Working Paper Series

Choosing Prior Hyperparameters

WP 16-09

Pooyan Amir-Ahmadi
Goethe University Frankfurt

Christian Matthes
Federal Reserve Bank of Richmond

Mu-Chun Wang
University of Hamburg

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Choosing Prior Hyperparameters

Pooyan Amir-Ahmadi Christian Matthes
Mu-Chun Wang*

Working Paper No. 16-09

August 13, 2016

Abstract

Bayesian inference is common in models with many parameters, such as large VAR models, models with time-varying parameters, or large DSGE models. A common practice is to focus on prior distributions that themselves depend on relatively few hyperparameters. The choice of these hyperparameters is crucial because their influence is often sizeable for standard sample sizes. In this paper we treat the hyperparameters as part of a hierarchical model and propose a fast, tractable, easy-to-implement, and fully Bayesian approach to estimate those hyperparameters jointly with all other parameters in the model.
In terms of applications, we show via Monte Carlo simulations that in time series models with time-varying parameters and stochastic volatility, our approach can *drastically* improve on using fixed hyperparameters previously proposed in the literature.

Keywords: Priors, Bayesian inference, Bayesian VAR, Time variation

*Affiliations: Goethe University Frankfurt (Amir-Ahmadi), Federal Reserve Bank of Richmond (Matthes), and University of Hamburg (Wang). We would like to thank Luca Benati, Fabio Canova, Minsu Chan, Frank Diebold, Luca Gambetti, Thomas Lubik, Frank Schorfheide, Mark Watson, Alexander Wolman, and seminar participants at the University of Pennsylvania as well as conference participants at the EUI workshop on time-varying parameter models and the 9th Rimini Bayesian econometrics workshop for helpful comments. Andrew Owens and Daniel Tracht provided excellent research assistance. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1 Introduction

Multivariate time series models form the backbone of empirical macroeconomics. A common feature of all popular multivariate time series models is that, as researchers include more variables, the number of parameters quickly grows large. This feature is present to some degree not only in Vector Autoregressions (VARs, Sims (1980)), or VARs that feature time-varying parameters and/or stochastic volatility (Primiceri (2005), Cogley & Sargent (2005), Sims & Zha (2006)), but also in multivariate time series models that try to economize on the number of parameters such as factor models (see for example Stock & Watson (2002)) or Dynamic Stochastic General Equilibrium (DSGE) models (Smets & Wouters (2007)).

Arguably the most popular way to tackle the issue of the large number of parameters has been to use Bayesian inference, which, via its use of priors, allows researchers to avoid overfitting the observed sample (which comes at the cost of unrealistic out-of-sample behavior). Eliciting priors in such high-dimensional models is a daunting task, though. A common practice is to focus on prior distributions that themselves depend on a substantially smaller number of parameters (which we will call hyperparameters). One prominent example that uses this approach is the 'Minnesota' prior for VARs (Doan, Litterman & Sims (1984)), which is especially useful in applications with many observable variables (Banbura, Giannone & Reichlin (2010)).

The choice of hyperparameters is crucial because their influence is often sizeable for standard sample sizes. Nonetheless, the choice of those hyperparameters is often ad-hoc in the literature. In this paper, we propose a fast, tractable, and easy-to-implement Metropolis step that can easily

be added to standard posterior samplers such as the Metropolis-Hastings algorithm or the Gibbs sampler (Gelman, Carlin, Stern, Dunson, Vehtari & Rubin (2013)).[1]Researchers can use our approach with minimal changes in their code (and negligible increase in runtime) to estimate these hyperparameters. The estimation algorithm that we present in this paper exploits the hierarchical structure that is automatically present whenever prior hyperparameters are used, and thus can be used generally in *any* model with prior hyperparameters. Our approach interprets the structure implied by the interaction of parameters of the model and the associated prior hyperparameters as a hierarchical model, which is a standard model in Bayesian inference (Gelman et al. (2013)).

In terms of applications, we focus in this paper on one class of multivariate time series models where the influence of prior hyperparameters is especially pronounced: VARs that feature time-varying parameters and stochastic volatility, which have been introduced by the work of Cogley & Sargent (2005) and Primiceri (2005), building on earlier contributions such as Doan et al. (1984), Stock & Watson (1996) and Cooley & Prescott (1976). These models are commonly estimated using Bayesian methods and can feature a plethora of parameters, thus making the choice of priors important. The work of Cogley & Sargent (2005) and Primiceri (2005) has established a de-facto standard on how to set the priors for these models using a training sample and relatively few hyperparameters.[2] In this paper, we show in a Monte Carlo exercise that using our approach instead of relying on standard values for the hyperparameters can *dras-*

[1]The Gibbs sampler can be viewed as a special case of the Metropolis-Hastings algorithm (see again Gelman et al. (2013))

[2]Examples of papers that follow this structure are, among many others, Canova & Gambetti (2009), Clark & Terry (2010) Benati & Lubik (2014), Kliem, Kriwoluzky & Sarferaz (2013), Koop, Leon-Gonzalez & Strachan (2011), and D'Agostino & Surico (2011).

tically improve the estimation of the parameter and volatility paths. We also show that our approach can lead to substantially different conclusions in applications with real data - we study the joint dynamics of US and UK inflation for a monthly sample covering the last century.

The Gibbs sampler is already the standard approach to estimate models in this class and thus our approach fits naturally into the estimation approach used for these models.[3] Our approach allows researchers to estimate hyperparameters for the time variation in all parameters (including the volatilities of the residuals) without restricting those parameters to possibly take on only a small finite number of values.

The importance of hyperparameters in this class of models has been established by Primiceri (2005), who also estimates the hyperparameters (to our knowledge, the only other paper that does so in a Bayesian context for these models). Unfortunately, Primiceri's approach to estimating the prior hyperparameters is computationally very involved and requires focusing on only a small number of possible values for the hyperparameters.[4] Since the hyperparameters interact with the part of the prior that is set via the use of a training sample (which depends crucially on the specific data sample), it is also not clear that the same discrete grid of possible parameter values that Primiceri used should be employed for other applications.

