Forecasting With High Dimensional Panel VARs

Gary Koop[a], Dimitris Korobilis[b]

[a]*University of Strathclyde*
[b]*University of Glasgow*

Abstract

In this paper, we develop econometric methods for estimating large Bayesian time-varying parameter panel vector autoregressions (TVP-PVARs) and use these methods to forecast inflation for euro area countries. Large TVP-PVARs contain huge numbers of parameters which can lead to over-parameterization and computational concerns. To overcome these concerns, we use hierarchical priors which reduce the dimension of the parameter vector and allow for dynamic model averaging or selection over TVP-PVARs of different dimension and different priors. We use forgetting factor methods which greatly reduce the computational burden. Our empirical application shows substantial forecast improvements over plausible alternatives.

Keywords: Panel VAR, inflation forecasting, Bayesian, time-varying parameter model

1. Introduction

As the economies of the world become increasingly linked through trade and financial flows, the need for multi-country econometric models has increased. Panel Vector Autoroegressions (PVARs) which jointly model many macroeconomic variables in many countries are a popular way of fulfilling this need. In this paper, we develop econometric methods for TVP-PVARs which can help overcome the concerns about over-parameterization that arise in these models and do so in a computationally-feasible manner. The contributions made lie in the econometric methods and in the application.

To explain the significance of the econometric contributions of our paper, we note the combination of the growing interesting in high-dimensional multivariate time series models such as VARs (see, among many others, Banbura, Giannone and Reichlin, 2010, Carriero, Kapetanios and Marcellino, 2010, 2012, Koop, 2013 and Korobilis, 2013) with the recognition that allowing for coefficient change can be empirically necessary has led to a need for the development of econometric methods for large TVP-VARs. In the large VAR literature, it has become standard to work with models involving dozens or even a hundred or more dependent variables. This leads to over-parameterized models with

Email addresses: `Gary.Koop@strath.ac.uk` (Gary Koop), `Dimitris.Korobilis@glasgow.ac.uk` (Dimitris Korobilis)

huge parameter spaces. An unrestricted large TVP-VAR will additionally have parameters which control the time-variation in VAR coefficients. Such huge parameter spaces raise two issues: how to achieve parsimony and how to achieve efficient computation. Bayesian methods are commonly used with these models and prior shrinkage of various sorts[1] is typically used to address the first of these. In the standard homoskedastic VAR, priors can be used which lead to analytical posterior and predictive results and, thus, easy computation. But, once one leaves the restrictive homoskedastic, constant coefficient world, analytical results are not available and Markov Chain Monte Carlo (MCMC) methods are required. With larger models, the computational demands of MCMC methods can be prohibitive. This has led to various approximate methods being proposed. Our earlier work with large TVP-VARs, Koop and Korobilis (2013), offers one example of how the researcher can over-come the computational hurdle. In it, we use approximate methods involving forgetting factors, shrinkage priors and model switching (i.e. switching between more parsimonious models nested within the large TVP-VAR) which allow for computationally feasible inference in large TVP-VARs.

One econometric contribution of our paper lies in extending Koop and Korobilis (2013) to the panel context. TVP-PVARs, which involve a TVP-VAR for each of several countries, can be very large indeed. For instance, an unrestricted PVAR with 4 lags involving 7 dependent variables for each of 19 countries will have 70,756 PVAR coefficients and 8,246 free parameters in the error covariance matrix to estimate. Adding time-variation in PVAR coefficients and stochastic volatility in the errors increases this number substantially. A TVP-PVAR is, in a sense, simply a large TVP-VAR. Hence, the reader may wonder why the methods of Koop and Korobilis (2013), or some other method for estimating large TVP-VARs, cannot simply be used for the large TVP-PVAR. However, in over-parameterized models where the data information (i.e. the number of observations) is small relative to the number of parameters to be estimated, the role of prior information increases in importance. For the Bayesian, sensible prior elicitation is always an important issue, but in high-dimensional cases such as the TVP-PVAR it becomes absolutely essential. Simply using a standard VAR prior (e.g. the Minnesota prior) for the PVAR, ignoring the panel nature of the PVAR, can potentially have negative consequences. Similarly, using a standard TVP-VAR prior (e.g. assuming each coefficient evolves according to a random walk) with a TVP-PVAR, can lead to an excessively parameter-rich model and misleading inference. Accordingly, developing econometric methods for the TVP-PVAR is not simply a matter of using existing TVP-VAR methods. In this paper, we build on an approach suggested for PVARs in Canova and Ciccarelli (2009, 2013) to our TVP-PVAR. The result is a computationally efficient approach, suitable for estimating very large TVP-PVARs that explicitly takes the panel nature of the problem into account. In addition, we extend the approach of Canova and Ciccarelli (2009) to allow for time-varying error covariances. We rewrite the TVP-PVAR model in a form which allows us to shrink the (large) time-varying

[1]Priors used in this literature are often hierarchical or objective in the sense that key prior hyperparameters are estimated from the data. With TVP models, coefficients are often assumed to evolve according to a state equation. This state equation is a hierarchical prior.

error covariance matrix using a hierarchical prior.

A second econometric contribution lies in the treatment of model uncertainty. There are typically a large number of specification choices in the TVP-PVAR. Furthermore, the optimal specification may be changing over time. To deal with model uncertainty in a dynamic fashion, we adapt dynamic model averaging (DMA) and dynamic model selection (DMS) methods initially developed for use with single equation TVP regression models by Raftery, Karny and Ettler (2010), for the TVP-PVAR. We do this in relation to several aspects of the specification, but the most important relate to the dimension of the TVP-PVAR and hierarchical prior used to accommodate the panel nature of the problem. In our application, there is uncertainty over the appropriate TVP-PVAR dimension. Instead of just choosing a particular dimension, we consider a large set of TVP-PVARs where each is defined by a different combination of the dependent variables. When doing DMA, this allows us to attach more weight to the appropriate dimension in a dynamic fashion. So, for instance, at some points in time forecasts from a particular TVP-PVAR can receive most weight, then we can switch to a TVP-PVAR involving different dependent variables and attach more weight to its forecasts. Another advantage of using DMS is that our unrestricted TVP-PVAR is very large and potentially over-parameterized, but DMS can achieve parsimony by choosing a lower dimensional model.

With regards to the hierarchical prior, we use the one suggested in Canova and Ciccarelli (2009) as well as a new one which shrinks parameters toward country-specific VARs. Treating these two priors as defining two models, we can use DMA methods to decide which prior leads to better forecasts and adjust weights associated with the two priors appropriately in a dynamic fashion.

Our paper also seeks to contribute to the empirical literature on inflation forecasting in the eurozone. There are many linkages and inter-relationships between the economies of the eurozone countries. Modelling aggregate inflation for the eurozone as a whole will miss many interesting country-specific patterns since monetary policy can have differing impacts on different countries. These considerations justify why we want to forecast individual country inflation rates, but not using conventional VAR methods one country at a time. The panel VAR is an effective way of modelling the spillovers and inter-linkages between countries that, no doubt, exist for the eurozone countries.

Euro area inflation has been an important component in many recent policy discussions. Deflation has been a recent worry. For instance, in December 2014, 12 of the 18 countries that then comprised the eurozone[2] were experiencing deflation and no country registered an inflation rate above 1%. The impact of quantitative easing by the ECB in this context is of great policy interest. Although there has been a tendency for inflation rates in the various eurozone countries to converge to one another, there are still substantive cross-country differences, in particular around the time of the eurozone crisis. For instance, Delle Monache, Petrella and Venditti (2015) document the relative roles of country-specific and common shocks to euro area country inflation rates. Although commonalities predominate, country specific shocks play a large role. Furthermore, Delle

[2]Lithuania did not join the eurozone until January 2015.

Monache, Petrella and Venditti (2015) document substantial time variation in parameters, providing additional support for our model which allows for such variation.

The remainder of the paper is organized as follows. In Section 2, we describe our econometric methods, beginning with the panel VAR before proceeding to the TVP-PVAR and then our dynamic treatment of model uncertainty. Section 3 contains our empirical study of euro area inflation. We find substantial evidence of forecasting benefits, in particular from using DMA methods which average over different TVP-PVAR dimensions and different hierarchical priors. Section 4 concludes.

2. Econometric Model: The TVP-PVAR

2.1. The Panel VAR

In an increasingly globalized world, where financial or macroeconomic events in one country can spill over to another country, the need for models which accommodate such interlinkages has grown. We use the general term PVAR for models where VARs for each individual country are augmented with lagged dependent variables from other countries. Several different specifications are commonly used, including the multi-country VARs of Canova and Ciccarelli (2009) and the global VARs of, among others, Dees, Di Mauro, Pesaran and Smith (2007) and Feldkircher and Huber (2015). We begin by discussing some of the issues which occur with PVAR models before discussing TVP versions of them.

