Forecasting with High-Dimensional Panel VARs*

Gary Koop
University of Strathclyde

Dimitris Korobilis†
University of Essex

November 25, 2018

Abstract

This paper develops methods for estimating and forecasting in Bayesian panel vector autoregressions of large dimensions with time-varying parameters and stochastic volatility. We exploit a hierarchical prior that takes into account possible pooling restrictions involving both VAR coefficients and the error covariance matrix, and propose a Bayesian dynamic learning procedure that controls for various sources of model uncertainty. We tackle computational concerns by means of a simulation-free algorithm that relies on analytical approximations to the posterior. We use our methods to forecast inflation rates in the eurozone and show that these forecasts are superior to alternative methods for large vector autoregressions.

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

*We want to thank Ana Galvao, Tony Garratt, George Kapetanios, James Mitchell, Ivan Petrella, Rob Taylor and participants at the 9th European Central Bank Workshop on Forecasting Techniques; the Bank of England workshop on Time-Variation in Econometrics and Macroeconomics; and the 10th RCEA Bayesian Econometric Workshop, for helpful comments and discussions. All remaining errors are ours.

†Corresponding author: Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. email: d.korobilis@essex.ac.uk

1 Introduction

As the economies of the world become increasingly linked through trade and financial flows, the need for multi-country econometric modelling has increased. Panel Vector Autoregressions (PVARs), which jointly model many macroeconomic variables in many countries, are becoming a popular way of fulfilling this need. We use the general term PVAR for models where the dependent variables for all countries are modelled jointly in a single VAR and, thus, the VAR for each individual country is augmented with lagged dependent variables from other countries. In this paper, we develop econometric methods for PVARs of possibly large dimensions using a hierarchical prior which can help overcome the concerns about over-parameterization that arise in these models. The novelties of our approach are that i) we tackle in one integrated setting all concerns pertaining to out-of-sample forecasting, such as controlling for over-fitting and dealing with model uncertainty; ii) we allow for empirically relevant extensions that account for structural breaks and changing volatilities; and iii) we do so in a computationally efficient way, building on previous work by Koop and Korobilis (2013) for single-country VARs.

To explain the significance of the econometric contributions of this paper, we note a major characteristic of the existing literature on multi-country VAR models lies in the need to appropriately model linkages between countries. This literature includes Bayesian multi-country VARs (Canova and Ciccarelli, 2009; Koop and Korobilis, 2016), Global VARs (Dees, di Mauro, Pesaran and Smith, 2007; Feldkircher and Huber, 2016), multi-country factor models (Kose, Otrok and Whiteman, 2003), and spatial VARs (Chudik and Pesaran, 2011). The common ground of all these modelling approaches is the need to account for the panel structure in the data, and explicitly model inter-dependencies and commonalities in the units (countries or individuals). This is an important consideration as the dimension of panel VARs tends to grow rapidly: the case with only five variables for 10 countries results in a model with 50 endogenous variables and thousands of parameters. Therefore,

any panel-specific restrictions one can impose, such as clustering/pooling coefficients across units, could help identify parsimonious models that are useful in forecasting or structural inference. Alternative solutions to the overparametrization problem are offered in the literature on large (single-country) VARs – such as those estimated by Bańbura, Giannone and Reichlin (2010) and Koop and Korobilis (2013), that incorporate shrinkage estimators and efficient computational algorithms. However, large VAR methods typically rely on shrinkage priors that ignore the panel structure in the data and this loss of useful information might have adverse effects in forecasting.[1]

Consequently, our motivation and first econometric contribution is to develop efficient methods for VARs of large dimensions that feature panel-specific restrictions as well as time-varying parameters and stochastic volatility, and fill this particular gap in the VAR literature. Our starting point is the seminal contribution of Canova and Ciccarelli (2009) who introduce a hierarchical shrinkage prior for multi-country VARs with time-varying parameters and develop Markov chain Monte Carlo (MCMC) simulation methods to tackle estimation.[2] We extend their methods to account for stochastic volatility in the panel VAR error covariance matrix. We also propose a model formulation that allows us to introduce their hierarchical shrinkage prior to the time-varying error covariance matrix. Both these extensions are empirically relevant. There is ample evidence that volatility in empirical macroeconomic models is extremely important for forecasting (see, among many others, Clark and Ravazzolo, 2015 and Diebold, Schorfheide and Shin, 2017), in

[1] In particular, popular applications of large VARs rely on the Minnesota prior (Doan, Litterman and Sims, 1984) that places weak prior shrinkage on the intercepts and own autoregressive dynamics of each variable, but heavily shrinks cross-terms and more distant lags. In the case of panel VARs right-hand side lags include own lags of each variable of a given country, but also i) lags of the same variable of other countries; ii) lags of other variables of the same country; and iii) lags of other variables of other countries. The Minnesota prior would simply place equal prior weight on these three categories of right-hand side variables, which would result in discarding useful information about interdependencies and homogeneities among countries. Similar arguments would hold for any shrinkage prior or penalized estimator that is not developed specifically for panel VARs, such as the popular lasso of Tibshirani (1996).

[2] As we explain later in this paper, this hierarchical prior pools certain VAR coefficients by country or by variable, and results in a lower-dimensional VAR which greatly reduces the computational burden. However, the use of simulation methods means that the authors consider only structural analysis (not much more computationally-intensive recursive forecasting) using 27-variable PVARs.

which case it is imperative to relax the assumption of homoskedasticity used by Canova and Ciccarelli (2009). However, given that time-varying covariance matrices are non-parsimonious, some form of shrinkage is also needed for these parameters. Off-diagonal elements of the error covariance matrix of a panel VAR represent static/contemporaneous interdependencies among countries. This motivates our use the same pooling prior on these error covariance parameters that Canova and Ciccarelli (2009) only use for the VAR coefficients. Finally, we adapt the scalable state-space estimator of Koop and Korobilis (2013) to this time-varying parameter panel VAR setting. This estimator is simulation-free and, in combination with the pooling-shrinkage prior of Canova and Ciccarelli (2009), it allows estimation and forecasting with models with possibly hundreds of endogenous variables.

A second econometric contribution lies in the treatment of model uncertainty. We consider various sources of uncertainty that pertain to panel VARs with time-varying parameters, and we use a Bayesian dynamic learning prior that allows to learn interesting model features from the data. Given our interest in forecasting, we use posterior predictive likelihoods (an out-of-sample measure) to inform a number of decisions about optimal model structure. The Bayesian learning procedure we use is dynamic, meaning that at each point in time a different model structure might hold. For example, one consideration is whether to shrink using a pooling prior, or whether it is better for forecasting to reduce the model to a more parsimonious country-specific structure where within-country effects are the only drivers of fluctuations in the macroeconomic variables and between-country spillover effects are zero. Our procedure allows for the choice between these two structures. A second consideration is how much time-variation in parameters is optimal for forecasting. As mentioned previously, there is a great deal of evidence that stochastic volatility is extremely important for forecasting, but there is mixed evidence about whether allowing for time-varying intercepts and autoregressive coefficients is important. Our full model specification nests models with faster or slower drifts in coefficients, as well as a fully

time-invariant panel VAR structure. Thus we can estimate (in a time-varying manner) the amount of time variation in the error covariance and VAR coefficients. Finally, we take into account uncertainty about the size of the panel VAR. When forecasting, it might be the case that VARs with a small number of variables (or even VARs that include only the variable of interest and no additional predictors, such as Stock and Watson, 2005, do for international business cycle dynamics) might perform better. Existing evidence for forecasting inflation in the US suggests that this might be the case; see the popular unobserved components model of Stock and Watson (2007) that features no predictors but a time-varying trend and stochastic volatility. We introduce, therefore, a very flexible learning mechanism that looks at the most recent out-of-sample performance at each point in time and then shrinks the panel VAR model to an optimal parsimonious structure. We show that this novel modelling feature acts as a safeguard against overfitting and poor forecasting performance.