Since Primiceri (2005), most applications of models in this class use his estimated values for the hyperparameters or other fixed values on an

[3]Even if the model of interest has previously been estimated with another posterior sampler, our approach can be used in that case - the previous posterior sampler then forms one block of a multiple-block algorithm, with our proposed step forming the other.

[4]To be more specific, Primiceri (2005) uses a reversible jump MCMC algorithm to estimate the hyperparameters. To get proposal densities, he in a first step has to estimate his time-varying parameter model conditional on *each* possible combination of hyperparameters.

ad-hoc basis. Furthermore, these papers do not take into account that uncertainty about these hyperparameters could influence later inference, whereas our approach automatically takes this uncertainty into account. Some readers might wonder why the choice of prior hyperparameters is important. Shouldn't the importance of the prior vanish as the data size increases? In this paper, we show that the hyperparameters influence estimation outcomes for the class of models we consider and standard sample sizes available for macroeconomic analysis.[5]

While we present our approach using a time-varying parameter VAR with stochastic volatility, the extension to other time-varying parameter models such as factor augmented VARs with time varying parameters and stochastic volatility (see Ellis, Mumtaz & Zabczyk (2014) and Baumeister, Liu & Mumtaz (2013), for example) or factor models with time-varying parameters and stochastic volatility is straightforward because of the modular structure of the Gibbs sampler.[6] Our approach will also be of interest to researchers trying to estimate multivariate extensions of the univariate stochastic volatility model presented in Kim, Shephard & Chib (1998). As will become evident in our discussion of the models we study, this class of models allows for the stochastic increments in the stochastic volatility processes for different variables to be correlated, whereas in the finance literature on multivariate stochastic volatility, those increments are often assumed to be independent (Chib, Nardari & Shephard (2006)). Our approach can also be used for DSGE models with stochastic volatility as introduced by Justiniano & Primiceri

[5]This echoes the results in Reusens & Croux (2015), who carry out an extensive Monte Carlo study of prior sensitivity using a VAR with time-varying parameters, but no stochastic volatility.

[6]In the appendix, we show how to augment the standard Gibbs sampler for fixed coefficient VARs to include the estimation of prior hyperparameters.

(2008).[7]

Stock & Watson (1996) propose a frequentist approach to estimate scaling parameters in the law of motion for time-varying parameter models. Benati (2015) adapts their approach to a time-varying parameter VAR model without stochastic volatility. Benati's approach is computationally more involved than ours and a mix of Bayesian and frequentist approaches, thus making it harder to interpret in the otherwise Bayesian estimation of these models. Furthermore, he focuses on the hyperparameter for the coefficients (since his model does not feature stochastic volatility), while our approach is general enough to estimate the hyperparameters in the law of motion for stochastic volatilities.

Lubik, Matthes & Owens (2014) calibrate the hyperparameters using a prior predictive approach: They simulate paths for parameters and volatilities from the prior and choose fixed values for the hyperparameters so that a pre-specified set of moments of the simulated parameters and volatilities approximate as close as possible the corresponding moments obtained by estimating fixed coefficients VARs on a set of rolling training samples.

Our paper is more generally related to the literature on choosing prior hyperparameters in Bayesian inference. Giannone, Lenza & Primiceri (2013) estimate prior hyperparameters for time-invariant VARs with conjugate Normal-Inverse Wishart priors by exploiting the fact that in this case the density of the data conditional on the hyperparameters is known in closed form, which they maximize with respect to the hyperparameters. Therefore they do not have to use a Gibbs sampler. On the other hand, their approach only delivers a point estimate of the hyperparam-

[7] Incorporating stochastic volatility in DSGE models substantially improves the forecasting performance, as highlighted by Diebold, Schorfheide & Shin (2016).

eters and thus does not take into account the uncertainty surrounding the parameters. Our approach can be applied to any model in which prior hyperparameters are present and thus presents an alternative to the approach in Giannone et al. (2013) for the fixed coefficient VARs they study. In the models we focus on in this paper, there is no closed form for the marginal data density. The approach by Giannone et al. (2013) can thus not be easily extended to time-varying parameter models. As highlighted by Giannone et al. (2013), their approach is an empirical Bayes approach, while our approach focuses on the hierarchical structure imposed by the use of prior hyperparameters.

In an early attempt to tackle the problem of estimating prior hyperparameters, Lopes, Moreira & Schmidt (1999) propose an alternative procedure to estimate hyperparameters using sampling importance resampling. Their approach requires the calculation of the marginal likelihood conditional on the hyperparameters of interest, i.e. the density of data conditional *only* on the hyperparameters, with all other parameters integrated out. Computing even one such marginal likelihood is a computationally daunting task in the models we focus on in this paper (to the point that marginal likelihoods are basically never computed in this literature). The approach in Lopes et al. (1999) would require the computation of such a marginal likelihood for *every* unique draw of the hyperparameters, thus rendering it infeasible for the applications we are interested in. Furthermore, in the class of models we study, researchers regularly use loose priors. It is well known (Gelman et al. (2013)) that in the case of loose priors, the exact specification of those priors has a substantial influence on the value of the marginal likelihood, even though point estimates and error bands are largely unaffected. Our approach,

on the other hand, is not sensitive to the priors for the other parameters in the model.

Korobilis (2014) estimates some prior parameters in a VAR with time-varying parameters and stochastic volatility. He restricts the prior covariance of the innovations to the parameters to be diagonal. Those diagonal elements are then estimated in a Gibbs sampling step. His approach could be combined with ours since Korobilis (2014) relies on prior hyperparameters for the prior covariance matrix of the innovations to the parameters.