Suppose we have N countries (in our application, these are the 19 countries that comprise the eurozone as of 2015) and G variables for each country (in our application, these are inflation plus six additional variables which may be useful for predicting inflation) which are observed for T time periods. Let $Y_t = (y'_{1t}, y'_{2t}, ..., y'_{Nt})$ for $t = 1, ..., T$ be the $NG \times 1$ vector of dependent variables where y'_{it} is the $G \times 1$ vector of dependent variables of country i, $i = 1, ..., N$. The i-th equation of the PVAR with p lags takes the form

$$y_{it} = A_i^1 Y_{t-1} + ... + A_i^p Y_{t-p} + u_{it}, \tag{1}$$

where A_i^j for $j = 1, .., p$ are $G \times NG$ matrices PVAR coefficients for country i. Additionally, u_{it} is a $G \times 1$ vector of disturbances, uncorrelated over time, where $u_{it} \sim N(0, \Sigma_{ii})$. The errors between countries may be correlated and we define $E(u_{it} u_{jt}) = \Sigma_{ij}$ and Σ to be the entire $NG \times NG$ error covariance matrix for $u_t = (u_{1t}, .., u_{Nt})'$. Let $A^j = \left(A_1^j, ..., A_N^j\right)$ for $j = 1, ..., p$ and $\alpha = \left(vec(A^1)', ..., vec(A_t^p)'\right)'$. Note that, for notational simplicity, we have not added an intercept or other deterministic terms nor exogenous variables, but they can be added with the obvious modifications to the formulae below. In our empirical work, we include an intercept.

2.2. Methods of Ensuring Parsimony in the PVAR

The unrestricted PVAR given in (1) is likely over-parameterized, involving $K = p \times (N \times G)^2$ unknown autoregressive parameters and $\frac{N \times G \times (N \times G + 1)}{2}$ error covariance terms. Plausible choices for N, G and p can lead to very large parameter spaces. In our application, the main variable of interest is the inflation rate. Forecasting inflation is hard, with many researchers finding univariate forecasting models hard to beat (see Faust and Wright, 2013, for a survey of the inflation forecasting literature). Thus, it is likely that many coefficients in an unrestricted PVAR will be zero. In fact, it is quite possible that a more parsimonious panel model involving only the inflation rates (i.e. our PVAR with $G = 1$ variable: the inflation rate) may forecast better than the less parsimonious model with $G = 7$. Or perhaps a Phillips curve model involving only the inflation and unemployment rates (and, thus, $G = 2$) will forecast well. The point to stress is that we are uncertain about how many variables it is worth including in the PVAR.

These considerations suggest that one way to achieve parsimony would be to investigate PVARs of smaller dimension. This is what we do in this paper. We use a large model space involving PVARs of different dimension and using DMA methods to average over them. To be precise, the unrestricted PVAR has $N \times G$ equations. If we restrict attention to the PVARs which include the G_C core variables of interest (in our case, these are inflation, unemployment and industrial production) for every country, but every other variable may or may not be included, then there are $N \times 2^{G-G_C}$ possible restricted PVARs of interest. In theory, we could average over all these restricted PVARs in a DMA exercise to select the best one in a dynamic fashion. That is, we can potentially give more weight to a PVAR of a particular dimension at one point in time, but attach more weight to a different dimension at other points in time.[3] In practice, searching over every possible restricted PVAR is too computationally demanding in our application. Nevertheless, we do consider a range a PVARs of different dimensions. In particular, the empirical results in our most general model do DMA for $G = 3, 4, .., 7$.[4] In this subsection, we are discussing constant coefficient PVARs, but the same issues hold with TVP-PVARs and will be addressed in the same way in this paper.

Another approach to dimension reduction in PVARs is described in Canova and Ciccarelli (2009, 2013) who use a hierarchical prior so as to work with restricted versions of (1), a practice which we follow in this paper. In particular, we can assume a factor structure for the PVAR coefficients and write:

$$\begin{aligned} \alpha &= \Xi_1\theta_1 + \Xi_2\theta_2 + .. + \Xi_q\theta_q + e \\ &= \Xi\theta + e \end{aligned}$$

where $\Xi = (\Xi_1, .., \Xi_q)$ are known matrices and $\theta = \left(\theta_1', .., \theta_q'\right)'$ is an $R \times 1$ vector of unknown parameters with $R < K$ and e is uncorrelated with u_t and distributed as $N(0, \Sigma \otimes V)$ where $V = \sigma^2 I$.

Suppose, for instance, that the elements of α are made up of a common factor, a factor specific to each country and a factor specific to each variable. This is the factor structure used in Canova and Ciccarelli (2009). Then $q = 3$ and Ξ_1 will be a $K \times 1$ vector of ones,

[3] Koop (2014), in a VAR application, refers to this is dynamic dimension selection.

[4] Allowing for different countries to have different Gs is also theoretically possible, but this becomes much more computationally demanding.

θ_1 a scalar. Ξ_2 will be a $K \times N$ matrix containing zeros and ones defined so as to pick out coefficients for each country and θ_2 is an $N \times 1$ vector. Ξ_3 will be a $K \times G$ matrix containing zeros and ones defined so as to pick out coefficients for each variable and θ_3 is an $G \times 1$ vector. For instance, if $N = G = 2$ and $p = 1$ then

$$\Xi_2 = \begin{bmatrix} \iota_1 & 0 \\ \iota_1 & 0 \\ 0 & \iota_2 \\ 0 & \iota_2 \end{bmatrix} \text{ and } \Xi_3 = \begin{bmatrix} \iota_3 & 0 \\ 0 & \iota_4 \\ \iota_3 & 0 \\ 0 & \iota_4 \end{bmatrix}$$

where $\iota_1 = (1,1,0,0)'$, $\iota_2 = (0,0,1,1)'$, $\iota_3 = (1,0,1,0)'$ and $\iota_4 = (0,1,0,1)'$. Thus, the K dimensional α is dependent on a much lower dimensional vector of parameters, since θ is of dimension $R = 1 + N + G$ with e being left to model any residual variation in the parameters.

Such a strategy can be used to greatly reduce the dimensionality of α and help achieve parsimony. However, such a method may come at a cost if the factor structure is not chosen correctly. The latter could lead either to over-parameterization concerns or to mis-specification concerns. In the previous example, where the coefficients are assumed to depend on a common factor, a country specific factor and a variable specific factor, it could be, e.g., that no common factor exists ($\theta_1 = 0$) and a specification which ignored this restriction would over-parameterized. On the other hand, our example of a factor structure might be too restrictive and mis-specification might result. The K distinct elements of α may be so heterogeneous that a factor structure with only $N + G + 1$ parameters may not be adequate.

These considerations suggest that the model space could be augmented using different choices of Ξ. In theory, DMA could be done over a huge range of possible structures for Ξ. In practice, computational concerns lead us in this paper to consider two structures. The first is identical to that used in Canova and Ciccarelli (2009). The second we call the country-specific VAR factor structure. To explain what we mean by this, let $p = 1$ and consider the NG^2 coefficients in the VAR for country i. G^2 of these coefficients are on lags of country i variables, with the remaining $(N-1)\,G^2$ being on lags of other countries' variables. We define Ξ such that its accompanying θ loads only on the G^2 coefficients that are on lags of country i variables. Thus, if $e = 0$, the coefficients on other country variables are zero and the PVAR breaks down into N individual VARs, one for each country (apart from any inter-linkages which occur through Σ). The impacts of other countries' variables on country i are only allowed for through the presence of e. Intuitively, this structure for Ξ captures the idea that working with VARs one country at a time comes close to being adequate (i.e. most of the coefficients on lagged country j variables in the country i VAR will be zero), but there are occasional inter-linkages which can be captured through e. When we move to the TVP-PVAR in the next section, this definition of Ξ will imply the same intuition, except in terms of individual-country TVP-VARs.

2.3. Moving from the PVAR to the TVP-PVAR

There are many ways that the coefficients in the PVAR can be made to be time varying. In this paper, we use a specification suggested in Canova and Ciccarelli (2009) which uses the factor structure described in the preceding subsection. The basic idea is allow for random walk behavior of θ instead of the high-dimensional α, resulting in a state space model which can be estimated using standard state space methods.