Our paper also seeks to contribute to the empirical literature on inflation modelling 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 may 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, most of the eurozone were experiencing deflation and no country registered an inflation rate above 1%. Subsequently euro area inflation has increased, but as of 2018 it still remains low by historic standards. 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 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 case of time-varying parameters and stochastic volatility, 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 dynamic learning methods which average over different PVAR dimensions and different hierarchical priors. Section 4 concludes.

2 Econometric Methodology

In this section we detail the underlying econometric methodology for estimation in large panel VARs. We begin by discussing some of the issues which occur with PVAR models and introduce the hierarchical prior we adopt, before discussing time-varying parameter versions of PVARs. We then propose modifications of the time-varying parameter PVAR model in order to allow hierarchical modelling of the error covariance matrix. We conclude this section with a discussion of our treatment of model uncertainty and provide a detailed explanation of how the dynamic learning procedure works.

2.1 Methods of Ensuring Parsimony in the PVAR

Assume N countries and G variables for each country which are observed for T time periods. In our empirical application these are 10 eurozone countries and we use inflation plus an additional eight country-specific predictor variables, observed for 216 months. 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, $i = 1, ..., N$.[3] 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 $E\left(u_{it}u'_{jt}\right) = \Sigma_{ij}$ 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 $\alpha = \left(vec\left(A^1\right)', ..., vec\left(A_t^p\right)'\right)'$. 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.

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. A popular approach to dimension reduction in Bayesian vector autoregressions is to use hierarchical priors that induce shrinkage in the parameters. Examples of such priors include the scale mixture of Normals priors used in George, Sun and Ni (2008) and the priors from general equilibrium models of Del Negro and Schorfheide (2004). However, adopting any of the existing priors developed for a single-country VAR will result in discarding valuable information due to the fact that one would expect partial pooling of coefficients in a panel setting. Incorporating such information could improve forecasts. Furthermore, these methods require the use of MCMC methods which make them computationally infeasible

[3] In our empirical work, we also include the oil price in the vector of dependent variables and its dimensionality is thus $NG + 1$.

with high-dimensional models, particularly in the context of a forecasting exercise.

A popular way to introduce such pooling in PVARs that will result in efficient shrinkage, is described in Canova and Ciccarelli (2009, 2013). These authors use a certain hierarchical prior so as to work with restricted versions of (1), a practice which we adopt in this paper. In particular, we assume the following reduced rank representation of the PVAR coefficients:

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

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 $V = \sigma^2 I$. Due to the fact that the high-dimensional vector of coefficients α is projected into a lower dimensional vector of latent parameters θ, we will refer to this second layer regression as a "factor model" for the PVAR coefficients (this will become clearer later when α, θ become time-varying). This specification can be thought of as a hierarchical prior for the PVAR model of the form $\alpha|\Sigma \sim N(\Xi\theta, \Sigma \otimes V)$ and $\theta \sim N(0, Q)$, which is of the natural conjugate form for α due to the conditioning on Σ.

How can we use this specification to extract meaningful lower-dimensional factors of α that take into account the panel structure in the data? 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, 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 should be augmented using different choices of Ξ and an algorithm developed to choose between them. This is what we do in this paper. In theory, one could devise a huge range of possible structures for Ξ, e.g. allowing own lag coefficients to be unrestricted; impose "core" and "periphery" clusters on the coefficients; global VAR restrictions, and so on. In practice, we have found that for our euro area data two specific structures clearly beat a range of several other options in terms of in-sample and out-of-sample fit. The first is one begins with the same pooling structure proposed in Canova and Ciccarelli (2009) but leaves intercepts and first autoregressive

lagged coefficients to be unrestricted – meaning that these sets of coefficients are not determined through the common factors. This can simply be achieved if we extract one factor for each coefficient we want to leave unrestricted and set the state variance of that parameter to zero. If, as an example, the j-th element of α is an intercept, we extract a new factor θ_i with associated "loadings vector" Ξ_i that has it's j-th scalar element equal to one, and all remaining $K-1$ elements equal to zero. That way, we add $2 \times N \times G$ factors to the ones that already Canova and Ciccarelli suggest.[4]

The second way of specifying the factor structure is by restricting Ξ in such a way that the PVAR collapses to a country-specific VAR 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 time-varying parameter PVAR (TVP-PVAR) in the next section, this definition of Ξ will imply the same intuition, except in terms of individual-country TVP-VARs.

[4]It is not surprising that we find important for forecasting *not* to pool the coefficients on first own lags with other VAR coefficients through a common factor structure. Intercepts and first own lags are always important components of good time-series forecasts, and the Minnesota prior is also based on the same principle and shares a similar structure for intercepts and AR(1) coefficients.

2.2 Moving from the PVAR to the TVP-PVAR

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 time-varying parameter 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. At each point in time t the number of PVAR parameters, $p \times (N \times G)^2$, could run into the thousands or more. The fact that in the case of time-variation we have T such high-dimensional parameter vectors only complicate computations. Adding to the mix the fact that typical estimation would rely on MCMC methods (e.g. Cogley and Sargent, 2005, use MCMC), means that we have to repeat thousands of times any burdensome algorithmic operations involving these high-dimensional parameter vectors. 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. As discussed in Koop or Korobilis (2013), estimating TVP-VARs using MCMC methods can easily become computationally infeasible unless the number of forecasting models and their dimension are both small.

In order to achieve parsimony, 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, $R \ll K$, Ξ is defined as in the preceding subsection and $w_t \sim N(0, W_t)$ where W_t is an $R \times R$ covariance matrix, and extending Canova and Ciccarelli (2009)'s homoskedastic specification we let $e_t \sim N(0, \Sigma_t \otimes V)$ where $V = \sigma^2 I$. The factors θ_t evolve according to a random walk, in order to be consistent with the bulk of the time-varying parameter VAR literature we cited above. 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). However, the generic MCMC sampler of Chib and Greenberg (1995), when applied to the hierarchical prior above, proves to be computationally inefficient. This is because their algorithm requires many draws from the Normal conditional posterior of α_t $\forall t$, which proves to be extremely demanding. In this case, a certain conjugacy assumption of Canova and Ciccarelli (2009) can help bypass explicit inference for the high-dimensional α_t, resulting in huge computational gains. To be specific, the assumption that the prior for α_t in equation (3) is of natural conjugate form (conditional on Σ_t) comes in handy.

As a consequence, we can simplify the TVP-PVAR given by (2), (3) and (4) into the following simpler, two-equation form (see Canova and Ciccarelli, 2013, page 22, for a proof):

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

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

where $\widetilde{X}_t = X_t\Xi$ and $v_t = X_t'e_t + u_t$ with $v_t \sim N(0, (I + \sigma^2 X_t'X_t) \times \Sigma_t)$. Therefore, in this form the TVP-PVAR is written as a Normal linear 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 and the reader is referred to the online technical Appendix of this paper, as well as Koop and Korobilis (2013) for further computational details. 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 when we take model uncertainty into account (see the next subsection). The following paragraphs elaborate on these points.