In the next section, we describe the general algorithm before turning to time-varying parameter models in section 3. We then carry out a simulation study in section 4 before showing the effect of estimating prior hyperparameters on a real data set of historical inflation data for the US and the UK in section 5.

2 How to Estimate Prior Hyperparameters

In this section, we derive a Metropolis step to estimate prior hyperparameters. The model is given by a likelihood function $p(Y|\Omega, H, h)$ where Y is the matrix of data (note that here the data is not necessarily time series data), Ω is the set of all parameters except for the parameter block H associated with the hyperparameter vector h. The prior for H, $p(H|h)$, depends on the hyperparameter h. To give a specific example, it might be useful to think of h as the scaling parameters for the Minnesota prior used in the Bayesian estimation of VARs - then H would be the intercepts and the coefficients on lagged observables. We assume that Ω and H are estimated via Gibbs-sampling or the (possibly multiple-block) Metropolis-

Hastings algorithm, as described, for example, in Gelman et al. (2013). The augmented algorithm that includes the estimation of the hyperparameters then alternates between draws from the algorithm to draw Ω and H (both those steps condition on a value for h) and the drawing of h conditional on H and Ω, which we describe in this section.[8] The prior beliefs about the hyperparameter h are encoded in a prior distribution $p(h)$. From a conceptual point of view, a researcher could introduce another level of hierarchy and make the prior for h depend on more hyperparameters as well. Since we are concerned with applications where the dimensionality of h is already small (such as the time-varying parameter models we describe later), we will not pursue this question further in this paper - our approach could be extended in a straightforward matter if a researcher was interested in introducing additional levels of hierarchy. We focus here on drawing one vector of hyperparameters, but other hyperparameters could be included in Ω (which could be high-dimensional, as in our time-varying parameter VAR later). Draws for those other hyperparameters would then be generated using additional Metropolis steps that have the same structure. We assume that the following assumptions hold (assumption 1 is only necessary if multiple vectors of hyperparameters are present in the model):

Assumption 1 *The different vectors of hyperparameters are* a priori *independent of each other.*

Assumption 2 *All parameters of the model except for the parameter block directly linked to a specific hyperparameter are* a priori *independent of that specific hyperparameter.*

[8] In the appendix, we lay out the estimation algorithm for a VAR with time-varying parameters and stochastic volatility in detail.

Neither of these assumptions are restrictive. If assumption 1 is violated, the dependent vectors of hyperparameters just have to be grouped into one larger vector of hyperparameters.[9] The modifications for the algorithm in this case are straightforward. Violations of the second assumption can be handled similarly: The different parameter blocks whose priors depend on the same hyperparameters have to be grouped together in one larger parameter vector, which then depend on the same vector of hyperparameters.

Deriving a Metropolis step for h amounts to deriving a formula for the acceptance probability in the Metropolis-Hastings step. We draw a realization from the proposal density q, which will be accepted with probability α^i at iteration i of the algorithm. This acceptance probability in the Metropolis-within-Gibbs step at iteration i [10] is given by

$$\alpha^i = \min\left(1, \frac{p(\Omega, h, H|Y)q(h^{prop}|h^{i-1})}{p(\Omega, h^{i-1}, H|Y)q(h^{i-1}|h^{prop})}\right) \tag{1}$$

a superscript $prop$ denotes a proposed value, a superscript $i-1$ denotes values from iteration $i-1$ of the algorithm and superscripts are dropped for H and Ω for convenience. We now simplify α^i in this general environment.

First, we rewrite $p(\Omega, h, H|Y)$:

$$p(\Omega, h, H|Y) \propto p(Y|\Omega, h, H)p(\Omega|h, H)p(H|h)p(h) \tag{2}$$

By the hierarchical nature of the model (the hyperparameters only enter the prior for H), $p(Y|\Omega, h, H)$ does not depend on h since it conditions on

[9] We later spell out these assumptions in more detail for our VAR model.

[10] We call this step a Metropolis-within-Gibbs step even if all other parameter blocks in the model are estimated with Metropolis steps themselves.

H. Thus, $p(Y|\Omega, h, H)$ cancels out in the numerator and denominator of α^i. By assumption 2 and the hierarchical nature of the hyperparameter structure (and, if necessary, assumption 1), the term $p(\Omega|h, H)$ equals $p(\Omega|H)$, which then also cancels out in the fraction determining α^i. We are left with

$$\alpha^i = \min\left(1, \frac{p(H|h^{prop})p(h^{prop})q(h^{prop}|h^{i-1})}{p(H|h^{i-1})p(h^{i-1})q(h^{i-1}|h^{prop})}\right) \tag{3}$$

A key insight to this equation is that all identities that need to be valuated are either the proposal density q or prior densities ($p(h)$ is the prior density for h while $p(H|h)$ is the prior density of H, which depends on the hyperparameter h). Generally those densities are known in closed form and thus fast to evaluate, thus making our algorithm computationally efficient.

3 The VAR Model and the Estimation of Hyperparameters

This section presents the class of models we us in our applications and the necessary additional steps in the Gibbs-sampling algorithm for time-varying parameter VARs to estimate the prior scale parameters. As another specific example, the appendix shows how to use our approach with a popular specification for fixed coefficient VARs.