We begin by putting t subscripts on all the PVAR coefficients in (1) and, thus, $\alpha_t = \left(vec\left(A_t^1\right)', ..., vec\left(A_t^p\right)'\right)'$ is the $K \times 1$ vector collecting all PVAR parameters at time t. We write the TVP-PVAR in matrix form as:

$$Y_t = X_t'\alpha_t + u_t, \tag{2}$$

where $X_t = I \otimes \left(Y_{t-1}', ..., Y_{t-p}'\right)'$, and $u_t \sim N\left(0, \Sigma_t\right)$. An unrestricted TVP-VAR would typically assume α_t to evolve as a random walk (see, e.g., Doan, Litterman and Sims, 1984, Cogley and Sargent, 2005, or Primiceri, 2005). However, in the multi-country TVP-PVAR case this may lead to an extremely over-parameterized model and burdensome (or even infeasible) computation. The over-parameterization concerns can be clearly seen. Even in the constant coefficient case, the number of PVAR parameters, $p \times (N \times G)^2$, could run into the thousands or more. Allowing them to be time-varying greatly increases their number. With regards to computation, TVP-VARs are typically estimated using MCMC methods. Even with small TVP-VARs, MCMC running time for a single model is measured in minutes or hours on a modern PC. Repeatedly running such an algorithm on an expanding window of data, as is typically done in a recursive forecasting exercise, multiplies this burden by hundreds in many applications. And repeating this whole process for hundreds or thousands or more models multiplies it many times more. These statements hold for small TVP-VARs, but are proportionally worse for large TVP-VARs. As discussed in Koop or Korobilis (2013), estimating TVP-VARs using MCMC methods can easily become computationally infeasible unless the number of models and their dimension are both small.

In order to achieve parsimony, we use the same methods for reducing the dimension of the TVP-VAR using DMA methods described in the preceding section. As a further step towards reducing over-parameterization problems, we follow Canova and Ciccarelli (2009) and extend the factorization of the PVAR coefficients described in the preceding subsection to the time-varying case using the following hierarchical prior:

$$\alpha_t = \Xi\theta_t + e_t \tag{3}$$
$$\theta_t = \theta_{t-1} + w_t, \tag{4}$$

where θ_t is an $R \times 1$ vector of unknown parameters, Ξ is defined as in the preceding subsection and $w_t \sim N\left(0, W_t\right)$ where W_t is an $R \times R$ covariance matrix.

The hierarchical representation of the panel VAR using equations (2), (3) and (4) resembles the hierarchical time-varying parameter SUR specified in Chib and Greenberg (1995). Canova and Ciccarelli (2009) use this representation to reduce the dimensionality of the $K \times 1$ coefficient vector α_t by assuming it to be driven by θ_t which is of significantly lower dimensions, that is, $R \lll K$. Extending Canova and Ciccarelli (2009)'s homoskedastic specification we let $e_t \sim N(0, \Sigma_t \otimes V)$ where $V = \sigma^2 I$.

An equivalent way of writing the TVP-PVAR given by (2), (3) and (4) is:

$$Y_t = \widetilde{X}_t' \theta_t + v_t, \tag{5}$$

$$\theta_t = \theta_{t-1} + w_t, \tag{6}$$

where $\widetilde{X}_t = X_t \Xi$ and $v_t = X_t' e_t + u_t$ with $v_t \sim N\left(0, \left(I + \sigma^2 X_t' X_t\right) \times \Sigma_t\right)$. Therefore, in this form the TVP-PVAR is written as a linear Gaussian state-space model consisting of the measurement equation in (5) and the state equation (6).

For known values of Σ_t, W_t and σ^2, standard methods for state space models based on the Kalman filter can be used to obtain the predictive density and posterior distribution for θ_t. Thus, we will not repeat the relevant formulae here (see, e.g., Durbin and Koopman, 2001). A typical Bayesian analysis would involve using MCMC methods to draw Σ_t, W_t and σ^2 and then, conditional on these draws, use such state space methods. However, in our case, the computational burden of MCMC methods will be prohibitive. Accordingly, we use: i) forgetting factor methods to provide an estimate of W_t, ii) Exponentially Weighted Moving Average (EWMA) methods to estimate Σ_t and, iii) use a grid of values for σ^2 and interpret each value as defining a particular model and, thus, include them in our model space over which we do DMA. The following paragraphs elaborate on these points.

Dynamic model averaging methods were pioneered in Raftery, Karny and Ettler (2010) and, within each model included in the model space, forgetting factor methods were used for forecasting. The benefit of this was to provide an estimate of W_t, thus avoiding the need for use of MCMC methods. We refer the reader to Raftery, Karny and Ettler (2010) or to Koop and Korobilis (2013) for a motivation and discussion of the properties of forgetting factor methods. The main idea is to estimate W_t as

$$\widehat{W}_t = \left(\frac{1}{\lambda} - 1\right) var\left(\theta_t | \mathcal{D}_{t-1}\right),$$

where $\mathcal{D}_{t-1}$ denotes data available through period $t-1$, $0 < \lambda \leq 1$ is the forgetting factor and $var\left(\theta_t | \mathcal{D}_{t-1}\right)$ is a quantity readily available from the Kalman filter iteration at time $t-1$. Typically, λ is set to a number slightly below one. The forgetting factor approach allows estimation of systems with large number of variables in seconds, and, hence, is computationally attractive for recursive point and density forecasting or any other state space modelling exercise that can become infeasible using MCMC methods.

In contrast to Canova and Ciccarelli (2009), we allow for the TVP-PVAR error covariance matrix to be time-varying and use EWMA filtering methods to estimate it as:

$$\widehat{\Sigma}_t = \kappa \widehat{\Sigma}_{t-1} + (1-\kappa) \widetilde{u}_t \widetilde{u}_t',$$

where $\widetilde{u}_t \widetilde{u}_t' = \left(I + \sigma^2 X_t' X_t\right)^{-1} \left[\left(Y_t - \widetilde{X}_t' E\left(\theta_t | \mathcal{D}_{t-1}\right)\right)\left(Y_t - \widetilde{X}_t' E\left(\theta_t | \mathcal{D}_{t-1}\right)\right)'\right]$, $E\left(\theta_t | \mathcal{D}_{t-1}\right)$ is produced by the Kalman filter and $0 < \kappa \leq 1$. κ is referred to as a decay factor. We define the $\kappa = 1$ case to be $\widehat{\Sigma}_t = \frac{\sum_{\tau=1}^{t} \widetilde{u}_\tau \widetilde{u}_\tau'}{t}$. In order to initialize $\widehat{\Sigma}_t$, we set $\widehat{\Sigma}_0 = 0.1 \times I$ which is a relatively diffuse choice.

The forgetting factor λ and decay factor κ control the amount of time variation in the system. Lower (higher) values of λ, κ imply faster (slower) changes over time in the values of θ_t and Σ_t, respectively. When $\lambda = \kappa = 1$ then both θ_t and Σ_t become time invariant and we have the constant parameter homoskedastic PVAR. In the most general model used in our empirical work, we let $\lambda \in \{0.99, 1\}$ and $\kappa = \{0.94, 0.96, 1\}$, interpret each grid point as defining a model and use DMS to select the optimal value. Thus, the data can select either the constant coefficient PVAR or homoskedastic PVAR at any point in time, or can select a greater degree of variation in coefficients or error covariance matrix. We adopt a similar strategy for σ^2, using a grid of $\sigma^2 \in \{0.001, 0.003, 0.005, 0.007, 0.009,\ 0.01, 0.03, 0.05, 0.07, 0.09,\ 0.1, 0.3, 0.5, 0.7, 0.9,\ 1, 3, 5, 7, 9\}$.

The Kalman filter provides us with a one-step ahead predictive density. Since we wish to forecast at horizon $h > 1$ and calculate predictive likelihoods, we use predictive simulation. To do this, we draw Y_{T+1} from its Gaussian predictive density with mean and variance given by the Kalman filter (these are assumed to be constant and fixed during predictive simulation), then simulate Y_{T+2} from its Gaussian predictive density conditional on the drawn Y_{T+1}, etc. up to h.

2.4. A Hierarchical Prior for the Error Covariance Matrix

As we have seen, the error covariance matrix of the TVP-PVAR can also be huge, leading to a desire for shrinkage on it as well. In this subsection, we extend the hierarchical prior of Canova and Ciccarelli (2009) to allow for such shrinkage. We decompose the error covariance matrix as $\Sigma_t = B_t^{-1} H_t \left(H_t B_t^{-1} \right)'$ where B_t is a lower triangular matrix with ones on the diagonal, H_t is a diagonal matrix and write the VAR as

$$\begin{aligned} Y_t &= X_t' \alpha_t + B_t^{-1} H_t \varepsilon_t \implies \\ B_t Y_t &= X_t' \gamma_t + H_t \varepsilon_t. \end{aligned}$$

where $\gamma_t = B_t \alpha_t$ and $\varepsilon_t \sim N(0, I)$. We can write the model in the following form (see also the Appendix of Primiceri, 2005) as:

$$Y_t = X_t' \gamma_t + Z_t' \beta_t + H_t \varepsilon_t,$$

where Z_t is the matrix

$$Z_t = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ -Y_{1t} & 0 & \cdots & 0 \\ 0 & [-Y_{1t}, -Y_{2t}] & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \left[-Y_{1t}, \ldots, -Y_{(NG-1)t}\right] \end{bmatrix}.$$

With this specification we have an equivalent model where the error covariance matrix shows up as contemporaneous regressors in the RHS of the TVP-PVAR. We can extend our previous approach and introduce a hierarchical prior on both γ_t (which are now the VAR coefficients in this formulation) and β_t of the form:

$$\begin{aligned} \delta_t &\equiv \begin{bmatrix} \gamma_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \Xi^\gamma & 0 \\ 0 & \Xi^\beta \end{bmatrix} \theta_t + u_t \equiv \Xi\theta_t + u_t, \\ \theta_t &= \theta_{t-1} + v_t. \end{aligned} \tag{7}$$

where now $u_t \sim N\left(0, H_t \otimes (\sigma^2 I)\right)$, which has a diagonal covariance matrix since both H_t and $\sigma^2 I$ are diagonal matrices. Under the additional assumption that Ξ is block diagonal and γ_t and β_t load on separate factors (rows of θ_t), then we have prior independence between the two sets of coefficients.[5] The econometric methods described in the preceding subsection can be used directly, with a slight simplification due to the diagonality of H_t.

