 10: Summary of out-of-sample results (*)

	Evaluation sample					
	AR	ES	GVAR	HNKPC	**Atheoretical**	**Economics**
h=1	1	-	-	5	1	5
h=6	1	1	-	4	2	4
h=12	2	1	-	3	3	3
h=24	4	1	-	1	5	1

(*) Figures indicate times that a model is the best given an horizon.
Source: Author's elaboration.

5 Summary and concluding remarks

In this article, it is analysed the multihorizon predictive power of the HNKPC making use of a compact-scale GVAR for the headline inflation. The GVAR includes five developed countries and one region (the US, Canada, Japan, Switzerland, the UK plus the Euro Zone) exhibiting different inflationary experiences. The used monthly sample covers from 2000.1 until 2014.12 (180 observations), divided in the estimation sample (2000.1-2005.12, 72 observations) and two evaluation samples. Therefore, the whole evaluation spans (2006.1-2014.12, 108 observations) plus a shortened span (2006.1-2008.8; 32 observations). Special attention is given to this shortened period given some atypical projections obtained with the GVAR. The analysed forecast horizons are h={1,6,12,24} months ahead. The marginal cost proxy variable is the output gap with a special treatment for the *end-of-sample* problem.

The key element of this article is the use of direct measures of inflation expectations–CF–embedded in a GVAR environment for inflation forecasting purposes; using a HNKPC specification. As CF is a fixed-horizon prediction–for December of the current and the next year–a special adjustment is made.

The GVAR forecasts are statistically compared to several benchmarks using the RMSFE statistic and the GW testing procedure. These benchmarks are the AR, ES, and the RW acting as a pivot benchmark. One last economics-based benchmark is the *closed economy* univariate HNKPC.

The results indicate that the GVAR is a valid forecasting procedure especially in the short-run. This is the case for the Euro Zone, Japan, and Switzerland for h=6. For most cases (countries and horizons), the most accurate forecasts are obtained with the AR and especially with the HNKPC. In the long-run, the ES model also appears as a better alternative rather than the RW.

When the forecast errors across the time are depicted, it is noticed that especially at short-run horizons, the MSPE is mainly driven by the unanticipated effects of the financial crisis started in the US in 2008.9. To take this limitation into account, in the shortened evaluation sample the GVAR appears as a valid alternative to the RW also in the long-run for the US, the Euro Zone, Switzerland, and the UK for h=24, and the Euro Zone again at h=12. The most robust forecasting device across countries and horizons is the HNKPC, suggesting that there is a role for economic fundamentals when forecasting inflation.

Note that the results provide heterogeneous results across the countries. This suggests that an averaging scheme may be fruitful for accuracy purposes. Also, and given the AR results, its estimation with different methods could also improve the forecasts. Finally, the inclusion of more countries and global variables to the GVAR may capture better domestic inflation dynamics. All these features may be of interest and are left for further research.

Acknowledgements

I thank the suggestions, comments, and help to Professor Kevin C. Lee and Pablo Medel. Nevertheless, I exclude them for any error or omission that remains at my own responsibility.

Disclosure

No other interest rather than an economic research question on applied economics has motivated this article. There is no any conflict of interest of any kind involved in the production of this article.

References

1. Abbas, S.K. and P.M. Sgro, 2011, "New Keynesian Phillips Curve and Inflation Dynamics in Australia," *Economic Modelling* **28**(4): 2022-2033.

2. Aiolfi, M., C. Capistrán, and A. Timmermann A., 2011, *Forecast Combinations*, in M.P. Clements and D.F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*, Oxford University Press, US.

3. Agénor, P.R. and N. Bayraktar, 2010, "Contracting Model of the Phillips Curve Empirical Estimates for Middle-Income Countries," *Journal of Macroeconomics* **32**(2): 555-570.