The observable vector y_t is modeled as:

$$y_t = \mu_t + \sum_{j=1}^{L} A_{j,t} y_{t-j} + e_t \tag{4}$$

where the intercepts μ_t, the autoregressive matrices $A_{j,t}$ and the covariance matrix of e_t are allowed to vary over time. To be able to parsimoniously describe the dynamics of our model, we define $X_t' \equiv I \otimes (1, y_{t-1}', ..., y_{t-L}')$ and rewrite (4) in the following state space form[11]:

$$y_t = X_t' \theta_t + e_t \tag{5}$$

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

The observation equation (5) is a more compact expression for (4). The state equation (6) describes the law of motion for the intercepts and autoregressive matrices. The covariance matrix of the innovations in equation (5) is modeled following Primiceri:

$$e_t = \Lambda_t^{-1} \Sigma_t \varepsilon_t \tag{7}$$

The covariance state Λ_t is a lower triangular matrix with ones on the main diagonal and representative non fixed element λ_t^i. Σ_t is a diagonal matrix with representative non fixed element σ_t^j. The dynamics of the non fixed elements of Λ_t and Σ_t are given by:

$$\lambda_t^i = \lambda_{t-1}^i + \zeta_t^i \tag{8}$$

$$\log \sigma_t^j = \log \sigma_{t-1}^j + \eta_t^j \tag{9}$$

To conclude the description of our model, we need to make distributional assumptions on the innovations ε_t, u_t, η_t, and ζ_t, where η_t and ζ_t are vectors of the corresponding scalar innovations in the elements of Σ_t and Λ_t. We assume that all these innovations, which govern the time variation

[11] I denotes the identity matrix.

for the different parameters in this models, are normally distributed with covariance matrix V, which we, following Primiceri, restrict as follows:

$$V = Var\left[\begin{pmatrix} \varepsilon_t \\ u_t \\ \zeta_t \\ \eta_t \end{pmatrix}\right] = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix} \tag{10}$$

S is further restricted to be block diagonal with J blocks, which simplifies inference (this is inconsequential for our extension to the standard Gibbs sampler, but we decided to use the standard model in the literature). Note that W, on the other hand, is not restricted, allowing the increments in the stochastic volatility processes to be correlated.

We will now describe the estimation of general prior hyperparameters in this setting before turning to the specific prior hyperparameters used by Primiceri (2005) and the subsequent literature. We repeat, for expository purposes, the derivation of the Metropolis-Hastings acceptance probability from the previous section as applied to this specific model.

The priors for Q, S, and W are given by:

$$Q \sim p_Q(h_Q) \tag{11}$$

$$W \sim p_W(h_W) \tag{12}$$

$$S_j \sim p_{Sj}(h_{Sj})\ j = 1, \ldots, J \tag{13}$$

where $h_i,\ i \in (Q, W, S_j)$ denotes the vectors of hyperparameters for each set of matrices. S_j is the j-th block of S.

We are interested in estimating the hyperparameters $h_Q, h_W, \{h_{Sj}\}_{j=1}^J$. To do so, we attach priors $p_X(X)$ to the hyperparameters ($X = \{h_Q, h_W, \{h_{Sj}\}_{j=1}^J\}$).

In our empirical applications, we assume that the prior specification for all other parameters are the same as in Primiceri (2005), but this is inconsequential for our algorithm. We denote by Ω all parameters to be estimated except for the prior hyperparameters themselves and the associated covariance matrices Q, W, and $\{S_j\}_{j=1}^{J}$. Our approach builds on the insight that equations (11) to (13) can be interpreted as a hierarchical model, which in our case is embedded in a larger model, the VAR with time-varying parameters and stochastic volatility. Bayesian inference lends itself naturally to the estimation of hierarchical models because they can be analyzed with a Gibbs sampler or the multiple-block Metropolis-Hastings algorithm. This ease of estimation and the flexibility of hierarchical models explains their popularity (see for example Gelman et al. (2013)).

We now restate the two assumptions used to derive our algorithm for the specific model at hand:

Assumption 3 *The different vectors of hyperparameters are* a priori *independent of each other:*

$$p(h_Q, h_W, h_{S1}, ..., h_{SJ}) = p_{h_Q}(h_Q)p_{h_W}(h_W)p_{h_{S1}}(h_{S1}) \cdots p_{h_{SJ}}(h_{SJ})$$

Assumption 4 *All parameter blocks of the model except for the parameter block directly linked to a specific hyperparameter (via one of the equations 11 through 13 in this model) are* a priori *independent of that specific hyperparameter (e.g.* W *and* S_j $j = 1, \ldots, J$ *are* a priori *independent of* h_Q*).*

As long as we assume that p_Q, p_W, and p_{Sj} are all inverse Wishart distributions (as is standard in the literature), the drawing of the covariance

matrices themselves can be carried out just as in the algorithm described in Del Negro & Primiceri (2015) once we condition on the hyperparameters.

To estimate the hyperparameters, we use a Metropolis-within-Gibbs step (Geweke (2005)) for each vector of hyperparameters. We focus here on the estimation of h_Q because the other blocks are conceptually the same. The acceptance probability α^i at iteration i of the Metropolis-within-Gibbs algorithm is given by:

$$\alpha^i = \min\left(1, \frac{p(\Omega, h_Q^{prop}, Q, \{h_{Sj}\}, \{S_j\}, h_W, W|y^T)q(h_Q^{prop}|h_Q^{i-1})}{p(\Omega, h_Q^{i-1}, Q, \{h_{Sj}\}, \{S_j\}, h_W, W|y^T)q(h_Q^{i-1}|h_Q^{prop})}\right) \tag{14}$$

where a superscript $prop$ denotes the prosed value and a superscript $i-1$ the value from the previous iteration (superscripts are dropped for all other parameters for ease of reading). y^T is the history of observables used for estimation ($y^T = \{y_t\}_{t=1}^T$). q is the proposal density.

For each vector of hyperparameters, we need to calculate the following posterior:

$$p(\Omega, h_Q, Q, \{h_{Sj}\}, \{S_j\}, h_W, W|y^T) \propto \tag{15}$$
$$p(y^T|\Omega, h_Q, Q, \{h_{Sj}\}, \{S_j\}, h_W, W)p(\Omega, \{h_{Sj}\}, \{S_j\}, h_W, W|h_Q, Q)p(Q|h_Q)p(h_Q)$$

Two observations are key to simplifying this formula: First, once we condition on Ω, the conditional density of y^T is independent of h_Q and thus cancels out in the calculation of the Metropolis-Hastings acceptance probability because it appears in both the denominator and the numerator. This can be seen by studying equation (4) and noting that Ω contains the sequence of $\{\mu_i, \{A_{j,i}\}_{j=1}^k, \Lambda_i, \Sigma_i\}_{i=1}^T$. Once we know Ω and Q, h_Q is not needed for the calculation of this density.