The two choices for Ξ^γ are those described at the end of Section 2.2. For Ξ^β we also use these two choices with the trivial adaptation required by the structure for Z_t. In our DMA exercise, we allow for Ξ^γ and Ξ^β to be different. Using 1 subscripts to denote the form suggested in Canova and Ciccarelli (2009) and 2 subscripts to denote the country-specific VAR factor structure, we define four models which have: i) Ξ_1^γ and Ξ_1^β, ii) Ξ_1^γ and Ξ_2^β, iii) Ξ_2^γ and Ξ_1^β and iv) Ξ_2^γ and Ξ_2^β.

2.5. *Dynamic Treatment of Model Uncertainty*

The previous subsections discussed the estimation of single TVP-PVARs and defined our model space. Our most general approach contains models which differ in the dimension of the TVP-PVAR (5 choices are used, see subsection 2.2), choice of factor structure for Ξ (4 choices are used, see subsection 2.4), choice of forgetting factor, λ (two choices are used, see subsection 2.3), choice of decay factor, κ (three choices are used, see subsection 2.3) and choice of σ^2 (20 choices are used, see subsection 2.3). Overall, this approach involves 2400 models. We wish to average over them or select between them in a dynamic fashion. This subsection describes our methods for doing so.

Let $M^{(i)}$ for $i = 1, ..., J$ be the set of models under consideration. In our application, J is very large and we use DMA and DMS methods so as to navigate this vast model space in a dynamic fashion. When forecasting time t inflation rates using data available at time $t-1$, these methods (see Raftery, Karny and Ettler, 2010) involve calculation of a model probability, $p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$, for each model. DMS forecasts using the single model with the highest value for $p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$. DMA uses forecasts averaged over all models with model i receiving weight $p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$ in the average. We use forgetting factor methods to estimate $p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$.

[5] The assumption that the errors in the various state-space equations that characterize different blocks of time-varying parameters are uncorrelated with each other is typically made in the TVP-VAR literature. Here we want the independence assumption for the additional reason that the state-space model for β_t has Y_t showing up both on the LHS and the RHS (through Z_t) of (7). However, as Primiceri (2005) shows on page 845, given the lower triangularity of B_t, if we allow $var\left(\beta_t\right)$ to be diagonal (or block diagonal), then the state space model for β_t is conditionally linear and the Kalman filter can be applied.

The forgetting factor literature (e.g. Kulhavý and Kraus, 1996 and Raftery, Karny and Ettler, 2010) provides derivations and additional motivation for how sensible estimates for $p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$ can be produced in a fast, recursive manner, in the spirit of the Kalman filtering approach. Here we outline the basic steps, following the exponential forgetting factor approach of Kulhavý and Kraus (1996). Let $\omega_{t|t-1}^{(i)} = p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$ be the probability associated with model i for forecasting Y_t using data available through time $t-1$. The general version of the algorithm combines a prediction step

$$\omega_{t|t-1}^{(i)} = \frac{\left(\omega_{t-1|t-1}^{(i)}\right)^{\mu}}{\sum_{i=1}^{J}\left[\left(\omega_{t-1|t-1}^{(i)}\right)^{\mu}\right]}, \tag{8}$$

with an updating step

$$\omega_{t|t}^{(i)} \propto \omega_{t|t-1}^{(i)} p\left(Y_t|M^{(i)}, \mathcal{D}_{t-1}\right), \tag{9}$$

with a normalizing constant to ensure the $\omega_{t|t}^{(i)}$ sum to one. $p\left(Y_t|M^{(i)}, \mathcal{D}_{t-1}\right)$ is the predictive density produced by the Kalman filter, evaluated at the realized value for Y_t. The recursions begin with an initial condition for the weights, which we set at $\omega_{0|0}^{(i)} = \frac{1}{J}$ (i.e. all models have equal prior probability).

The quantity $0 < \mu \leq 1$ is a forgetting factor used to discount exponentially more distant observations in a similar fashion to λ. Since $p\left(Y_t|M^{(i)}, \mathcal{D}_{t-1}\right)$ is a measure of forecast performance, it can be seen that this approach attaches more weight to models which have forecast well in the recent past. To see this clearly, note that (8) can be written as

$$\omega_{t|t-1}^{(i)} \propto \prod_{i=1}^{t-1}\left[p\left(Y_t|M^{(i)}, \mathcal{D}_{t-1}\right)\right]^{\mu^i}.$$

With monthly data and $\mu = 0.99$, this equation implies that forecast performance one year ago receives about 90% as much weight as forecast performance last period, two years ago receives about 80% as much weight, etc. This is the value used by Raftery, Karny and Ettler (2010) and in our empirical work.

We alter this algorithm in a minor way to take account for the fact that some of our models differ in Y_t (see subsection 2.2). To surmount this problem, $p\left(Y_t|M^{(i)}, \mathcal{D}_{t-1}\right)$ is replaced by $p\left(Y_t^C|M^{(i)}, \mathcal{D}_{t-1}\right)$ where Y_t^C is the set of variables which are common to all models. In our application, these are the three variables which are included in our smallest TVP-PVAR. These are inflation, the unemployment rate and industrial production.

It is possible to do either DMA or DMS over any or all of the modelling choices described at the beginning of this subsection. In our empirical work, we do DMS over λ, κ and σ^2 (i.e. decay and forgetting factors and parameters) and DMA over VAR dimension and choices for Ξ^{γ} and Ξ^{β} (i.e. these might be thought of as more conventional model specification choices). We refer to the approach which does the things as TVP-PVAR-DMA. We follow Canova and Ciccarelli (2009) and set $p = 1$ for all our specifications involving TVP-PVARs.

3. Forecasting Euro Area Inflation

3.1. Data

We use $G = 7$ macroeconomic series for $N = 19$ Eurozone countries for the period 1999M1-2014M12. All variables are transformed so as to be rates (e.g. inflation rate, unemployment rate, etc.), as shown in the last column of the following table, where $\Delta \ln$ denotes first log differences (growth rates), and lev denotes that the variable remains in levels and is not transformed. All variables are seasonally adjusted. Thus, the largest models we work with have 133 dependent variables. We also consider smaller models with $G = 3$ (inflation, unemployment and industrial production), $G = 4$ (adding REER to the $G = 3$ choices), $G = 5$ (adding SURVEY1 to $G = 4$) and $G = 6$ (adding SURVEY2 to $G = 5$). The 19 countries are: Austria (AT), Belgium (BE), Cyprus (CY), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (GR), Ireland (IE), Italy (IT), Latvia (LV), Lithuania (LT), Luxembourg (LU), Malta (MT), Netherlands (NL), Portugal (PT), Slovakia (SK), Slovenia (SI) and Spain (ES).

Variables	**Explanation**	**Source**	**Tr**
HICP	Indices of Consumer Prices	Eurostat	$\Delta \ln$
IP	Industrial production index	IMF IFS	$\Delta \ln$
UN	Harmonised unemployment rates (%)	Eurostat	lev
REER	Real Effective Exchange Rate	Eurostat	$\Delta \ln$
SURVEY1	Financial situation over the next 12 months	Eurostat	lev
SURVEY2	General economic situation over the next 12 months	Eurostat	lev
SURVEY3	Price trends over the next 12 months	Eurostat	lev

Figure 1 plots the inflation rates for the 19 countries. This figure is included only to give the reader a rough impression of the patterns in country-specific inflation rates (labelling individual countries would make the figure even more difficult to read). The main point we wish to note is that there are clearly some general co-movements between the series, but that individual country movements are also very important. For example, the line where the quarter-on-quarter inflation rate reaches almost 6% in 1999 is for Slovakia. It is much more erratic, particularly early in the sample, than other countries. Similarly, at the very end of the sample while all countries are struggling with deflation, two countries in particular have much lower inflation rates. These are Greece and Cyprus.