4. Ager, P., M. Kappler, and S. Osterloh, 2009, "The Accuracy and Efficiency of the Consensus Forecasts: A Further Application and Extension of the Pooled Approach," *International Journal of Forecasting* **25**(1): 167-181.

5. Akaike, H., 1974, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control* **19**(6): 716-723.

6. Andersson, M., G. Karlsson, and J. Svensson, 2007, "The Riksbank Forecasting Performance," *Economic Review* **3**: 59-75.

7. Andrews, D.W.K., 1993, "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica* **61**(1): 139-165.

8. Andrews, D.W.K. and H.-Y. Chen, 1994, "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics* **12**(2): 187-204.

9. Ang, A., G. Bekaert, and M. Wei, 2007, "Do Macro Variables, Assets Markets or Surveys Forecast Inflation Better?" *Journal of Monetary Economics* **54**(4): 1163-1212.

10. Assenmacher, K., 2013, *Forecasting the Swiss Economy with a Small GVAR Model*, in F. di Mauro and M.H. Pesaran (Eds.), *The GVAR Handbook*, Oxford University Press, UK.

11. Assenmacher-Wesche, K. and M.H. Pesaran, 2008, "Forecasting the Swiss Economy Using VECX* Models: An Exercise in Forecast Combination Across Models and Observation Windows," *National Institute Economic Review* **203**(1): 91-108.

12. Assenmacher-Wesche, K. and D. Geissmann, 2012, "*Forecasting Swiss Inflation and GDP with a Small Global VAR Model*," *manuscript*, Swiss National Bank.

13. Atkeson, A. and L.E. Ohanian, 2001, "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review* **25**(1): 2-11.

14. Bai, J. and S. Ng, 2008, "Forecasting Economic Time Series using Targeted Predictors," *Journal of Econometrics* **146**: 304-317.

15. Balakrishnan, R. and J.D. López-Salido, 2002, "*Understanding UK Inflation: The Role of Openness*," Working Paper 164, Bank of England.

16. Barrell, R., K. Dury, I. Hurst, and N. Painl, 2001, "*Modelling the World Economy: The National Institute's Global Econometric Model, NiGEM*," paper presented at the Workshop on Simulation Properties of Macroeconometric Models, European Network of Economic Policy Research Institutes (ENEPRI), Paris, July 2001.

17. Bates, J.M. and C.W.J. Granger, 1969, "The Combination of Forecasts," *Operational Research Quarterly* **20**(4): 451-468.

18. Batini, N., B. Jackson, and S. Nickell, 2005, "An Open-Economy New Keynesian Phillips Curve for the UK," *Journal of Monetary Economics* **52**(6): 1061-1071.

19. Bloom, N., 2009, "The Impact of Uncertainty Shocks," *Econometrica* **77**(3): 623-685.

20. Bobbitt, L. and M.C. Otto, 1990, "*Effects of Forecasts on the Revisions of Seasonally Adjusted Values Using the X-11 Seasonal Adjustment Procedure*," Proceedings of the Business and Economic Statistics Section, American Statistical Association, 449-453.

21. Box, G.E.P. and G.M. Jenkins, 1970, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, US.

22. Brissimis, S.N. and N.S. Magginas, 2008, "Inflation Forecasts and the New Keynesian Phillips Curve," *International Journal of Central Banking* **08**(June): 1-22.

23. Calhoun, G. and G. Elliott, 2012, "*Why Do Nonlinear Models Provide Poor Macroeconomic Forecasts?*" paper presented at the Seventh ECB Workshop on Forecasting Techniques - New Directions for Forecasting, Frankfurt am Main, Germany, 4-5 May 2012.

24. Calhoun, G., 2014, "*Out-of-Sample Comparisons of Overfit Model*," Working Paper 11002, Department of Economics, Iowa State University, US.

25. Calvo, G.A., 1983, "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics* **12**(3): 383-398.

26. Canova, F., 2007, "G-7 Inflation Forecasts: Random Walk, Phillips Curve or What Else?" *Macroeconomic Dynamics* **11**: 1-30.