The second point worth noting is that conditional on Q, h_Q does not carry any information about any of the other parameters. This is due to the hierarchical nature of the model and the two assumptions made before. Thus we get that

$$p(\Omega, \{h_{Sj}\}, \{S_j\}, h_W, W|h_Q, Q) = p(\Omega, \{h_{Sj}\}, \{S_j\}, h_W, W|Q) \tag{16}$$

As a result, we find that $p(\Omega, \{h_{Sj}\}, \{S_j\}, h_W, W|Q)$ also cancels out in the acceptance probability since it is not a function of h_Q. The acceptance probability then simplifies to

$$\alpha^i = \min\left(1, \frac{p(Q|h_Q^{prop})p(h_Q^{prop})q(h_Q^{prop}|h_Q^{i-1})}{p(Q|h_Q^{i-1})p(h_Q^{i-1})q(h_Q^{i-1}|h_Q^{prop})}\right) \tag{17}$$

$p(Q|h_Q)$ is the prior density for Q described above (which is usually an inverse Wishart density) and $p(h_Q)$ is the prior on h_Q. Once we have fixed a proposal density for h_Q, evaluating the acceptance probability is thus straightforward. Not only can the same argument be made for the other hyperparameters introduced before, but for any hyperparameter since the logic used for deriving the acceptance probability only hinges on the hierarchical nature of the model with respect to the prior hyperparameters.

Now turning to the exact specification in Primiceri (2005), the priors for Q, S and W are set as follows:

$$Q \sim IW(k_Q^2 df_Q V_Q, df_Q) \tag{18}$$

$$W \sim IW(k_W^2 df_W V_W, df_W) \tag{19}$$

$$S_j \sim IW(k_S^2 df_{Sj} V_{Sj}, df_{Sj}) \tag{20}$$

where df denotes the degrees of freedom, IW is the inverse Wishart distribution, V are prior scaling matrices, and $k_X, X \in \{Q, W, S\}$ are the scalar hyperparameters we want to estimate.

In this paper, we focus on the estimation of low-dimensional hyperparameters. In theory, our algorithms could be adapted to estimate the prior scaling matrices V; however, for most practical applications the V matrices are high-dimensional objects, so we focus instead on picking the V matrices using a training sample, as is standard in the literature.

One difference relative to the general algorithm above is that, to be in line with Primiceri and the subsequent literature, we use the same k_S for all blocks of S. For the different blocks of S, we use the fact that conditional on k_S the priors for the different blocks are independent inverse-Wishart densities. Thus, in that case we get

$$P(S|k_S) = \prod_{j=1}^{J} P(S_j|k_S) \tag{21}$$

4 Simulation Study

We use a univariate AR(1) process in a simulation study to demonstrate the properties of our approach. We plot results for 5 representative samples, but also report summary statistics based on 500 samples. The data-generating process is an AR(1) process with time-varying AR parameter and stochastic volatility:

$$\begin{aligned} y_t &= \phi_t y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} N(0, \sigma_t^2) \\ \phi_t &= \phi_{t-1} + e_t, \qquad e_t \overset{iid}{\sim} N(0, Q) \\ \log(\sigma_t) &= \log(\sigma_{t-1}) + u_t \qquad u_t \overset{iid}{\sim} N(0, W) \end{aligned}$$

We set $Q = 0.01$ and $W = 0.001$. The initial values are $\phi_0 = 0.5$ and $\sigma_0 = 0.1$. In addition, we constrain the AR-coefficient ϕ_t to be strictly smaller than unity and the stochastic volatility σ_t^2 to be strictly smaller than 100.[12] We generate 450 observations and discard the first 100, so that the final sample size is $T = 350$ (from which we use 40 periods as a training sample to initialize the prior in the same fashion as Primiceri (2005)). We approximate the posterior using 5000 draws and use an additional 5000 draws to adaptively set the proposal densities for the hyperparameters. The simulated parameters exhibit substantial movements in both the AR parameter and residual volatility. We report simulation results using 5 separate simulations.

We use independent inverse gamma priors for the hyperparameters. The parameters of those priors are chosen such that the prior mode is at 0.05 and the variance is infinite. Flat priors give similar results in this application.

Figure (1) shows actual and estimated paths of the AR coefficient. We compare the estimation procedure with fixed tuning parameters (fixed at the values used in Primiceri (2005), $k_Q = 0.01$ and $k_W = 0.01$) to our approach.[13] The bold lines represent the posterior medians at each point in the sample, while the dashed lines are the 68 % posterior bands (the blue lines are the results for our approach, while the red lines are the estimated parameter paths with fixed hyperparameters). Evidently, if the true data-generating process exhibits a substantial degree of time variation in the AR parameter, fixing the hyperparameters can lead to a *dramatic* underestimation of the degree of time variation of the AR

[12] We do this to generate paths of observables that resemble standard economic time series.

[13] The title of each figure in this section shows whether or not the true data-generating process features fixed coefficients or time-varying coefficients.