3.2. Estimation Using TVP-PVAR-DMA

Before comparing the forecasting performance of TVP-PVAR-DMA to some popular alternatives, it is useful to see which specification choices are receiving the most weight in our model averaging exercise.

Figure 1 sheds light on which choices for Ξ are supported by the data. It plots the probability DMA attaches to each of the four possible combinations of the form for Ξ

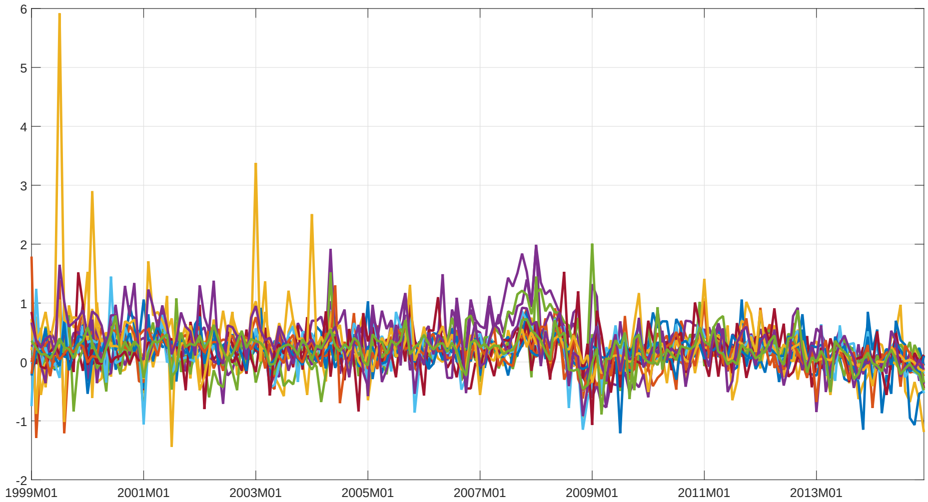

Figure 1: Figure 1: Inflation rates (quarter-on-quarter) for 19 eurozone countries

suggested by Canova and Ciccarelli (2009) and our country-specific VAR factor structure. The most important finding is that these probabilities are changing substantially over time. Any methodology that uses a single Ξ choice to hold over the entire time period risks misspecification and poor forecast performance. For much of the time, the model which uses the Canova and Ciccarelli (2009) choice for Ξ, but the country-specific VAR factor structure for the parameters controlling the error covariances receives strong support. It is also interesting that before the financial and eurozone crises, the combination $\{\Xi_2^{\gamma}, \Xi_2^{\beta}\}$ has high probability in some periods. This is the structure which says all factors are country-specific and dynamic interdependencies between countries are weak. However, after the crises hit, the combination $\{\Xi_1^{\gamma}, \Xi_1^{\beta}\}$ becomes important, suggesting that co-movements between countries have increased.

Figure 2 presents evidence on VAR dimension. Until 2009 there is strong support for small models: the TVP-PVAR using only three variables for each country is selected with probability near one in virtually every time period. However, after 2009 a great deal of dimension switching occurs. In less stable times, it seems that additional sources of information included in these higher dimensional models can be useful when forecasting.

Figure 3 relates to σ^2 and plots the optimal value selected by DMS at each point in time for the four different combinations of $\Xi^{\gamma}, \Xi^{\beta}$. Here again we can see a great deal of time variation in the optimal choice for this parameter. There is also a fair degree of sensitivity

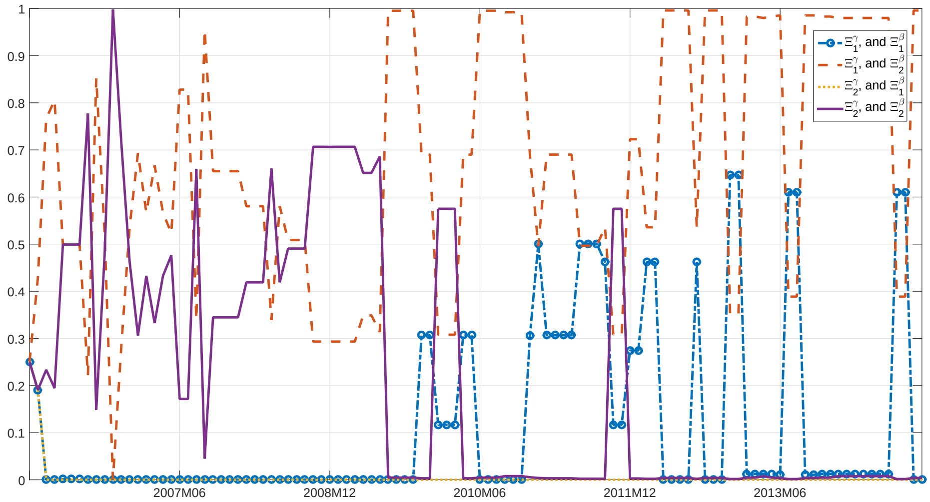

Figure 2: Figure 2: Probability DMA attaches to different choices for Ξ

to the structure of Ξ. When we use Canova and Ciccarelli (2009)'s choice of Ξ, then the optimal value of σ^2 is much lower than when using using the country-specific factor structure. This pattern is somewhat consistent with Canova and Ciccarelli (2009) who in their empirical work set $\sigma^2 = 0$. However, when we use country-specific restrictions, σ^2 tends to be different from zero. It is clearly worth estimating this parameter from the data in this case.

In this subsection, we have presented evidence that TVP-VAR-DMA captures important patterns in the data in a manner that a single model could not. But, ultimately, the test of our approach lies in forecasting and it is to this we now turn.

3.3. Forecasting using TVP-PVAR-DMA

3.3.1. Models for Comparison

We compare our TVP-PVAR-DMA approach to several potential competitors: i) individual country TVP-PVARs (19-TVP-VAR), ii) a large TVP-VAR (LTVP-VAR) using the methods of Koop and Korobilis (2013), iii) a dynamic factor model (DFM), iv) the single TVP-PVAR without DMA (TVP-PVAR) and v) the single TVP-PVAR extended to place a factor structure on the error covariances as in (7) which we refer to as TVP-PVAR-X. We describe in this subsection how we forecast with these models. To aid in comparability, we use the same forgetting factor and EWMA approaches to estimating time variation in coefficients and error covariance matrices in all models and set $\lambda = 0.99$ and $\kappa = 0.96$ in all cases. Some of the approaches use a Minnesota prior for the initial states of time-varying parameters (see Koop and Korobilis, 2013). In such cases, we use a single shrinkage

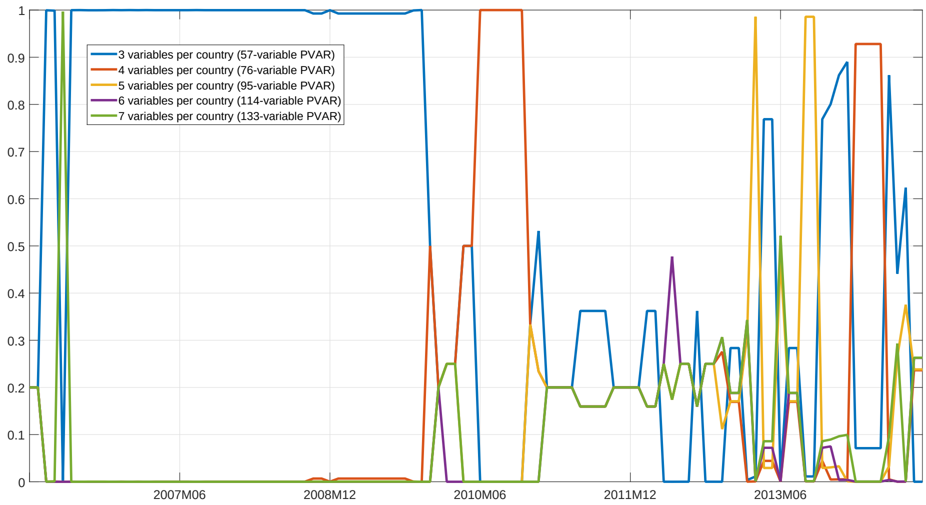

Figure 3: Figure 3: Probability DMA attaches to different choices for G

parameter as in, e.g., Banbura, Giannone and Reichlin (2010). We use a grid of eight different values for this shrinkage parameter: $[1e-10, 1e-5, 0.001, 0.01, 0.05, 0.1, 1, 5]$ and select the optimal one at each point in time as in Giannone, Lenza and Primiceri (2015).

For the 19-TVP-VARs, we forecast with 19 separate 3-variable TVP-VARs: one for each country. For the initial states for the time-varying VAR coefficients, we use a Minnesota prior. For these smaller models we use a maximum of four lags, given that the shrinkage Minnesota prior will take care any overparametrization concerns.