27. Carrière-Swallow, Y. and C.A. Medel, 2011, "*Global Uncertainty over the South Pacific*," [in Spanish] Working Paper 16-2001, Banco Central de Reserva del Perú.

28. Carrière-Swallow, Y. and L.F. Céspedes, 2013, "The Impact of Uncertainty Shocks in Emerging Economies," *Journal of International Economics* **90**(2): 316-325.

29. Charemza, W., C. Díaz, and S. Makarova, 2015, "*Conditional Term Structure of Inflation Forecast Uncertainty: The Copula Approach*," Working Paper 15/07, Department of Economics, University of Leicester, UK.

30. Christiano, L.J., M. Eichenbaum, and C.L. Evans, 2005, "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy* **113**(1): 1-45.

31. Chudik, A. and M.H. Pesaran, 2014, "*Theory and Practice of GVAR Modeling*," Cambridge Working Papers in Economics 1408, Faculty of Economics, University of Cambridge, UK.

32. Clements, M.P. and D.F. Hendry, 2011, *The Oxford Handbook of Economic Forecasting*, Oxford University Press, US.

33. Cobb, M. and C.A. Medel, 2010, "An Estimation of the Impact of the Trading Day Effect on Seasonally Adjusted Chilean Series of Activity and Demand," [in Spanish] *Economía Chilena* **13**(3): 95-103.

34. Cochrane, J., 2001, *Asset Pricing*, Princeton University Press, US.

35. Collard, F. and H. Dellas, 2004, "*The New Keynesian Model with Imperfect Information and Learning*," Working Paper 273, Institut d'Économie Industrielle (IDEI), Toulouse, France.

36. Corsetti, G., L. Dedola, and S. Leduc, 2010, *Optimal Monetary Policy in Open Economies*, in B.M. Friedman and M. Woodford (Eds.), *Handbook of Monetary Economics, Volume 3*, Elsevier, North-Holland.

37. Corberán-Vallet, A., J.D. Bermúdez, E. Vercher, 2011, "Forecasting Correlated Time Series with Exponential Smoothing Models," *International Journal of Forecasting* **27**(2): 252-265.

38. Croushore, D. and T. Stark, 2001, "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* **105**(1): 111-130.

39. Dees, S., F. di Mauro, M.H. Pesaran, and L.V. Smith, 2007 [DdMPS], " Exploring the International Linkages of the Euro Area: A Global VAR Analysis," *Journal of Applied Econometrics* **22**(1): 1-38.

40. Dees, S., S. Holly, M.H. Pesaran, and L.V. Smith, 2007, "Long Run Macroeconomic Relations in the Global Economy," *Economics: The Open-Access, Open-Assessment E-Journal* **3**(2007): 1-58.

41. Dees, S., M.H. Pesaran, L.V. Smith, and R.P. Smith, 2009, "Identification of New Keynesian Phillips Curves from a Global Perspective," *Journal of Money, Credit and Banking* **41**(7): 1481-1502.

42. Dennis, R. and J.A. López, 2004, "Policy Applications of a Global Macroeconomic Model," *FRBSF Economic Letter* 2004-12.

43. De Waal, A., R. Van Eyden, and R. Gupta, 2015, "Do We Need a Global VAR Model to Forecast Inflation and Output in South Africa?," *Applied Economics* **47**(25): 2649-2670.

44. di Mauro, F. and M.H. Pesaran (Eds.), 2013, *The GVAR Handbook. Structure and Applications of a Macro Model of the Global Economy for Policy Analysis*, Oxford University Press, UK.

45. Dickey, D.A. and S.G. Pantula, "Determining the Order of Differencing in Autoregressive Processes," *Journal of Business and Economic Statistics* **5**(4): 455-459.

46. Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004, "Least Angle Regression," *Annals of Statistics* **32**(2): 407-499.