First, in any state-space problem estimation of the state variance (W_t in our case) can be quite demanding. In a VAR context, Primiceri (2005) discusses how the prior plays a key role in determining this parameter when the information in the likelihood is weak. Here we use instead a method introduced in engineering many decades ago involving the use of forgetting factors. 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 these methods. The main idea behind this method is that of exponentially discounting past state variances in order to obtain an estimate of the current period's variance. This means that W_t can be estimated using the following formula

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

where $\mathcal{D}_{t-1}$ denotes data available through period $t-1$, $0 < \lambda \leq 1$ is the forgetting factor (typically set to a value lower than one but close to it) and $var\left(\theta_t | \mathcal{D}_{t-1}\right)$ is a predicted

variance readily available at time t from the Kalman filter iteration of the previous time period, $t-1$. Thus, at the cost of using a method which is approximate, we gain huge benefits in terms of computational simplicity and stability.[5] The forgetting factor approach allows estimation of systems with large numbers 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.

Improving on the inference of 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 $\widetilde{u}_t \widetilde{u}_t' = (I + \sigma^2 X_t' X_t)^{-1} \left[\left(Y_t - \widetilde{X}_t' E\left(\theta_t | \mathcal{D}_{t-1}\right) \right) \left(Y_t - \widetilde{X}_t' E\left(\theta_t | \mathcal{D}_{t-1}\right) \right)' \right]$, $E\left(\theta_t | \mathcal{D}_{t-1}\right)$ 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 $\widehat{\Sigma}_t = \frac{\sum_{\tau=1}^{t} \widetilde{u}_\tau \widetilde{u}_\tau'}{t}$ (i.e. equivalent to least squares methods in a homoskedastic model). 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 our empirical work, we let $\lambda = \{0.990, 0.992, 0.994, 0.996, 0.998, 1.000\}$ and $\kappa = \{0.92, 0.94, 0.96, 0.98, 1.00\}$, interpret each grid point as defining a model and use dynamic model selection methods (to be described below) 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

[5] Likelihood-based methods, such as Bayesian and maximum likelihood inference, rely on formulae involving $var\left(\theta_t | \mathcal{D}_t\right)$. However, at time t this quantity is uknown and needs to be estimated reliably, which will directly affect the way θ_t is updated and estimated.

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 for longer forecast horizons. To do this, we draw Y_{T+1} from its Normal 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 Normal predictive density conditional on the drawn Y_{T+1}, etc. up to h.

2.3 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

$$Y_t = X_t' \alpha_t + B_t^{-1} H_t \varepsilon_t,$$

where $\varepsilon_t \sim N(0, I)$. We can write the model in the following form as:

$$Y_t = X_t' \alpha_t + W_t' \beta_t + H_t \varepsilon_t, \tag{7}$$

where W_t is the matrix

$$W_t = \begin{bmatrix} 0 & \dots & \dots & 0 \\ \varepsilon_{1t} & 0 & \dots & 0 \\ 0 & [\varepsilon_{1t}, \varepsilon_{2t}] & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \left[\varepsilon_{1t}, ..., \varepsilon_{(NG-1)t}\right] \end{bmatrix}.$$

With this specification we have an equivalent model where the error covariances show up as contemporaneous regressors on the right hand side of the TVP-PVAR. This model cannot be estimated as a multivariate system using standard filtering methods described previously. To see this, note that elements Y_t show up both on the left-hand side, and the right-hand side of the PVAR (via the matrix of contemporaneous error terms, W_t). In this case, the state-space system is nonlinear and multivariate estimation would need to rely on computationally intensive simulation methods. A potential solution to this problem would be to follow Carriero, Clark and Marcellino (2016) and estimate the model equation-by-equation: the first equation does not contain any contemporaneous information on the right-hand side so can be estimated independently of other equations using a linear filter; the second equation is dependent on ε_{1t} which can be replaced by residuals from the first equation; the third equation is dependent on $\varepsilon_{1t}, \varepsilon_{2t}$ which can also be replaced by residuals, and so on until equation NG which also depends on residuals available from the previous $NG-1$ equations.[6] However, such an option is not available to us, since the pooling prior

[6] In a previous version of this manuscript we were instead working with the model

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

where $\gamma_t = \left(vec\left(B_t A_t^1\right)', ..., vec\left(B_t A_t^p\right)'\right)'$ and Z_t had the same structure as W_t but with elements $-Y_{jt}$ in place of $\widehat{\varepsilon}_{jt}$ – see also the definition of the respective matrix Z_t in the Appendix of Primiceri (2005). This previous formulation of the VAR and the one we currently use in equation (7) are observationally equivalent, however, the latter offers the advantage of having the original VAR coefficient α_t remaining the vector of VAR coefficients as in our original specification in (2). In contrast, the VAR form described above is in terms of coefficients γ_t which consist of products of VAR coefficients A_t and the VAR covariance B_t.

we adopt clusters coefficients among different equations. As a consequence the parameters of different PVAR equations will not be independent a-posteriori, and equation-by-equation estimation is not available. We overcome this issue in a fashion similar to the problem of estimating time-varying covariance matrices using the EWMA specification described in the previous subsection. That is, when constructing W_t we replace ε_t with the one-step ahead residuals from each Kalman filter iteration, namely $\widetilde{\varepsilon}_t = Y_t - X_t' E(\alpha_t|\mathcal{D}_{t-1}) - W_t' E(\beta_t|\mathcal{D}_{t-1})$. Doing so allows each time period t to have all right-hand-side variables observed and proceed with the estimation methods described in the previous section.

In order to complete this new but equivalent PVAR specification that treats elements of the error covariance matrix as exogenous predictors, we can extend our previous approach and introduce a hierarchical prior on both α_t and β_t of the form:

$$\begin{aligned} \delta_t &\equiv \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \Xi^{\alpha} & 0 \\ 0 & \Xi^{\beta} \end{bmatrix} \theta_t + u_t \equiv \Xi\theta_t + u_t, \\ \theta_t &= \theta_{t-1} + v_t. \end{aligned} \tag{8}$$

where now $u_t \sim N(0, H_t H_t \otimes (\sigma^2 I))$, 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 α_t and β_t load on separate factors (rows of θ_t), then we have prior independence between the two sets of coefficients. This exact prior independence of the sets of coefficients – an assumption that is used extensively in many Bayesian applications (see Primiceri, 2005) – is a crucial assumption that allows for equation-by-equation estimation of the PVAR using the linear Kalman filter. 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^{α} are those described at the end of Section 2.1. For Ξ^{β} we also use these two choices with the trivial adaptation required by the structure for W_t. In the

In this case, applying the pooling prior on γ_t instead of the original VAR coefficients α_t is a feature that is not desirable. The new, equivalent formulation solves this problem.

forecasting exercise we allow for Ξ^{α} and Ξ^{β} to potentially be different. Using P (for pooled) subscripts to denote the form that builds on Canova and Ciccarelli (2009), and CS (for country-specific) subscripts to denote the country-specific VAR factor structure, we define four possible specification for the Ξ matrices: i) Ξ_P^{α} and Ξ_P^{β}, ii) Ξ_P^{α} and Ξ_{CS}^{β}, iii) Ξ_{CS}^{α} and Ξ_P^{β} and iv) Ξ_{CS}^{α} and Ξ_{CS}^{β}. As we explain next in the following subsection, selection of the best specification pair for Ξ^{α} and Ξ^{β} is part of a more general dynamic procedure that selects the values of various hyperparameters that are optimal for forecasting.