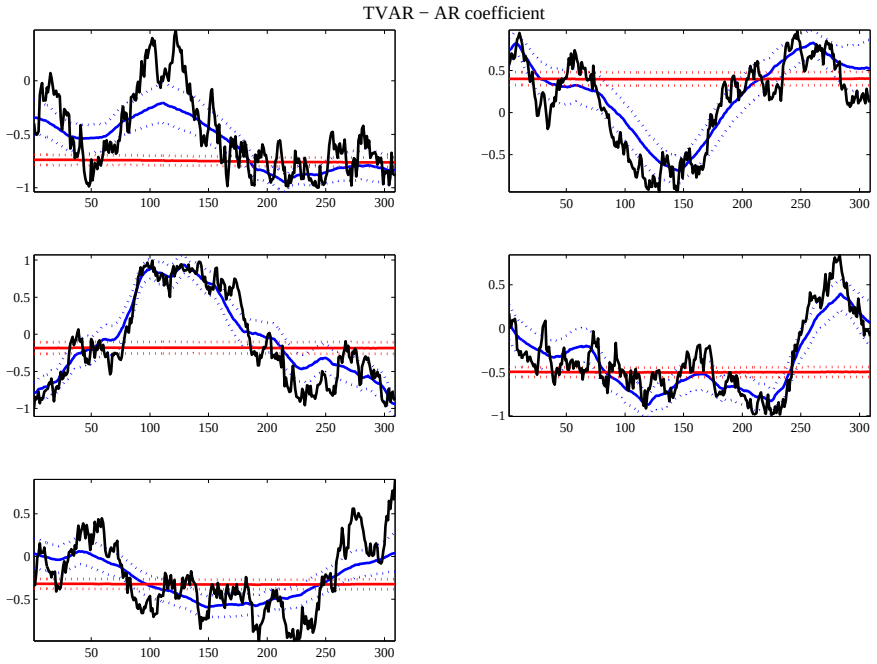

Figure 1: Actual (in black) and estimated AR coefficients (for both estimated (in blue) and fixed (in red) values for the hyperparameters).

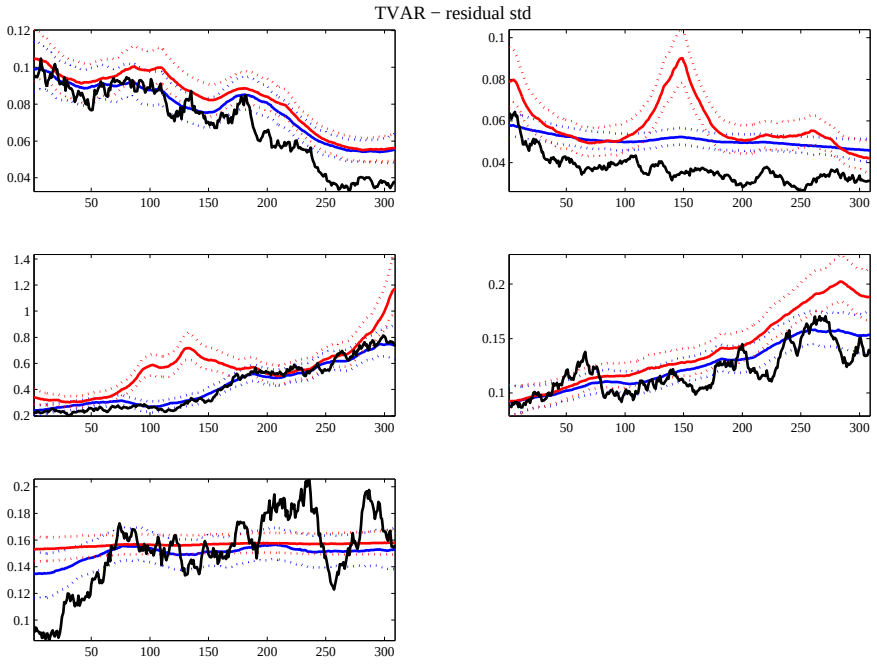

Figure 2: Actual (in black) and estimated standard deviations (for both estimated (in blue) and fixed (in red) values for the hyperparameters).

coefficient. In fact, the estimation results for the case with fixed hyperparameters suggest that there is basically no time variation in the AR

parameter at all. Our approach, instead, can capture the time variation present in our simulations. As can be seen in figure 1, estimating the hyperparameters does *not* lead to estimates of time variation in parameters that overstate the true amount of time variation. In fact, the estimated parameter paths still resemble smooth approximations to the true paths of the AR parameter. It is worth pointing out here that many studies that use time-varying parameter VARs with stochastic volatility indeed find that many parameters do not vary substantially over time (Cogley & Sargent (2005)). We will see in the empirical application below that the findings in this section carry over to real-world examples.

The effects of underestimating the time variation in the AR parameter can be seen when looking at the estimated parameter paths for σ_t in figure 2, which can show a substantial upward bias in the case of fixed hyperparameters. This is particularly striking in the upper right panel and the middle left panel of figures 1 and 2, where we can see that with fixed hyperparameters and substantial time variation in parameters (which is most present in those two panels), the model interprets large changes in the AR coefficients as changes in the residual covariances. The main takeaway from this exercise is *not* that the hyperparameters in Primiceri (2005) are 'wrong' in any sense, but rather that, if a researcher is interested in a very different application (including, but certainly not limited to, a different number of observables), then that researcher should think carefully about the prior hyperparameters. We offer one data-driven way to take the dependence of the results on the prior hyperparameters into account.

To get a sense how much more time variation our approach allows, it is instructive to study the estimated values for the hyperparameters. Across

parameter	fixed hyperparameters	estimated hyperparameters	ratio
ϕ_t	0.46	0.23	2
σ_t	0.03	0.01	1.79

Table 1: RMSEs for the median parameter estimates

simulations the estimated posterior median values for k_Q vary from $.21$ to $.63$, whereas the estimated median values for k_W range from $.04$ to $.07$. Contrasting this with the values obtained in Primiceri (2005) and the associated parameter paths, we can see that larger values for the scaling parameters substantially influence the estimated paths of parameters, even with a large sample and few parameters as in our simulations.

To make sure that our results are not driven by the small number of 5 samples, we now report root mean squared errors (RMSEs) for 500 simulations using the same setup as described before. The root mean squared errors are constructed using the posterior median at each point in time as the point estimate. Table 1 confirms the pattern that we saw in the figures: With estimated hyperparameters, the estimates of time-varying parameters and volatilities are much better able to track the true values (the root mean squared errors for the parameter ϕ_t are twice as large with fixed hyperparameters and the corresponding number for the volatilities σ_t is 79 % larger).