47. Elliott, G., C.W.J. Granger, and A. Timmermann (Eds.), 2006, *Handbook of Economic Forecasting, Volume 1*, Elsevier, North-Holland.

48. Elliott, G. and A. Timmermann, 2008, "Economic Forecasting," *Journal of Economic Literature* **46**(1): 3-56.

49. Erceg, C.J. and A.T. Levin, 2003, "Imperfect Credibility and Inflation Persistence," *Journal of Monetary Economics* **50**(4): 915-944.

50. Ericsson, N.R. and E.L. Reisman, 2012, "Evaluating a Global Vector Autoregression for Forecasting," *International Advances in Economic Research* **18**:247-258.

51. Faust, J. and J. Wright, 2013, *Forecasting Inflation*, in G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 2*, Elsevier, North-Holland.

52. Findley, D.F, B.C. Monsell, W.R. Bell, M.C. Otto and B.-C. Chen, 1998, "New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program," *Journal of Business and Economic Statistics* **16**(2): 127-152.

53. Fuhrer, J.F., 2011, *Inflation Persistence*, in B.M. Friedman and M. Woodford (Eds.), *Handbook of Monetary Economics, Volume 3*. Elsevier, North-Holland.

54. Galí, J. and M. Gertler, 1999, "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics* **44**(2): 195-222.

55. Galí, J., M. Gertler, and J.D. López-Salido, 2001, "European Inflation Dynamics," *European Economic Review* **45**(7): 1237-1270.

56. Galí, J., M. Gertler, and J.D. López-Salido, 2005, "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve," *Journal of Monetary Economics* **52**(6): 1107-1118.

57. Galí, J. and T. Monacelli, 2005, "Monetary Policy and Exchange Rate Volatility in a Small Open Economy," *Review of Economic Studies* **72**: 707-734.

58. Garratt, A., K.C. Lee, M.H. Pesaran, and Y. Shin, 2006, *Global and National Macroeconometric Modelling: A Long-Run Structural Approach*, Oxford University Press, Oxford, UK.

59. Garratt, A., K.C. Lee, E. Mise, and K. Shields, 2008, "Real-Time Representations of the Output Gap," *Review of Economics and Statistics* **90**(4): 792-804.

60. Giacomini, R. and H. White, 2006, "Tests of Conditional Predictive Ability," *Econometrica* **74**(6): 1545-1578.

61. Ghysels E., D. Osborn, and P.M. Rodrígues, 2006, *Forecasting Seasonal Time Series*, in G. Elliott, C.W.J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, *Volume 1*, Elsevier, North Holland.

62. Granger, C.W.J., 1982, "Acronyms in Time Series Analysis (ATSA)," *Journal of Time Series Analysis* **3**(2): 103-107.

63. Granger, C.W.J. and Y. Jeon, 2007, "Evaluation of Global Models," *Economic Modelling* **24**(6): 980-989.

64. Granger, C.W.J. and Y. Jeon, 2011, "The Evolution of the Phillips Curve: A Modern Time Series Viewpoint," *Economica* **78**: 51-66.

65. Groen J.J.J., G. Kapetanios, and S. Price, 2009, "A Real Time Evaluation of Bank of England Forecasts of Inflation and Growth," *International Journal of Forecasting* **25**: 74-80.

66. Groen, J.J.J., R. Paap, and F. Ravazzolo, 2013, "Real-Time Inflation Forecasting in a Changing World," *Journal of Business and Economic Statistics* **31**(1): 29-44.

67. Gross, M., 2013, "*Estimating GVAR Weight Matrices*," Working Paper 1523, European Central Bank.

68. Gruen, D., T. Robinson, A. Stone, 2002, "*Output Gaps in Real Time: Are They Reliable Enough to Use for Monetary Policy?*," Research Discussion Paper 2002-26, Reserve Bank of Australia.

69. Hansen, L.P., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* **50**(4): 1029-1054.

70. Hansen, B.E., 1999, "The Grid Bootstrap and the Autoregressive Model," *Review of Economics and Statistics* **81**(4): 594-607.