It is worth digressing briefly to discuss the role of the ordering of variables in our model when we decompose Σ_t using a lower triangular B_t and diagonal H_t. We stress first that we are not using this decomposition to identify structural shocks, merely to estimate a reduced form forecasting model. Next we note that, conditional on Σ_t, the posterior of the VAR coefficients does not depend on the ordering of variables. That is, it will be exactly the same for any ordering of the variables. The only possible sensitivity to ordering relates to Σ_t itself and the prior we use. That is, we elicit priors separately on B_t and H_t and the resulting prior for Σ_t has the potential to differ slightly depending on the ordering of the variables. This issue is inherent to all the many papers which use this decomposition (e.g. Primiceri, 2005). But, we would argue, it is a small problem relative to the benefits of working with this way of decomposing the error covariance matrix. In particular, it allows for fast equation-by-equation computation and for straightforward development of restrictions on covariances through our Ξ^{α} matrices.

2.4 Dynamic Treatment of Model Uncertainty

The previous subsections discussed the estimation of single time-varying parameter PVARs and defined our model space. Our most general approach involves a model space where the models differ in various features, in order to explicitly control for model uncertainty. The features defining the models include the different choices for λ, κ and σ^2 described in

sub-section 2.2, the different combinations of Ξ^{α} and Ξ^{β} described in sub-section 2.3; and different PVAR dimensions (to be described in sub-section 3.1). In total we have 16,800 PVAR or TVP-VAR models to choose between. In this sub-section, we describe how to do so in a dynamic manner such that the chosen model may change over time. Our methods use posterior model probabilities constructed in a dynamic manner.

We can use such posterior model probabilities to either do model selection or averaging. In this paper, we adopt the view that some of our specification choices relate to concepts which can be interpreted as parameters (i.e. λ, κ and σ^2). For these we do model selection since this is similar to estimating them (e.g. if we select a model with $\lambda = 0.99$ this is the same as estimating λ to be 0.99). We also do model selection for the different choices for Ξ_{α} and Ξ_{β} since they define different priors and presenting results which average over very different priors would reduce the interpretability of results. With regards to choosing the dimension of the PVAR, this is more like a conventional modelling choice. For this, we use model averaging methods following standard Bayesian practice. Of course, the econometric methods developed in this paper could be used to use a single approach throughout (e.g., do only dynamic model selection over all models in our model space).

To be precise, for Ξ_{α}, Ξ_{β}, λ, κ and σ^2, we choose the values that maximize posterior model probabilities. Conditional on the optimal choice of these, we then estimate TVP-PVARs with different numbers of endogenous variables and provide forecasts which average over them. That is, we produce forecasts from every model and then our final forecast is a weighted average of them, where the weights are given by their respective posterior model probabilities.

Calculating posterior model probabilities can also be computationally burdensome, especially with a vast array of models. In addition, most conventional methods are not dynamic (e.g. simply calculating the marginal likelihood for each model). In the remainder of this sub-section, we outline a computationally simple method for the calculation of posterior model probabilities in a dynamic fashion, comparable with the Kalman filter

update rules used for parameter estimation (i.e. we predict time t model probabilities given information at time $t-1$, and then update these probabilities when time t information becomes observed).

Let $M^{(i)}$ for $i = 1, ..., J$ be the set of models under consideration, which in our application are all nested and J is a very large number. In order to be able to learn about the optimal configuration for forecasting we need to quantify a measure of belief for each single model. We follow Raftery et al (2010) and do so by calculating dynamic model probabilities, $p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$, for each model. We use forgetting factor methods to estimate $p\left(M^{(i)}|\mathcal{D}_{t-1}\right)$. 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

$$\omega_{t|t-1}^{(i)} = \frac{\left(\omega_{t-1|t-1}^{(i)}\right)^{\mu}}{\sum_{j=1}^{J}\left(\omega_{t-1|t-1}^{(j)}\right)^{\mu}}, \tag{9}$$

with an updating step

$$\omega_{t|t}^{(i)} \propto \omega_{t|t-1}^{(i)} p\left(Y_t|M^{(i)}, \mathcal{D}_{t-1}\right), \tag{10}$$

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 (9) 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 (i.e. they have a different number of endogenous variables). 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 (see sub-section 3.1) for every country. We refer to the approach as the TVP-PVAR with a dynamic learning prior: TVP-PVAR (DLP). We use this terminology since the prior is hierarchical and dynamic so we can learn about which panel structure is appropriate to impose on the coefficients. Finally, note that it is also possible to include lag length selection as another specification choice and do dynamic model selection over p in a time-varying fashion. However, we do not do so to keep the computational burden manageable. Canova and Ciccarelli (2009) set $p = 1$ for all their specifications involving TVP-PVARs. Allowing for $p > 1$ is possible and we found that $p = 2$ provides optimal inflation forecasts compared to other choices. Hence, our empirical results use $p = 2$.

3 Forecasting Euro Area Inflation

This Section presents the performance of the proposed methods in a standard pseudo out-of-sample forecasting exercise. However, an important feature of our methodology is assessing the in-sample fit. For the sake of brevity we present in-sample estimation results in the online Appendix. There, the reader can find evidence that variances and time-varying parameters are changing, as well as evidence about the panel structure and dimensions of the VAR model changing.

The next subsections present the data, model setup and forecasting results for the problem of forecasting Euro-Area inflation.

3.1 Data

We use $G = 9$ macroeconomic series for $N = 10$ major euro zone countries for the period 1999M1 to 2016M12. The countries are Austria (AT), Belgium (BE), Finland (FI), France (FR), Germany (DE), Greece (GR), Italy (IT), Netherlands (NL), Portugal (PT) and Spain (ES). 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. We also add a 10th variable, the oil price, into all models. But since this does not vary across countries, it is entered only once in the vector of dependent variables. Thus, the largest models we work with have 91 dependent variables. We also consider smaller models with $G = 3, .., 9$ variables (plus the oil price). These models choose variables according to their ordering in the following table (i.e. the model with $G = 3$ uses the first three variables, $G = 4$ uses the first four, etc.).

Our choice of a long term interest rate, as opposed to a short term interest rate (e.g. a monetary policy instrument or the euro overnight index average inter-bank rate or EONIA),

is motivated by our wish to exploit cross-country as well as inter-temporal variation in variables in our panel forecasting exercise. In the euro zone, long term interest rates exhibit substantial variation across countries, whereas short term interest rates do not. Using long term interest rates thus seems sensible in a forecasting exercise, as opposed to a structural analysis where it might be important to include a monetary policy instrument.

Variables	**Explanation**	**Source**	**Tr**
HICP	Indices of Consumer Prices[a]	Eurostat	$\Delta \ln$
UN	Harmonised unemployment rates (%)	Eurostat	*lev*
LTI	Long-Term Interest Rate (10 year, Euro denominated)	Eurostat	*lev*
REER	Real Effective Exchange Rate	Eurostat	$\Delta \ln$
IP	Industrial production index	IMF IFS[b]	$\Delta \ln$
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*
SURVEY4	Consumer Confidence Indicator, SA	Eurostat	*lev*
OIL.PRICE	European Dated BFOE Crude Oil Spot Price[a]	ECB SDW[c]	$\Delta \ln$

[a] Variables that are not seasonally adjusted by the provider, are adjusted by the authors using the X11 filter in Eviews.