The LTVP-VAR is estimated as in Koop and Korobilis (2013) using a single 57 variable TVP-VAR, that is, the three core variables (inflation, unemployment and industrial production) for the 19 countries. We use a Minnesota prior for the initial states. Due to the dimension of this model we use a maximum of one lag.

The DFM is given by

$$\begin{aligned} Y_t &= F_t'\Lambda + \varepsilon_t, \\ F_t &= F_{t-1}'\Phi_{1t} + ... + F_{t-4}'\Phi_{4t} + u_t. \end{aligned}$$

The factors in the first equation, F_t, are extracted using principal components methods. The second equation is needed for multi-step ahead forecasting. That is, if we let subscripts $t+h|t$ denote forecasts of $t+h$ variables using data available at time t, then $Y_{t+h|h} = F_{t+h|h}'\Lambda$. Y_t contains the same 57 variables as in the LTVP-VAR. The time-variation in $\Phi_t = (\Phi_{1t}, \Phi_{2t}, \Phi_{3t}, \Phi_{4t})$ is estimated using the forgetting factor approach described at the beginning of this subsection and its initial state has a Minnesota prior.

The final two approaches are special cases of our TVP-PVAR-DMA. Neither does DMA over Ξ nor over VAR dimension. Both use Canova and Ciccarelli (2009)'s choice for Ξ and $G = 7$. The first of these approaches, which we label TVP-PVAR in the tables, uses

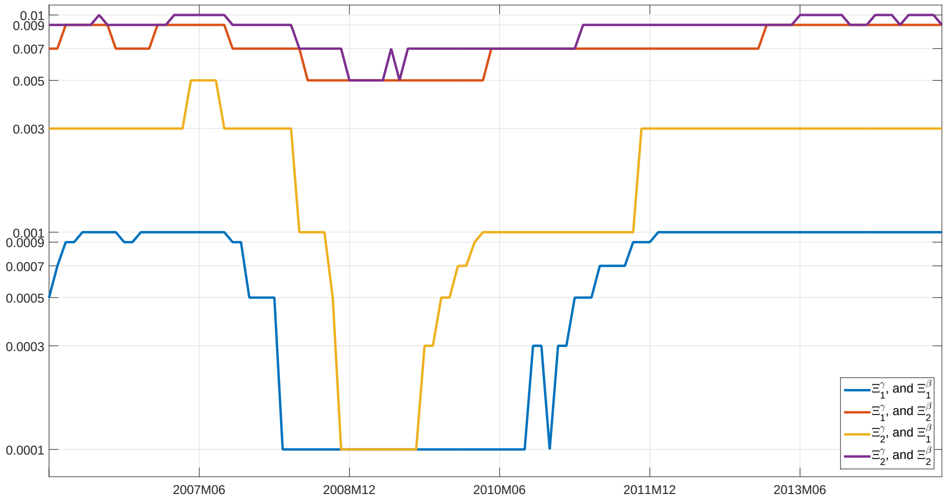

Figure 4: Figure 4: Optimal values for σ^2 selected by DMS for different choices for Ξ

the methods of Section 2.3 which do not impose any factor structure on error covariances. The second, which we label TVP-PVAR-X, uses the methods of Section 2.4 and does impose the factor structure on the error covariances.

3.3.2. Forecasting Results

Table 1 presents Mean Squared Forecast Errors (MSFEs) evaluated over the period 2006M1-2014M12 for the various approaches relative to univariate random walk (RW) forecasts for various forecast horizons for the 19 countries in the eurozone. Using MSFEs as a metric, every one of our approaches beats random walk forecasts for almost every country and forecast horizon. The approach with the best forecast performance is highlighted in bold. With 19 countries and 4 forecast horizons, we have 76 forecasts to compare.

It can be seen that different approaches tend to forecast best in different cases. But overall, our TVP-PVAR-DMA approach is forecasting best. It has the lowest MSFE in 29 of the 76 cases. And even in cases where its MSFE is not the lowest, its MSFEs are never more than a few percent higher than the best. The method that is perhaps the second best in our application is 19-TVP-VARs and involves just using individual country VARs. This often forecasts well, but occasionally forecasts very poorly (see, e.g., results for Ireland). We are finding TVP-PVAR-DMA to be a robust approach which often forecasts best but, if not, it never goes too far wrong. It is worth noting that TVP-PVAR-DMA forecasts particularly well at longer ($h = 12$) horizons.

Factor methods, as embodied with our DFM approach, also tend to forecast well, but are typically beaten by TVP-PVAR-DMA. However, the LTVP-VAR, which ignores the panel structure of the data and simply throws all the variables together into one large TVP-VAR, does not forecast well. This reinforces one of the original motivations for this paper: that taking the panel structure of the data into account when designing a hierarchical prior can be important.

Our TVP-PVAR approaches which do not use DMA methods also tend to forecast well (especially TVP-PVAR-X which imposes a factor structure on the error covariance matrix). However, they are almost uniformly beaten by the TVP-PVAR-DMA approach which averages over different choices for Ξ and different VAR dimensions.

Table 2 presents averages (over time) of the log predictive likelihoods.[6] Unlike MSFEs, which are based on point forecasts, predictive likelihoods evaluate the entire predictive distribution. But, in our case, predictive likelihoods and MSFEs are telling roughly the same story. Probably this is due to the fact that all of our approaches use EWMA methods to estimate a time-varying error covariance matrix. Allowing for time-varying volatility is usually important in getting a reasonable estimate of the predictive variance. Homoskedastic models can often provide good point forecasts (and, thus, good MSFEs), but often fail to provide good predictive likelihoods. Nevertheless it is worth noting that, when using predictive likelihoods, the DFM is occasionally forecasting quite poorly (see, e.g., results for Latvia and Greece). But, overall, TVP-PVAR-DMA is still producing forecasts which are often the best and, when not, are not too far from being best. And this finding is occurring despite the fact that working with $G = 7$ variables for all countries is probably excessive. But, a key benefit of TVP-PVAR-DMA is that it is finding out this fact in a data-based fashion and deciding to put more weight on forecasts from more parsimonious models.

Tables 1 and 2 present evidence on average forecast performance over the entire forecast evaluation period and showed that our TVP-PVAR-DMA tends to forecast well. Figure 5 presents evidence on when it does so. It presents cumulative sums of log predictive likelihoods where the sums are taken both over time and across countries for different forecast horizons. These measure the overall forecast performance across all countries. Note first that the line corresponding to the TVP-PVAR-DMA almost always lies above the lines for all of the other approaches. This reinforces the story of Tables 1 and 2 that it is the best overall forecast method for this data set. However, these benefits mostly occur after 2009. Thus, we have a story where, before the eurozone crisis, any sensible forecasting method can work well. But after 2009, the benefits of our approach become clear. DMA and DMS methods, which allow for model switching, work well in unstable times.

[6]We present averages to give the reader an idea of the difference in log predictive likelihoods for a typical observation. Cumulative sums can be obtained by multiplying by the number of observations in the forecast evaluation period (i.e. $504 - h$). Log Bayes factors comparing two approaches are approximately the difference in the cumulative sums of their log predictive likelihods.