71. Hansen, P.R., 2009, "*In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and its Implications for Model Selection*," *manuscript*, version of 23 April, 2009, Department of Economics, Stanford University, US.

72. He, Q., H. Shen, and Z. Tong, 2012, "Investigation of Inflation Forecasting," *Applied Mathematics and Information Sciences* **6**(3): 649-655.

73. Hendry, D.F., S. Johansen, and C. Santos, 2008, "Automatic Selection of Indicators in a Fully Saturated Regression," *Computational Statistics* **23**(2): 317-335.

74. Henzel, S. and T. Wollmershaeuser, 2008, "The New Keynesian Phillips Curve and the Role of Expectations: Evidence from the CESifo World Economic Survey," *Economic Modelling* **25**(5): 811-832.

75. Holan, S.H., R. Lund, and G. Davis, 2010, "The ARMA Alphabet Soup: A Tour of ARMA Model Variants," *Statistics Survey* **4**: 232-274.

76. Huff, D., 1993, *How to Lie with Statistics*, Reissue Edition, W.W. Norton and Company, UK.

77. Hyndman, R.J., A.B. Koehler, J.K. Ord, and R.D. Snyder, 2008, *Forecasting with Exponential Smoothing. The State Space Approach*. Springer Series on Statistics, Berlin, Germany.

78. Jean-Baptiste, F., 2012, "Forecasting with the New Keynesian Phillips Curve: Evidence from Survey Data," *Economics Letters* **117**(3): 811-813.

79. Kaiser, R. and A. Maravall, 1999, "Estimation of the Business Cycle: A Modified Hodrick-Prescott Filter," *Spanish Economic Review* **1**: 175-206.

80. Kichian, M. and F. Rumler, 2014, "Forecasting Canadian Inflation: A Semi-Structural NKPC Approach," *Economic Modelling* **43**: 183-426.

81. Kim, J.H., 2003, "Forecasting Autoregressive Time Series with Bias-corrected Parameter Estimators," *International Journal of Forecasting* **19**(4): 493-502.

82. Koehler, A.B. and E.S. Murphree, 1988, "A Comparison of the Akaike and Schwarz Criteria for Selecting Model Order," *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **37**(2): 187-195.

83. Kolassa, S., 2011, "Combining Exponential Smoothing Forecasts using Akaike Weights," *International Journal of Forecasting* **27**(2): 238-251.

84. Kostenko, A.V. and R.J. Hyndman, 2008, "*Forecasting Without Significance Tests?*" *manuscript*, Monash University, Australia.

85. Kuester, K., G.J. Müller, and S. Stöling, 2009, "Is the New Keynesian Phillips Curve Flat?" *Economics Letters* **103**(1): 39-41.

86. Kuha, J., 2004, "AIC and BIC: Comparison of Assumptions and Performance," *Sociological Methods and Research* **33**(2): 188-229.

87. Kullback, S. and R.A. Leibler, 1951, "On Information and Sufficiency," *Annals of Mathematical Statistics* **22**: 79-86.

88. Lawless, M. and K. Whelan, 2011, "Understanding the Dynamics of Labour Shares and Inflation," *Journal of Macroeconomics* **33**(2): 121-136.

89. Leith, C. and J. Malley, 2007, "Estimated Open Economy New Keynesian Phillips Curves for the G7," *Open Economies Review* **18**(4): 405-426.

90. Levin, A., A. Onatski, A. Williams, and J. Williams, 2005, *Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models*, in M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual*, MIT Press, US.

91. Lindé, J., 2005, "Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach," *Journal of Monetary Economics* **52**(6): 1135-1149.

92. Lovell, M., 2008, "A Simple Proof of the FWL Theorem," *Journal of Economic Education* **39**(1): 88-91.

93. Loungani, P., 2001. "How Accurate are Private Sector Forecasts? Cross-country Evidence from Consensus Forecasts of Output Growth," *International Journal of Forecasting* **17**(3): 419-432.