[b] International Monetary Fund, International Financial Statistics

[c] European Central Bank, Statistical Data Warehouse

3.2 Models for Comparison

We compare our TVP-PVAR (DLP) approach to several potential competitors. There are, of course, an enormous number of models used for forecasting inflation which we could consider. To focus on the potential benefits of our approach, we emphasize five of its features which, in combination, make our approach distinct from the existing literature. We choose comparators which differ according to one or more of these five features. The five features may briefly be summarized as: i) incorporation of panel structure in the prior in a dynamic fashion, ii) time variation of parameters, iii) type of panel structure used in the prior, iv) allowing for multi-country linkages and v) use of conventional Bayesian hierarchical priors (as opposed to machine learning methods as discussed below). We offer an intuitive

explanation of our alternative approaches in the following paragraphs and provide complete specification details of all approaches in the online Appendix accompanying this article.

First, relative to other approaches for high-dimensional VARs, our dynamic learning prior allows for the model to learn which panel prior is appropriate. This motivates a comparison with other Big Data approaches which do not allow for this sort of learning. Thus, we include a conventional large Minnesota prior VAR similar to the popular specification of Bańbura, Giannone and Reichlin (2010) but with optimal degree of shrinkage estimated as in Giannone, Lenza and Primiceri (2015). We also include a dynamic factor model and a factor augmented VAR. We abbreviate these three approaches as BVAR, DFM and FAVAR, respectively.

Second, relative to other large VAR approaches which do incorporate a panel structure in the hierarchical prior, our approach allows for time-variation in parameters. This motivates comparison with a model with a panel structure similar to our own, but without time-variation in the parameters. We consider a version of the constant parameter model of Canova and Ciccarelli (2009) which is nested within our TVP-PVAR (DLP) approach. This uses their choices of Ξ_P^{α} and Ξ_P^{β} and sets $\lambda = \kappa = 1$, thus ensuring a homoskedastic model with no time-variation in PVAR coefficients. We do not do model averaging over VAR dimensions with this approach. All other specification and modelling choices (including treatment of σ^2) is the same as in our TVP-PVAR (DLP) approach. This is labelled PVAR (CC09) in the tables.

Third, our TVP-PVAR (DLP) considers two different panel structures in the hierarchical prior. However, several other structures have been proposed in the literature. An influential one allows for the investigation of whether there are dynamic or static interdependencies between countries as described in Canova and Ciccarelli (2013). We consider a hierarchical prior which allows for the selection or omission of these types of interdependencies between countries. We adopt the approach of our earlier work, Koop and Korobilis (2016), which develops a simulation algorithm called stochastic search

specification selection (SSSS) to find such interdependencies if they exist and refer to the approach as PVAR (SSSS). We also consider a restricted version of our approach which leads to a single TVP-PVAR of the form considered in Canova and Ciccarelli (2009). That is, we use the full set of endogenous variables, select Ξ_P^{α} and Ξ_P^{β} as done in their paper and set $\lambda = 0.99$ and $\kappa = 1$. Thus, we have a model which is homoskedastic, but allows for time-variation in the PVAR coefficients in the same manner as Canova and Ciccarelli (2009). Our use of forgetting factor methods (as opposed to MCMC) mean estimation and forecasting are computationally feasible.

Fourth, our approach differs from many approaches that estimate a model for each country individually. As a representative of this class of models, we present forecasts from country specific VARs and abbreviate this approach as CS-VAR. We also consider two popular univariate models which we run one country at a time. These are the unobserved components stochastic volatility (UCSV) model of Stock and Watson (2007) and an extension of the UCSV model which allows for AR lags with time varying coefficients on the right hand side. The latter model, which we label TVP-AR, was found by Pettenuzzo and Timmermann (2017) to provide good forecasts of inflation.

Finally, in Big Data models there is a growing interest in use of machine learning methods. Such methods are starting to work their way into the VAR literature and it is of interest to compare our approach to something from this emerging literature. The general idea of this literature is to allow an algorithm to automatically search through the myriad possible specification choices without much input from the economist. This contrasts with conventional approaches used in econometrics where the researcher carefully designs a hierarchical prior based on empirical insight into the problem at hand (e.g. the factor structure Σ_P^{α} reflects the empirical wisdom expressed in Canova and Ciccarelli, 2009, as to how VARs in different countries might be related to one another). In Koop, Korobilis and Pettenuzzo (2017), we developed a particular type of machine learning algorithm using random compression methods for large VARs. In the present paper, we forecast with it

and refer to it as BCVAR. See the Technical Appendix or Koop, Korobilis and Pettenuzzo (2017) for exact details on how this method works, but the key point to stress here is that it does not reflect the multi-country nature of our data set. Rather each equation in the VAR receives an identical treatment and it is left to the algorithm to uncover the nature of any inter-linkages across countries. Complete details of these models are given in the online Appendix.

For the sake of brevity, we do not offer a more extensive comparison to the many models nested within our approach since one of the key advantages of our dynamic learning prior is that it can choose these if it wishes. For instance, it could have chosen the TVP-PVAR with the hierarchical prior of the Canova and Ciccarelli (2009) but, as we saw in the preceding sub-section it did not always do so. Our approach could have chosen constant coefficient versions of our models, either with regards to the VAR coefficients (by choosing $\lambda = 1$) or the error covariance (by choosing $\kappa = 1$), but it does not. Our approach could have chosen smaller PVARs (by choosing $G = 3$), but it does not always do so. Thus, comparison with these types of restrictions on our models is less interesting than comparison with other plausible non-nested alternatives which are described above.

3.3 Forecasting Results

We evaluate statistically forecasts of inflation from the models introduced in the previous sub-section relative to a benchmark approach which produces forecasts using individual AR(2) models for each country's inflation series. We forecast $p_{t+h} - p_t$ where p_t is the log of a country's prices at time t for various forecast horizons. Our forecasts are produced recursively on an expanding window of data. We present Mean Squared Forecast Errors (MSFEs) and Averages of Log Predictive Likelihood (ALPLs) in order to evaluate point and density forecasts, respectively. In the tables in this sub-section, we present results for the 10 inflation rates in the 10 countries in our sample as well as an average across all

countries. For the MSFEs, we use a simple average (and divide by the comparable figure using the AR(2) benchmarks). For the ALPLs, the cross-country average is based on the multivariate predictive density for the 10 inflation series and the comparable figure from the AR(2) benchmark is subtracted off. The forecast evaluation period is 2006M1-2016M12.

It can be seen that, when we look at cross-country averages, our TVP-PVAR (DLP) approach is forecasting best particularly at short horizons. This holds true regardless of whether we use MSFEs or ALPLs as measures of forecast performance. We elaborate on these points in the remainder of this section in relation to the five features mentioned at the beginning of the preceding sub-section.

Incorporating the panel structure in the prior in a dynamic fashion is helping improve forecast performance. A comparison of the TVP-PVAR (DLP) results to a conventional Minnesota prior large VAR indicates the former is forecasting better at all horizons, regardless of whether MSFE or ALPL is used as a forecast metric. The same holds for the DFM and FAVAR approaches.

If we compare our approach to other approaches which do have a panel structure in the prior but do not allow for time variation (i.e. compare TVP-PVAR (DLP) to PVAR (CC09)) we find, without exception, evidence in favour of the TVP-PVAR (DLP). This holds true regardless of whether MSFEs or ALPLs are used as a measure of forecast performance. This suggests that it is not enough to allow for selection of different panel priors in the context of a constant coefficient models, one must do so in a dynamic fashion allowing for the panel structure to change.