Table 1: MSFEs relative to random walk

	AT	BE	CY	EE	FI	FR	DE	GR	IE	IT	LV	LT	LU	MT	NL	PT	SK	SI	ES
									Horizon, $h=1$										
RW	0.070	0.111	0.191	0.166	0.137	0.041	0.107	0.225	0.085	0.047	0.201	0.185	0.186	0.292	0.107	0.143	0.079	0.216	0.118
19-TVP-VAR	0.61	0.65	0.84	0.95	0.42	0.73	0.46	0.60	0.88	0.72	0.92	0.83	0.64	0.55	0.54	0.61	1.15	0.73	0.75
LTVP-VAR	0.79	0.65	1.19	1.14	0.69	0.73	0.54	0.74	0.85	4.34	1.64	1.23	0.86	0.63	0.58	0.70	3.81	0.82	0.69
DFM	**0.56**	**0.56**	**0.63**	**0.63**	**0.38**	0.68	**0.41**	0.74	**0.62**	**0.55**	**0.70**	**0.61**	0.67	0.48	**0.53**	**0.44**	1.32	0.66	**0.54**
TVP-PVAR	0.68	0.64	0.83	0.97	0.49	0.74	0.44	**0.65**	0.70	0.75	1.13	0.89	0.63	0.54	0.59	0.62	**0.76**	0.66	0.70
TVP-PVAR-X	0.58	0.59	0.70	0.89	0.45	**0.67**	0.46	0.67	0.80	0.64	1.36	1.04	**0.58**	0.49	**0.53**	0.61	0.83	**0.62**	0.71
TVP-PVAR-DMA	**0.56**	0.57	0.71	0.77	0.45	0.72	0.43	0.67	0.64	0.79	0.87	0.85	0.68	**0.45**	0.54	0.61	0.77	0.74	0.73
									Horizon, $h=3$										
RW	0.069	0.111	0.309	0.199	0.122	0.045	0.103	0.278	0.083	0.059	0.217	0.207	0.218	0.323	0.116	0.123	0.098	0.269	0.149
19-TVP-VAR	0.63	0.67	0.56	0.98	0.49	**0.68**	**0.49**	0.51	1.17	0.64	1.10	0.92	0.56	0.48	0.51	0.75	1.10	0.61	0.63
LTVP-VAR	0.70	0.75	0.78	1.00	0.69	0.75	0.56	0.62	0.84	3.48	1.72	1.26	0.62	0.55	0.51	0.78	3.06	0.64	**0.57**
DFM	**0.58**	0.69	**0.52**	0.88	**0.46**	0.74	0.52	0.43	0.79	**0.58**	**0.88**	**0.74**	0.53	0.48	0.46	**0.68**	1.14	0.57	0.63
TVP-PVAR	0.69	0.70	0.60	0.95	0.57	0.77	0.50	0.48	0.81	0.63	1.29	0.88	0.61	0.51	0.50	0.72	**0.71**	**0.55**	0.64
TVP-PVAR-X	0.61	**0.65**	0.55	0.86	0.51	**0.68**	**0.49**	0.60	0.94	0.60	1.38	0.95	**0.52**	0.47	0.49	0.72	0.73	0.56	0.63
TVP-PVAR-DMA	0.62	**0.65**	0.54	**0.81**	0.50	**0.68**	**0.49**	**0.41**	**0.75**	**0.58**	1.02	**0.74**	0.53	**0.39**	**0.44**	0.69	0.74	0.57	0.61
									Horizon, $h=6$										
RW	0.077	0.133	0.317	0.238	0.116	0.055	0.093	0.169	0.087	0.047	0.303	0.252	0.245	0.330	0.106	0.130	0.120	0.265	0.150
19-TVP-VAR	**0.56**	**0.56**	0.56	0.86	**0.54**	**0.57**	**0.54**	0.92	1.40	0.83	1.07	0.95	0.51	**0.45**	0.56	0.74	0.94	0.65	0.67
LTVP-VAR	0.62	0.60	0.77	0.79	0.70	0.61	0.68	1.03	**0.81**	4.16	1.28	1.10	0.52	0.52	0.61	0.76	2.57	0.70	**0.61**
DFM	0.59	0.61	**0.63**	**0.71**	0.57	0.63	0.56	0.83	0.90	0.84	**0.72**	**0.70**	0.54	0.54	0.58	0.68	0.98	0.65	0.64
TVP-PVAR	0.60	0.62	0.59	0.87	0.60	0.64	0.55	**0.80**	0.86	0.82	1.08	0.82	0.51	0.49	0.56	**0.66**	**0.57**	0.56	0.67
TVP-PVAR-X	**0.56**	0.57	0.56	0.75	0.56	0.59	0.55	1.02	0.95	0.79	1.00	0.82	**0.50**	0.46	0.55	0.72	0.67	0.58	0.66
TVP-PVAR-DMA	**0.56**	**0.56**	0.57	0.75	**0.54**	0.59	0.55	0.88	0.87	**0.74**	0.79	**0.70**	0.51	0.47	**0.54**	0.69	0.61	**0.56**	0.63
									Horizon, $h=12$										
RW	0.097	0.195	0.363	0.396	0.122	0.076	0.109	0.255	0.132	0.078	0.521	0.325	0.286	0.375	0.112	0.169	0.130	0.291	0.204
19-TVP-VAR	**0.45**	**0.41**	0.51	0.57	**0.52**	**0.42**	**0.48**	0.64	1.34	0.53	0.90	0.89	**0.43**	**0.40**	0.55	0.61	0.92	0.57	0.53
LTVP-VAR	0.52	0.49	0.67	**0.53**	0.74	0.44	0.63	0.73	0.59	2.65	0.81	0.83	0.58	0.44	0.56	0.61	2.23	**0.49**	0.49
DFM	0.61	0.54	0.51	0.56	0.62	0.59	0.53	0.62	0.61	0.54	0.66	**0.61**	0.54	0.47	0.59	0.58	0.98	0.58	**0.45**
TVP-PVAR	0.48	0.45	0.54	0.58	0.59	0.48	0.53	0.61	0.62	0.53	0.79	0.74	0.44	0.44	0.59	0.58	0.57	0.50	0.50
TVP-PVAR-X	**0.45**	0.42	0.52	0.46	0.57	0.43	0.49	0.73	0.67	0.51	**0.62**	0.66	**0.43**	**0.40**	0.54	0.60	0.65	0.51	0.51
TVP-PVAR-DMA	**0.45**	**0.41**	**0.50**	0.54	0.57	0.47	0.49	**0.59**	**0.54**	**0.49**	0.68	0.64	**0.43**	0.41	**0.53**	**0.56**	**0.54**	**0.49**	0.48

Table 2: Average log-predictive likelihoods

Model	AT	BE	CY	EE	FI	FR	DE	GR	IE	IT	LV	LT	LU	MT	NL	PT	SK	SI	ES
									Horizon, $h = 1$										
19-TVP-VAR	0.15	-0.10	-0.52	-0.46	**-0.02**	0.32	0.07	**-0.40**	-0.11	0.26	**-0.59**	-0.47	-0.36	-0.50	**-0.05**	**-0.22**	-0.22	-0.50	-0.23
LTVP-VAR	0.05	-0.09	-0.68	-0.55	-0.18	0.33	-0.02	-0.44	-0.04	-0.62	-0.79	-0.66	-0.48	-0.56	-0.11	-0.28	-0.81	-0.56	-0.18
DFM	**0.17**	**-0.01**	**-0.37**	**-0.36**	**-0.02**	**0.40**	0.09	-0.95	-0.10	0.19	-0.84	**-0.27**	-0.38	-1.24	-0.16	-0.34	-0.38	-0.60	**-0.04**
TVP-PVAR	-0.36	-0.42	-0.70	-0.69	-0.41	-0.20	-0.28	-0.62	-0.39	-0.21	-0.85	-0.73	-0.67	-0.67	-0.39	-0.48	-0.42	-0.67	-0.50
TVP-PVAR-X	0.16	-0.02	-0.45	-0.40	**-0.02**	**0.40**	0.10	**-0.40**	**-0.03**	**0.33**	-0.67	-0.48	**-0.30**	-0.46	-0.15	-0.24	**-0.10**	-0.44	-0.18
TVP-PVAR-DMA	0.16	-0.04	-0.53	**-0.36**	-0.03	0.37	**0.12**	-0.46	**-0.03**	0.25	-0.61	-0.44	**-0.30**	**-0.42**	-0.11	-0.24	**-0.10**	**-0.34**	-0.15
									Horizon, $h = 3$										
19-TVP-VAR	0.11	-0.13	**-0.58**	-0.58	**-0.06**	0.31	0.06	**-0.41**	-0.28	0.17	-0.76	-0.60	-0.39	-0.48	-0.07	-0.27	-0.30	-0.52	-0.29
LTVP-VAR	0.05	-0.14	-0.76	-0.58	-0.29	0.28	-0.04	-0.50	**-0.05**	-0.61	-0.85	-0.71	-0.44	-0.55	**-0.01**	-0.28	-0.82	-0.57	-0.23
DFM	**0.20**	**-0.03**	-0.65	-0.64	-0.16	0.25	0.07	-0.53	**-0.05**	0.14	**-0.67**	-0.54	-0.46	-0.83	-0.20	**-0.13**	-0.43	-0.28	**-0.13**
TVP-PVAR	-0.27	-0.12	-0.78	-0.77	-0.49	0.20	-0.12	-0.56	-0.20	-0.12	-0.97	-0.85	-0.65	-0.73	-0.30	-0.29	-0.44	-0.65	-0.52
TVP-PVAR-X	0.10	-0.12	-0.65	-0.50	-0.14	**0.33**	**0.08**	-0.47	-0.13	0.18	-0.76	-0.51	-0.36	**-0.46**	-0.12	-0.22	-0.16	-0.28	-0.32
TVP-PVAR-DMA	0.11	-0.11	-0.62	**-0.49**	-0.13	**0.33**	**0.08**	-0.43	-0.13	**0.20**	-0.72	**-0.50**	**-0.38**	-0.50	-0.08	-0.20	**-0.13**	**-0.22**	-0.21
									Horizon, $h = 6$										
19-TVP-VAR	**0.13**	**-0.15**	-0.60	-0.64	**-0.12**	0.28	0.04	-0.49	-0.46	0.15	-1.00	-0.78	-0.44	**-0.47**	-0.06	-0.29	-0.35	-0.55	-0.41
LTVP-VAR	0.06	**-0.15**	-0.77	-0.58	-0.29	0.28	-0.07	-0.54	**-0.07**	-0.60	-0.88	-0.78	-0.43	-0.50	**0.04**	-0.28	-0.85	-0.59	**-0.26**
DFM	0.06	-0.18	**-0.49**	-0.62	-0.17	0.16	-0.12	-1.06	-0.20	0.04	-1.54	**-0.51**	**-0.42**	-1.21	-0.18	-0.67	-0.48	-0.56	-0.40
TVP-PVAR	-0.26	-0.15	-0.78	-0.84	-0.51	0.20	-0.31	-0.60	-0.47	-0.21	-1.02	-0.90	-0.63	-0.65	-0.41	-0.50	-0.47	-0.63	-0.53
TVP-PVAR-X	0.10	-0.17	-0.72	**-0.54**	-0.23	0.28	**0.05**	-0.51	-0.22	0.14	-0.81	-0.58	-0.46	**-0.47**	-0.10	-0.29	-0.21	**-0.50**	-0.39
TVP-PVAR-DMA	0.09	-0.16	-0.67	-0.56	-0.20	**0.30**	**0.05**	**-0.46**	-0.21	**0.16**	**-0.79**	-0.59	-0.48	-0.48	-0.10	**-0.20**	**-0.19**	-0.55	-0.36
									Horizon, $h = 12$										
19-TVP-VAR	**0.14**	**-0.23**	-0.63	-0.68	**-0.16**	0.25	**-0.02**	-0.59	-0.83	0.07	-1.38	-0.92	**-0.46**	-0.50	**-0.07**	-0.35	-0.35	-0.53	-0.57
LTVP-VAR	0.07	-0.27	-0.64	-0.62	-0.40	**0.27**	-0.13	-0.60	**-0.11**	-0.65	-1.02	-0.87	**-0.46**	-0.56	-0.09	**-0.27**	-0.82	**-0.47**	**-0.33**
DFM	-0.22	-0.38	**-0.60**	-1.18	-0.47	-0.02	-0.39	-1.35	-0.46	-0.17	-2.59	-1.05	-0.62	-1.29	-0.17	-0.81	-0.53	-0.66	-0.61
TVP-PVAR	-0.29	-0.35	-0.79	-0.85	-0.48	0.25	-0.37	-0.66	-0.46	-0.26	-1.21	-1.02	-0.70	-0.78	-0.48	-0.57	-0.51	-0.68	-0.61
TVP-PVAR-X	0.03	-0.28	-0.72	**-0.56**	-0.28	0.24	-0.05	-0.61	-0.37	0.07	**-0.94**	**-0.65**	**-0.46**	-0.49	-0.13	-0.32	-0.20	**-0.47**	-0.52
TVP-PVAR-DMA	0.04	-0.27	-0.67	-0.62	-0.28	0.23	-0.03	**-0.54**	-0.30	**0.11**	-1.05	-0.72	**-0.46**	**-0.48**	-0.12	-0.29	**-0.13**	-0.51	-0.57