94. Lütkepohl, H., 1985, "Comparison of Criteria for Estimating the Order of a Vector Autoregressive Process," *Journal of Time Series Analysis* **6**(1): 35-52.

95. McAdam, P. and A. Willman, 2003, "New Keynesian Phillips Curves: A Reassessment using Euro Area Data," Working Paper 265, European Central Bank.

96. Marcellino, M., J.H. Stock, and M.W. Watson, 2006, "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics* **135**: 499-526.

97. Mazumber, S., 2010, "The New Keynesian Phillips Curve and the Cyclicality of Marginal Cost," *Journal of Macroeconomics* **32**(3): 747-765.

98. Mazumber, S., 2011, "The Long-Run Relationship Between Inflation and the Markup in the US," *Economics Bulletin* **31**(1): 473-484.

99. Medel, C.A. and M. Pedersen, 2010, "Uncertainty in Seasonally Adjusted Series of Activity and Demand in Chile," [in Spanish] *Economía Chilena* **13**(1): 63-72.

100. Medel, C.A. and S.C. Salgado, 2013, "Does the BIC Estimate and Forecast Better than the AIC?," *Economic Analysis Review* **28**(1): 47-64.

101. Medel, C.A., 2015a, "*Inflation Dynamics and the Hybrid Neo Keynesian Phillips Curve: The Case of Chile*," MPRA Paper 62609, University Library of Munich, Germany.

102. Medel, C.A., 2015b, "Classical Probability of Overfitting with Information Criteria: Estimates with Chilean Macroeconomic Series," [in Spanish] *Economic Analysis Review* **29**(1): 57-72.

103. Medel, C.A. and P. Pincheira, 2015, "The Out-of-sample Performance of an Exact Median-Unbiased Estimator for the Near-Unity AR(1) Model," forthcoming, *Applied Economics Letters*.

104. Mihailov, A., F. Rumler, and J. Scharler, 2011, "The Small-Open Economy New Keynesian Phillips Curve: Empirical Evidence and Implied Inflation Dynamics," *Open Economies Review* **22**(2): 317-337.

105. Mise, E., T.-H. Kim, and P. Newbold, 2005, "On Suboptimality of the Hodrick-Prescott Filter at Time Series Endpoints," *Journal of Macroeconomics* **27**(1): 53-67.

106. Muth, J., 1961, "Rational Expectations and the Theory of Price Movements," *Econometrica* **29**(3): 315-335.

107. Nason, J.M. and G.W. Smith, 2008, "Identifying the New Keynesian Phillips Curve," *Journal of Applied Econometrics* **23**(5): 525-251.

108. Neely, C.J. and D.E. Rapach, 2011, "International Comovements in Inflation Rates and Country Characteristics," *Journal of International Money and Finance* **30**(7): 1471-1490.

109. Newey, W.K. and K.D. West, 1987, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* **55**(3): 703-708.

110. Nunes, R., 2010, "Inflation Dynamics: The Role of Expectations," *Journal of Money, Credit and Banking* **42**(6): 1161-1172.

111. Orphanides, A., 2001, "Monetary Policy Rules Based on Real-Time Data," *American Economic Review* **91**(4): 964-985.

112. Orphanides, A. and S. van Norden, 2002, "The Unreliability of Output-Gap Estimates in Real Time," *The Review of Economics and Statistics* **LXXXIV**(4): 569-583.

113. Orphanides, A. and S. van Norden, 2005, "The Reliability of Inflation Forecasts based on Output Gap Estimates in Real Time," *Journal of Money, Credit and Banking* **37**(3): 583-601.

114. Paloviita, M. and D. Mayes, 2005, "The Use of Real-Time Information in Phillips-Curve Relationships for the Euro Area," *The North American Journal of Economics and Finance* **16**(3): 415-434.