Next we compare the forecast performance of our approach to the alternative panel structure in the prior of PVAR (SSSS). Recall that this latter algorithm was proposed by Koop and Korobilis (2016) in order to stochastically search for all combinations of static and dynamic interdependencies that can possibly occur among N countries. However, being a computationally expensive simulation-based approach, this approach only features constant coefficients and covariance. As a consequence, this alternative is also consistently beaten by

our current approach in terms of forecast performance. Similarly PVAR (CC09) which uses the single factor structure of Canova and Ciccarelli (2009) in a constant coefficient model provides forecasts inferior to our approach. Even allowing for time-variation in parameters (but constant volatilities) as in the TVP-PVAR (CC09) does not substantially improve things.

Working with individual country models such as CS-VAR, UCSV or TVP-AR also produces, on average, consistently worse forecast performance over all forecast horizons and both forecast metrics. Finally, the TVP-PVAR (DLP) is even forecasting better than the approach inspired by the machine learning literature: BCVAR.

The preceding conclusions are based on looking at cross-country averages in Tables 1 and 2. A careful examination of the individual country results also supports these conclusions. With only a few exceptions, the TVP-PVAR (DLP) approach forecasts best. The few exceptions occur mainly for Greece, Portugal and Spain. For these countries, which are often referred to as peripheral countries in the European Union, all of our forecasting models do relatively poorly. At longer forecast horizons, when using MSFEs as the forecast metric, the TVP-PVAR (DLP) does not even beat the AR(2) benchmark. Using ALPLs, the performance of TVP-PVAR (DLP) is better for Portugal and Spain, but not for Greece. But for the vast majority of countries, particular those at the core of Europe who were less affected by the eurozone crisis, our method is producing forecasts which beat plausible alternative forecasting models.

Figure 6 sheds light on the forecast performance of our TVP-PVAR (DLP) over time. It presents the cumulative sums of log predictive likelihoods where the sums are taken over time and across countries. It can be seen that, up to 2008 all of the approaches were forecasting roughly as well as one another. However, after some time in 2008 (depending on forecast horizon), TVP-PVAR (DLP) exhibits better forecasting performance than all the other approaches. After middle to late 2009 (depending on forecast horizon), few additional forecast gains are made by TVP-VAR (DLP). Thus, most of the benefits of

Table 1. MSFEs relative to AR(2) for the period 2006M1-2016M12

	AT.INF	BE.INF	FI.INF	FR.INF	DE.INF	GR.INF	IT.INF	NL.INF	PT.INF	ES.INF	AVERAGE
Panel A. Forecast Horizon $h = 1$											
TVP-PVAR (DLP)	0.620	0.785	0.651	0.719	0.881	0.964	0.676	0.854	0.970	0.869	0.799
Restricted versions of our model:											
PVAR(CC09)	0.752	0.891	0.707	0.949	0.890	1.014	0.773	0.931	1.005	1.021	0.893
TVP-PVAR (CC09)	0.729	0.873	0.692	0.926	0.871	0.990	0.759	0.915	0.983	1.011	0.875
Univariate benchmarks:											
UCSV	0.661	0.854	0.721	0.759	0.882	0.888	0.786	0.867	1.034	0.904	0.836
TVP-AR	0.653	0.845	0.711	0.750	0.879	0.880	0.775	0.861	1.036	0.901	0.829
Multivariate benchmarks:											
PVAR (SSSS)	0.782	0.808	0.788	0.758	0.840	0.987	0.747	0.851	1.047	1.093	0.870
CS-VAR	0.675	0.812	0.806	0.766	1.080	1.015	0.715	1.042	1.046	0.918	0.887
BCVAR	0.706	0.894	0.708	0.815	0.960	0.992	0.861	1.004	1.148	1.004	0.909
BVAR	0.716	0.877	0.700	0.917	0.882	0.987	0.763	0.906	0.974	1.015	0.874
DFM	0.720	0.933	0.679	1.022	0.822	0.902	0.859	0.821	1.027	0.948	0.873
FAVAR	0.701	0.892	0.741	0.786	0.913	0.912	0.829	0.896	1.066	0.937	0.867
Panel B. Forecast Horizon $h = 3$											
TVP-PVAR (DLP)	0.647	0.795	0.663	0.873	0.825	1.072	0.849	0.885	1.012	1.025	0.865
Restricted versions of our model:											
PVAR(CC09)	0.745	0.892	0.732	0.945	0.932	0.927	0.796	0.913	1.028	1.126	0.904
TVP-PVAR (CC09)	0.733	0.884	0.716	0.928	0.917	0.913	0.790	0.907	1.009	1.109	0.891
Univariate benchmarks:											
UCSV	0.694	0.806	0.687	0.889	0.856	1.148	0.901	0.931	1.075	0.974	0.896
TVP-AR	0.686	0.794	0.683	0.888	0.855	1.147	0.891	0.930	1.072	0.972	0.892
Multivariate benchmarks:											
PVAR (SSSS)	0.716	0.861	0.725	0.921	0.908	1.123	0.904	0.931	1.098	1.203	0.939
CS-VAR	0.713	0.837	0.720	0.929	0.878	1.140	0.908	0.955	1.082	1.173	0.933
BCVAR	0.872	1.061	0.786	0.993	0.981	0.971	0.935	1.026	1.082	1.077	0.978
BVAR	0.737	0.877	0.727	0.926	0.922	0.930	0.795	0.907	1.006	1.096	0.892
DFM	0.712	1.127	1.017	1.142	0.906	0.934	0.868	0.764	1.123	1.215	0.981
FAVAR	0.728	0.842	0.697	0.922	0.898	1.162	0.936	0.949	1.089	1.202	0.943
Panel C. Forecast Horizon $h = 6$											
TVP-PVAR (DLP)	0.609	0.749	0.628	0.767	0.833	0.982	0.774	0.836	1.007	1.073	0.826
Restricted versions of our model:											
PVAR (CC09)	0.734	0.870	0.712	0.942	0.966	0.932	0.743	0.938	1.046	1.175	0.906
TVP-PVAR (CC09)	0.722	0.858	0.698	0.927	0.951	0.917	0.734	0.913	1.025	1.165	0.891
Univariate benchmarks:											
USCV	0.675	0.813	0.686	0.881	0.869	1.084	0.831	0.880	1.055	0.953	0.873
TVP-AR	0.675	0.807	0.677	0.874	0.863	1.072	0.823	0.877	1.051	0.945	0.866
Multivariate benchmarks:											
PVAR (SSSS)	0.680	0.836	0.680	0.833	0.879	1.076	0.844	0.893	1.088	1.098	0.891
CS-VAR	0.662	0.799	0.702	0.839	0.880	1.095	0.818	0.893	1.029	1.091	0.881
BCVAR	0.762	0.974	0.712	0.935	0.948	0.943	0.795	0.973	1.212	1.080	0.934
BVAR	0.711	0.866	0.690	0.939	0.954	0.934	0.730	0.920	1.031	1.168	0.894
DFM	0.744	0.963	0.704	0.878	0.699	0.931	0.729	0.661	1.082	1.014	0.841
FAVAR	0.700	0.838	0.706	0.900	0.907	1.127	0.858	0.916	1.085	1.141	0.918
Panel D. Forecast Horizon $h = 12$											
TVP-PVAR (DLP)	0.688	0.856	0.835	0.907	0.818	0.931	0.890	0.887	1.095	1.062	0.897
Restricted versions of our model:											
PVAR (CC09)	0.768	0.886	0.881	0.877	0.854	1.061	0.925	0.939	1.096	1.145	0.943
TVP-PVAR (CC09)	0.742	0.868	0.857	0.858	0.834	1.033	0.910	0.923	1.078	1.123	0.923
Univariate benchmarks:											
USCV	0.644	0.769	0.722	0.862	0.781	1.109	0.929	0.940	0.976	0.941	0.867
TVP-AR	0.638	0.766	0.719	0.857	0.779	1.111	0.929	0.931	0.973	0.944	0.865
Multivariate benchmarks:											
PVAR (SSSS)	0.749	0.830	0.749	0.862	0.876	1.121	0.950	0.944	1.143	1.124	0.935
CS-VAR	0.676	0.831	0.722	0.900	0.876	1.103	0.914	0.939	1.098	1.130	0.919
BCVAR	0.836	1.098	0.863	0.945	0.986	0.956	1.004	1.087	1.315	1.073	1.016
BVAR	0.677	0.821	0.757	0.853	0.847	1.038	0.914	0.923	1.080	1.130	0.904
DFM	0.695	0.851	0.810	0.922	0.896	1.252	0.973	1.055	1.083	1.281	0.982
FAVAR	0.662	0.781	0.734	0.888	0.790	1.129	0.959	0.948	1.010	1.120	0.902