A possible concern for users of our approach might be that it might overestimate the amount of time variation if no or little time variation is present. To address this concern, we repeat the same exercise with a data-generating process that features fixed coefficients (fixed at the initial values used in the previous exercise). The results are shown in figures 3 and 4. While our approach leads to slightly more estimated time

variation for both parameters of the model[14], the results for both approaches are very similar. In particular, whenever the true parameter values are included in the posterior bands of one approach, the true parameter values are also included in the bands estimated using the other approach.

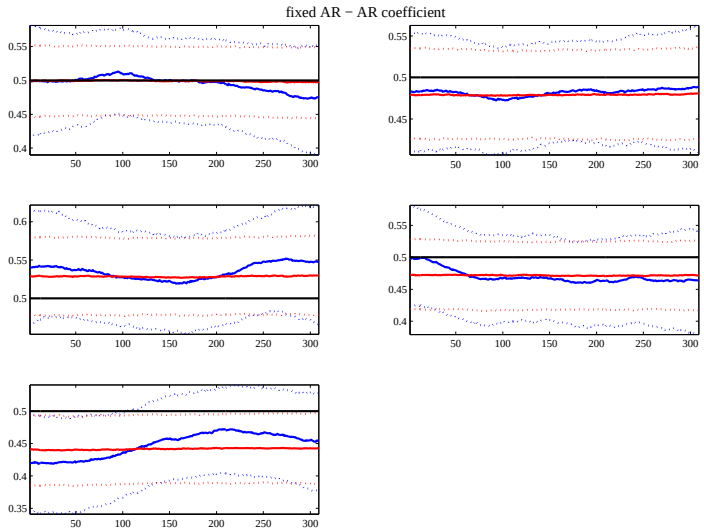

Figure 3: Actual (in black) and estimated AR coefficients (for both estimated (in blue) and fixed (in red) values for the hyperparameters).

5 Empirical Application

VARs with time-varying parameters and stochastic volatility have previously been used to analyze long historical time series - see, for example, Sargent & Surico (2011), D'Agostino & Surico (2011), and Ahmadi, Matthes & Wang (2016).[15] As the sample size increases, there seems more reason to allow for the possibility of changing parameters

[14]The estimated posterior median values for k_Q range from .04 to .05, whereas the values for k_W range from .03 to .04.

[15]Other papers that have used time series models with time-varying parameetrs and stochastic volatility (not necessarily VARs) include Cogley & Sargent (2015) and Schorfheide, Song & Yaron (2014).

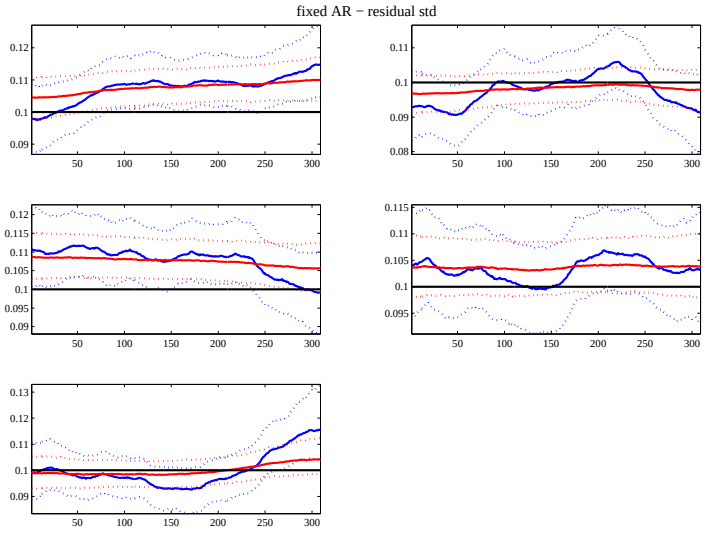

Figure 4: Actual (in black) and estimated standard deviations (for both estimated (in blue) and fixed (in red) values for the hyperparameters).

and volatilities. These changes can come from various sources - technological progress, changes in institutions, political changes, and international conflicts are just some of the reasons why we might suspect that constant parameter models are ill-suited for long historical samples. With long historical time series that are very different from the time series used by Primiceri (2005), there is little reason to believe a-priori that the hyperparameters chosen by Primiceri should reflect a researcher's view of the amount of time variation present in the data. To assess the importance of estimating the prior hyperparameters, a long historical dataset thus seems like a useful laboratory. We study a VAR with monthly year-over-year CPI inflation data for the US and the UK starting in 1915.[16]

The first 95 months are used as the training sample to initialize the priors. To be consistent with most of the literature, we use 2 lags.[17] We use inverse gamma priors for the hyperparameters. For simplicity, we use

[16]The data is from the Global Financial database.

[17]We use 50000 draws to adaptively pick the parameters of the proposal densities for the hyperparameters and another 50000 draws to actually approximate the posterior.

the same (loose) priors for all three hyperparameters that we have used in our Monte Carlo simulations. As a reminder, we pick the parameters for the prior of the hyperparameters such that the prior mode is at 0.05 and the variance is infinite. [18]