115. Paloviita, M., 2009, "Estimating Open Economy Phillips Curves for the Euro Area with Directly measured Expectations," *New Zealand Economic Papers* **43**(3): 233-254.

116. Pesaran, M.H., T. Schuermann, and S.M. Weiner, 2004, "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model," *Journal of Business and Economic Statistics* **22**(2): 129-162.

117. Pesaran, M.H., T. Schuermann, and L.V. Smith (2009), "Forecasting Economic and Financial Variables with Global VARs," *International Journal of Forecasting* **25**(4): 642-675.

118. Phillips, A.W., 1958, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957," *Economica* **25**: 283-299.

119. Pincheira, P. and R. Alvarez, 2009, "Evaluation of Short Run Inflation Forecasts and Forecasters in Chile," *Money Affairs* **XXII**(2)**:** 159-180.

120. Pincheira, P., 2012, "*Are Forecast Combinations Efficient?*," Working Paper 661, Central Bank of Chile.

121. Pincheira, P. and C.A. Medel, 2012, "*Forecasting Inflation with a Random Walk*," Working Paper 669, Central Bank of Chile.

122. Pincheira, P. and C.A. Medel, 2015. "Forecasting Inflation with a Simple and Accurate Benchmark: The Case of the US and a Set of Inflation Targeting Countries," *Czech Journal of Economics and Finance* **65**(1): 2-29.

123. Pollock, D.S.G., 2014, "*Econometric Filters*," Working Paper 14/07, Department of Economics, University of Leicester, UK.

124. Posch, J. and F. Rumler, 2015, "Semi-Structural Forecasting of UK Inflation Based on the Hybrid New Keynesian Phillips Curve," *Journal of Forecasting* **34**: 145-162.

125. Rabanal, P. and J.F. Rubio, 2005, "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics* **52**: 1151-1166.

126. Rissasen, J., 1978, "Modeling by Shortest Data Description," *Automatica* **14**(5): 465-471.

127. Roberts, J.M., 1997, "Is Inflation Sticky?," *Journal of Money, Credit and Banking* **39**(2): 173-196.

128. Robinson, T., A. Stone, and M. van Zyl, 2003, "The Real Time Forecasting Performance of Phillips Curves," Research Discussion Paper 2003-12, Reserve Bank of Australia.

129. Rudd, J. and K. Whelan, 2005, "New Tests of the New-Keynesian Phillips Curve," *Journal of Monetary Economics* **52**(6): 1167-1181.

130. Rudebusch, G.D. and L.E.O. Svensson, 1999, *Policy Rules for Inflation Targeting*, in J.B. Taylor (ed.), *Monetary Policy Rules*, University of Chicago Press, US.

131. Rumler, F., 2007, "Estimates of the Open Economy New Keynesian Phillips Curve for Euro Area Countries," *Open Economies Review* **18**(4): 427-451.

132. Rumler, F. and M.T. Valderrama, 2010, "Comparing the New Keynesian Phillips Curve with Time Series Models to Forecast Inflation," *The North American Journal of Economics and Finance* **21**(2): 126-144.

133. Schwarz, G.E., 1978, "Estimating the Dimension of a Model," *Annals of Statistics* **6**(2): 461-464.

134. Shibata, R., 1976, "Selection of the Order of an Autoregressive Model by Akaike Information Criterion," *Biometrika* **63**(1): 117-126.

135. Smets, F. and R. Wouters, 2003, "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association* **1**(5): 1123-1175.

136. Smets, F. and R. Wouters, 2005, "Comparing Shocks and Frictions in US and Euro Area Business Cycles: A Bayesian DSGE Approach," *Journal of Applied Econometrics* **20**(2): 161-183.

137. Smets, F. and R. Wouter, 2007, "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review* **97**(3): 586-606.

138. Smith, L.V. 2013, *Short and Medium-term Forecasting using "Pooling" Techniques*, in F. di Mauro and M.H. Pesaran (Eds.), *The GVAR Handbook*, Cambridge University Press, UK.