Table 2. ALPLs relative to AR(2) for the period 2006M1-2016M12

	AT.INF	BE.INF	FI.INF	FR.INF	DE.INF	GR.INF	IT.INF	NL.INF	PT.INF	ES.INF	AVERAGE
Panel A. Forecast Horizon $h = 1$											
TVP-PVAR (DLP)	3.43	2.52	4.72	5.87	1.35	-0.04	1.72	4.53	1.47	1.72	2.73
Restricted versions of our model:											
PVAR(CC09)	2.06	1.61	1.90	3.24	0.42	-0.64	1.13	1.16	0.45	1.36	1.27
TVP-PVAR (CC09)	2.87	2.17	3.66	5.10	1.09	-0.80	1.01	3.64	0.61	1.15	2.05
Univariate benchmarks:											
UCSV	1.88	2.01	2.50	2.37	0.68	-0.88	0.48	1.99	0.54	0.79	1.24
TVP-AR	1.75	1.88	2.06	3.47	0.65	-0.63	0.53	1.97	0.17	0.81	1.27
Multivariate benchmarks:											
PVAR (SSSS)	2.72	1.80	3.40	5.11	1.12	-0.78	0.80	3.14	0.61	0.80	1.87
CS-VAR	2.80	2.17	3.79	4.97	1.40	-0.97	1.00	3.34	0.57	1.29	2.04
BCVAR	1.78	1.96	2.46	2.35	0.35	-0.86	0.26	2.37	0.45	1.25	1.24
BVAR	1.96	1.90	2.27	3.62	0.66	-0.63	0.57	2.18	0.00	0.66	1.32
DFM	2.34	1.90	3.48	5.02	0.87	-0.70	0.70	3.41	0.56	0.87	1.85
FAVAR	2.14	1.26	3.08	4.46	-0.39	-1.78	0.14	3.39	0.38	0.44	1.31
Panel B. Forecast Horizon $h = 3$											
TVP-PVAR (DLP)	2.98	2.23	4.74	5.58	0.91	-0.23	1.54	4.71	0.86	1.57	2.49
Restricted versions of our model:											
PVAR(CC09)	1.76	1.24	1.99	3.42	0.35	-0.67	0.79	1.85	0.22	1.06	1.20
TVP-PVAR (CC09)	2.59	1.84	3.87	4.85	0.50	-0.76	1.10	3.62	0.40	1.16	1.92
Univariate benchmarks:											
UCSV	1.76	1.30	2.47	2.39	0.02	-0.83	0.78	1.67	0.28	0.58	1.04
TVP-AR	1.94	1.47	2.55	3.64	0.39	-0.98	0.55	2.27	0.17	0.73	1.27
Multivariate benchmarks:											
PVAR (SSSS)	2.65	2.20	3.86	4.65	0.59	-0.76	1.09	3.45	0.27	0.96	1.90
CS-VAR	2.61	2.10	3.49	4.93	0.32	-0.98	1.26	4.06	0.17	1.18	1.91
BCVAR	1.86	1.53	2.29	2.39	-0.34	-0.81	0.56	1.64	0.25	0.37	0.97
BVAR	1.81	1.29	2.99	3.53	0.87	-0.73	0.90	2.20	0.29	1.00	1.42
DFM	2.61	2.42	4.04	4.88	0.70	-0.46	0.99	3.45	0.31	0.89	1.98
FAVAR	1.62	0.60	3.43	4.03	-0.49	-1.86	0.04	2.92	-0.44	0.19	1.00
Panel C. Forecast Horizon $h = 6$											
TVP-PVAR (DLP)	2.40	1.85	4.42	4.98	0.46	-0.52	1.36	4.21	0.54	1.14	2.08
Restricted versions of our model:											
PVAR (CC09)	1.22	0.85	1.80	2.81	-0.10	-0.87	0.55	1.34	-0.06	0.64	0.82
TVP-PVAR (CC09)	1.90	1.42	3.73	4.58	-0.08	-0.96	0.81	3.54	-0.07	0.74	1.56
Univariate benchmarks:											
USCV	1.04	0.86	2.82	2.12	-0.30	-1.11	0.20	1.45	-0.05	0.65	0.77
TVP-AR	1.00	0.64	2.15	3.06	-0.27	-1.15	0.30	1.75	-0.33	0.43	0.76
Multivariate benchmarks:											
PVAR (SSSS)	1.77	1.40	3.40	4.24	-0.10	-1.03	0.86	3.20	0.15	0.47	1.44
CS-VAR	1.53	1.45	3.55	4.99	-0.23	-1.23	0.97	3.80	0.17	0.70	1.57
BCVAR	1.14	0.82	3.02	2.00	-0.31	-0.64	0.18	1.47	0.12	0.76	0.86
BVAR	0.62	0.42	2.00	3.15	-0.48	-0.93	0.57	1.58	-0.21	0.59	0.73
DFM	1.92	1.33	3.32	4.34	0.27	-1.17	0.61	3.24	0.17	0.74	1.48
FAVAR	0.70	0.07	2.65	3.22	-1.12	-1.97	0.11	1.81	-0.95	-0.06	0.45
Panel D. Forecast Horizon $h = 12$											
TVP-PVAR (DLP)	2.10	1.45	3.95	3.88	-0.18	-0.65	0.84	4.71	-0.38	0.75	1.65
Restricted versions of our model:											
PVAR (CC09)	0.99	0.71	1.52	2.37	-0.64	-1.08	0.10	2.43	-0.65	0.18	0.59
TVP-PVAR (CC09)	1.59	1.10	3.22	3.57	-0.57	-1.17	0.35	4.08	-0.61	0.17	1.17
Univariate benchmarks:											
USCV	0.96	0.69	2.27	1.19	-0.69	-1.27	-0.25	2.74	-0.43	-0.42	0.48
TVP-AR	0.78	0.83	1.57	2.43	-0.81	-1.15	0.02	2.37	-0.73	-0.07	0.52
Multivariate benchmarks:											
PVAR (SSSS)	1.58	1.26	2.74	3.52	-0.61	-1.17	0.16	3.64	-0.48	0.30	1.09
CS-VAR	1.80	0.98	2.91	3.79	-0.60	-1.00	0.40	4.02	-0.37	0.27	1.22
BCVAR	1.02	0.80	2.62	1.62	-0.56	-0.94	-0.19	3.09	-0.11	-0.36	0.70
BVAR	0.83	0.91	1.27	2.69	-0.98	-1.38	-0.33	2.18	-0.67	-0.25	0.43
DFM	1.81	1.08	2.67	3.58	-0.58	-0.91	0.13	3.80	-0.49	0.05	1.12
FAVAR	0.78	0.05	2.26	2.39	-1.49	-2.55	-0.74	3.61	-2.25	-0.76	0.13

TVP-VAR (DLP) are to be found at the times of the financial and euro zone crises.