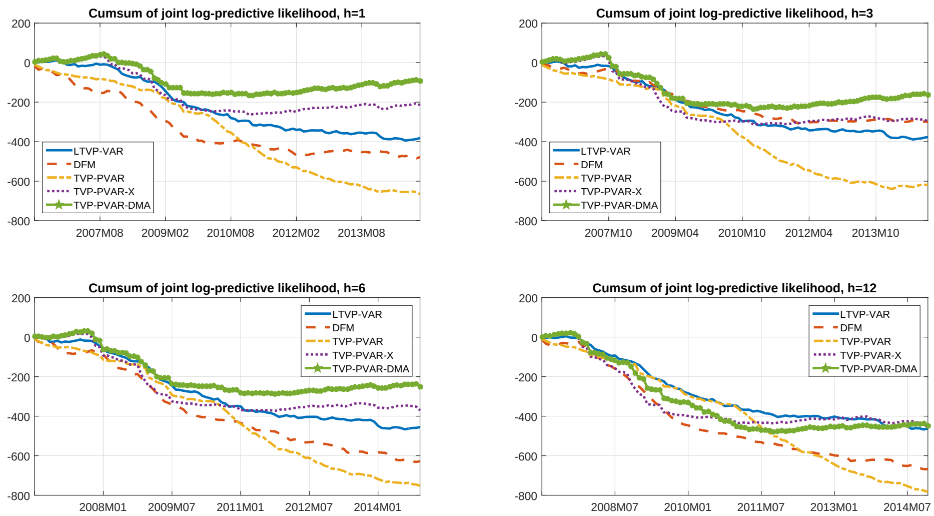

Figure 5: Figure 5: Cumulative sums of log predictive likelihoods (sums taken over time and across countries)

4. Conclusions

In this paper, we have developed Bayesian methods for estimating large TVP-PVARs and shown them to forecast well in an application involving euro area inflation rates. This development involved the design of plausible hierarchical priors for working with multi-country data which ensure parsimony without mis-specification and the design of computationally feasible forecasting methods using these priors. The latter we achieve using approximate forgetting factor and EWMA methods. For the former, we consider some alternative approaches, but our main innovation comes in the use of DMA methods. In an application such as ours, where there is uncertainty over what the appropriate structure of the hierarchical prior should be and what the appropriate VAR dimension should be, we have shown the benefits of use of DMA methods. These allow us to begin with a parameter rich model with a range of prior choices and automatically attach more weight to the best-forecasting, often more parsimonious, choices. Our forecasting exercise shows the benefits of our approach.

[1] Banbura, M., Giannone, D., and Reichlin, L. (2010), "Large Bayesian Vector Autoregressions," *Journal of Applied Econometrics,* 25, 71-92.

[2] Canova, F. (1993), "Modelling and Forecasting Exchange Rates using a Bayesian Time Varying Coefficient model," *Journal of Economic Dynamics and Control,* 17, 233-262.

[3] Canova, F., and Ciccarelli, M. (2009), "Estimating Multicountry VAR Models," *International Economic Review,* 50, 929-959.

[4] Canova, F., and Ciccarelli, M. (2013), "Panel Vector Autoregressive models: A Survey," European Central Bank Working Paper 1507.

[5] Carriero, A., Kapetanios, G., and Marcellino, M. (2010), "Forecasting Exchange Rates with a Large Bayesian VAR," *International Journal of Forecasting,* 25, 400-417.

[6] Carriero, A., Kapetanios, G. and Marcellino, M. (2012), "Forecasting Government Bond Yields with Large Bayesian VARs," *Journal of Banking and Finance*, 36, 2026-2047.

[7] Chib, S. and Greenberg, E. (1995), "Hierarchical Analysis of SUR models with Extensions to Correlated Serial Errors and Time-varying Parameter Models," *Journal of Econometrics*, 68, 339-360.

[8] Cogley, T. and Sargent, T. (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the post WWII U.S.," *Review of Economic Dynamics*, 8, 262-302.

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

[10] Delle Monache, D., Petrella, I. and Venditti, F. (2015), "Common Faith or Parting of Ways? A Time-varying Parameter Factor Analysis of Euro Area Inflation," *Advances in Econometrics*, forthcoming.

[11] Doan, T., Litterman, R., and Sims, C. (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews,* 3, 1-100.

[12] Durbin, J., and Koopman, S. (2001), *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

[13] Faust, J. and Wright, J. (2013), "Forecasting Inflation," chapter 1 in Handbook of Economic Forecasting, volume 2A, G. Elliot and A. Timmermann, eds., Amsterdam: North Holland.

[14] Feldkircher, M. and Huber, F. (2015), "The International Transmission of US shocks – Evidence from Bayesian Global Vector Autoregressions," *European Economic Review*, forthcoming.

[15] Giannone, D., Lenza, M. and Primiceri, G. (2015) "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, 97, 436-45.

[16] Koop, G., (2013), "Forecasting with Medium and Large Bayesian VARs," *Journal of Applied Econometrics,* 28, 177-203.

[17] Koop, G., (2014), "Forecasting with Dimension Switching VARs," *International Journal of Forecasting*, 30, 280-290.

[18] Koop, G. and Korobilis, D., (2013), "Large Time-varying Parameter VARs," *Journal of Econometrics,* 177, 185-198.

[19] Korobilis, D. (2013), "VAR Forecasting Using Bayesian Variable Selection," *Journal of Applied Econometrics,* 28, 204-230.

[20] Kulhavý, R., and Kraus, F. (1996), "On Duality of Regularized Exponential and Linear Forgetting," *Automatica*, 32, 1403-1415.

[21] Primiceri, G. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821-852.

[22] Raftery, A., Karny, M., and Ettler, P. (2010), "Online Prediction Under Model Uncertainty Via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, 52, 52-66.