139. Staiger, D., J.H. Stock, and M.W. Watson, 1997a, *How Precise are Estimates of the Natural Rate of Unemployment?*, in C. Romer and D. Romer (Eds.), *Reducing Inflation: Motivation and Strategy*, Chicago University Press.

140. Staiger, D., J.H. Stock, and M.W. Watson, 1997b, "The NAIRU, Unemployment and Monetary Policy," *Journal of Economic Perspectives* **11**(1): 33-49.

141. Stock, J.H., 2001, *Forecasting Economic Time Series*, in B. Baltagi (Ed.), *A Companion to Theoretical Econometrics*, Blackwell Publishing.

142. Stock, J.H. and M.W. Watson, 1999, "Forecasting Inflation," *Journal of Monetary Economics* **44**(2): 293-335.

143. Stock, J.H. and M.W. Watson, 2009, *Phillips Curve Inflation Forecasts*, in J.F. Fuhrer, Y. Kodrzycki, J.S. Little, and G. Olivei (Eds.), *Understanding Inflation and the Implications for Monetary Policy, A Phillips Curve Restrospective*, MIT Press, US.

144. Stone, M., 1979, "Comments on Model Selection Criteria of Akaike and Schwarz," *Journal of the Royal Statistical Society, Series B (Methodological)* **41**(2): 276-278.

145. van Ruth, F., 2014, "*Inflation Dynamics in the Netherlands. A Linear and Non-linear Analysis and the Influence of Economic Conditions*," Discussion Paper 22, Statistics Netherlands.

146. Vašíček, C., 2011, "Inflation Dynamics and the New Keynesian Phillips Curve in Four Central European Countries," *Emerging Markets Finance and Trade* **47**(5): 71-100.

147. Weakliem, L.D., 2004, "Introduction to the Special Issue on Model Selection," *Sociological Methods and Research* **33**: 167-186.

148. Zucchini W., 2000, "An Introduction to Model Selection," *Journal of Mathematical Psychology* **44**: 41-46.

A Data description and sources

In this Annex it is described the dataset in terms of its sources for further replication/checking purposes.

Table A1: Variable description (*)

Variable	Country	Unity	Scale	Descriptor	Source
Consumer Price Index (transformed to inflation series)	US	Index	2010=100	Consumer Prices - All Items	OECD Database
	CAN	Index	2010=100	Consumer Prices - All Items	OECD Database
	EUR	Index	2010=100	Harmonised CP (19 countries)	OECD Database
	JPN	Index	2010=100	Consumer Prices - All Items	OECD Database
	SWI	Index	2010=100	Consumer Prices - All Items	OECD Database
	UK	Index	2010=100	Consumer Prices - All Items	OECD Database
Inflation Expectations	US	Basis points	None	Ave. % chg. on prev. yr.	Consensus Economics
	CAN	Basis points	None	Ave. % chg. on prev. yr.	Consensus Economics
	EUR	Basis points	None	Ave. % chg. on prev. yr.	Consensus Economics
	JPN	Basis points	None	Ave. % chg. on prev. yr.	Consensus Economics
	SWI	Basis points	None	Ave. % chg. on prev. yr.	Consensus Economics
	UK	Basis points	None	Ave. % chg. on prev. yr.	Consensus Economics
Industrial Production (used for the output gap)	US	Index	2010=100	Production of total industry sa	OECD Database
	CAN	Index	2010=100	Production of total industry sa	OECD Database
	EUR	Index	2010=100	Production of total industry sa	OECD Database
	JPN	Index	2010=100	Production of total industry sa	OECD Database
	SWI	Index	2010=100	Total retail trade (volume)	OECD Database
	UK	Index	2010=100	Production of total industry sa	OECD Database

(*) "sa" stand for seasonally adjusted. Source: Author's elaboration.