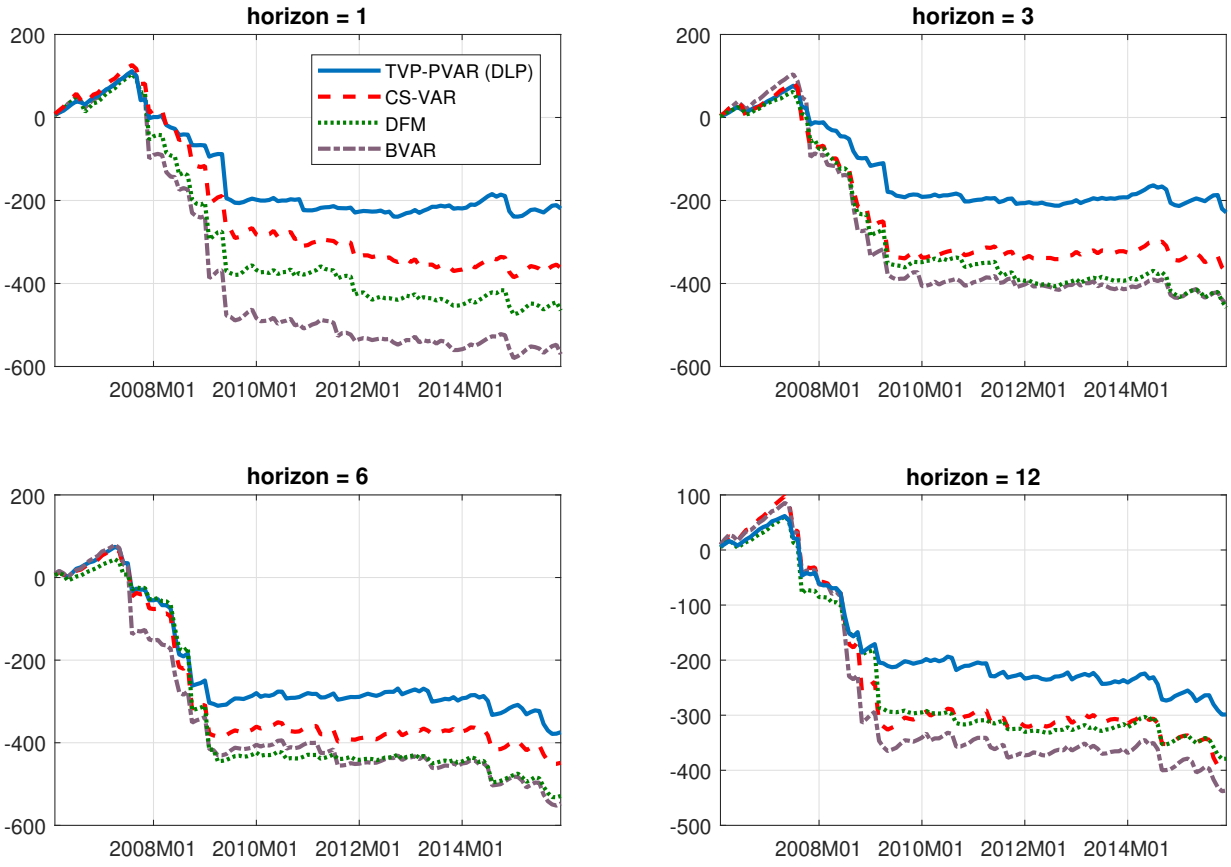

Figure 1: Cumulative sums of log predictive likelihoods (sums taken over time and across countries)

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.

References

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

[2] Bauwens, L., Koop, G., Korobilis, D., and Rombouts, J. (2015). "The Contribution of Structural Break Models to Forecasting Macroeconomic Series," *Journal of Applied Econometrics*, 30, 596-620.

[3] Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). "Measureing the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387-422.

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

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

[6] Canova, F., and Ciccarelli, M. (2013), "Panel Vector Autoregressive models: A Survey," *Advances in Econometrics*, 32, 205-246.

[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] Chudik, A. and Pesaran, H. (2011), "Infinite-Dimensional VARs and Factor Models," *Journal of Econometrics*, 163, 4-22.

[9] Clark, T. E. and Ravazzolo, F. (2015), "Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility " *Journal of Applied Econometrics*, 30, 551-575.

[10] 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.

[11] 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.

[12] Del Negro, M. and Schorfheide, F. (2004), "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643-673.

[13] 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.

[14] Diebold, F. X., Schorfheide, F. and Shin, M. (2017), "Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility," *Journal of Econometrics*, 201, 322-332.

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

[16] 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.

[17] Feldkircher, M. and Huber, F. (2016), "The International Transmission of US shocks – Evidence from Bayesian Global Vector Autoregressions," *European Economic Review*, 81, 167-188.

[18] Giannone, D., Lenza M. and Primiceri, G. (2015), "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, 97, 436-451.

[19] George, E. I., Sun, D. and Ni, S. (2008), "Bayesian Stochastic Search for VAR Model Restrictions," *Journal of Econometrics*, 142, 553-580.

[20] Harvey, D., Leybourne, S. and Newbold, P. (1997), "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting*, 13, 281-291.

[21] Koop, G. (2003), *Bayesian Econometrics.* Chichester: Wiley.

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

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

[24] Koop, G. and Korobilis, D. (2016), "Model Uncertainty in Panel Vector Autoregressive Models," *European Economic Review*, 81, 115-131.

[25] Koop, G., Korobilis, D. and Pettenuzzo, D. (2018), "Bayesian Compressed Vector Autoregressions," *Journal of Econometrics*, forthcoming.

[26] Korobilis, D. (2016), "Prior Selection for Panel Vector Autoregressions," *Computational Statistics and Data Analysis*, 101, 110-120.

[27] Kose, M. A., Otrok, C., and Whiteman, C. (2003), "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review*, 93, 1216-1239.

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

[29] Lütkephol, H. (2005) *New Introduction to Multiple Time Series Analysis.* Heidelberg: Springer Berlin.

[30] Pettenuzzo, D., and Timmermann, A. (2017), "Forecasting Macroeconomic Variables Under Model Instability," *Journal of Business and Economic Statistics*, 35, 183-201

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

[32] 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.

[33] Stock, J. and Watson, M. (2005). "Understanding Changes in International Business Cycle Dynamics" *Journal of the European Economic Association,* 3, 968-1006.

[34] Stock, J. and Watson, M. (2007). "Why has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking,* 39, 3-33.

[35] Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B*, 58, 267-288.