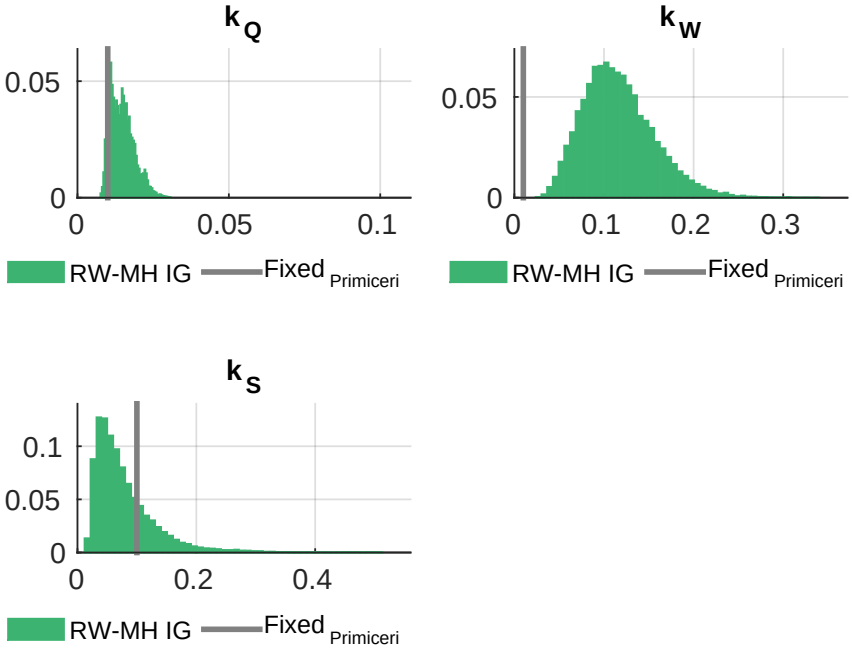

Figure 5: Marginal posterior distributions of the three hyperparameters along with common values used in the literature.

Figure 5 plots the marginal posteriors for the three hyperparameters (estimated using a random-walk Metropolis-Hasting algorithm with the inverse-Gamma prior) along with the values used in Primiceri (2005). We can see that there is substantial uncertainty surrounding the hyperparameters, which our approach takes into account. We also see that the

[18] As a robustness check, we will later show results for a prior with finite variance. Our choice of an inverse-gamma prior confronts the possible pile-up problem for the hyperparameters discussed by Stock & Watson (1996). We thus reduce the prior probability mass for the hyperparameters very close to zero. As seen in our Monte Carlo study (where we used the same priors), this does not preclude finding no time variation in parameters if indeed the data-generating process does not feature time-varying parameters. While the pile-up problem described in Stock & Watson (1996) was not an issue in our Monte Carlo exercise, we do suggest researchers use an informative prior to steer clear of any possibility of the pile-up problem appearing in real-world applications.

posterior modes are substantially different from the values used in Primiceri (2005). Primiceri estimated these values for a specific dataset. Our results hopefully caution the reader to not blindly use fixed values for other applications. As we saw in the case of our simulations, estimating the prior hyperparameters can have a substantial impact on parameter paths and, as a result, statistics derived from the estimated parameter and volatility paths such as impulse responses, forecasts, and variance decompositions. To illustrate this impact, we focus on the 'core inflation' statistic used by Cogley & Sargent (2005). Core inflation is defined here as the implied steady state value of inflation if the parameters of the VAR were fixed at each point in time at their posterior mean.[19] Thus, this measure, depicted in figure 6, can be interpreted as an estimate of trend inflation. Figure 6 shows that core inflation based on estimated hyperparameters, while still substantially smoother than the data, is not as smooth as the estimate using the fixed hyperparameter values of Primiceri (2005). To give one example, core inflation for the UK peaks around 10 % (in annual terms) with our approach, whereas it is only 6 % for the fixed hyperparameter-based estimates. The differences are especially pronounced during the 1930s and the 1970s (periods where inflation was large in absolute value). The greater amount of time variation in parameters implied by the posterior distribution shown before clearly has an impact on the core inflation statistic.

As a robustness check, we re-estimate our model using a different prior. For simplicity, we only change the prior for k_Q. We keep the prior mode fixed at the same value as before, but now use parameter values that imply a finite variance, namely .01 (whereas the prior variance was

[19] We checked that the VAR based on time t posterior mean parameter values is stationary for all t so that this calculation makes sense.

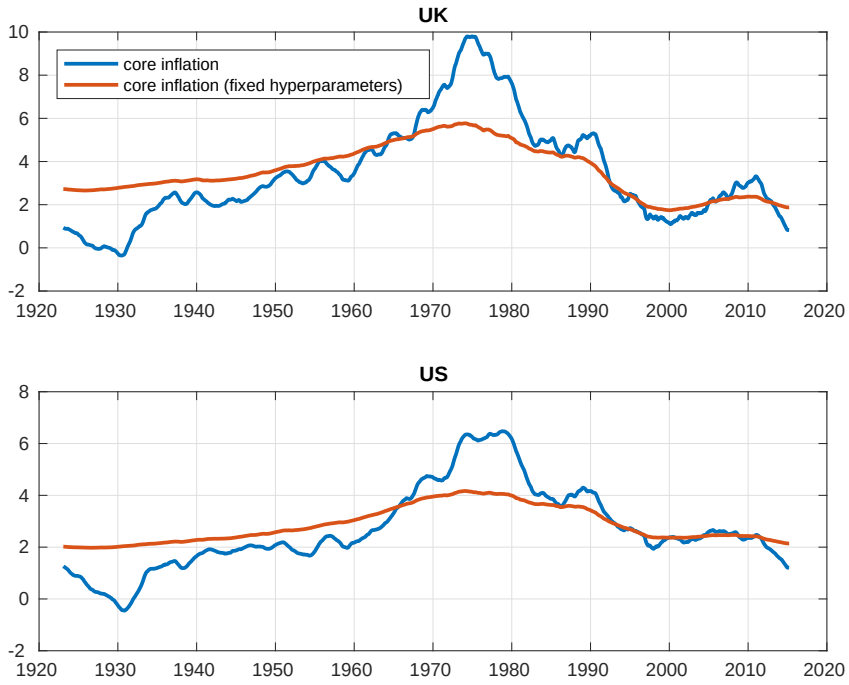

Figure 6: Core inflation for the US and UK, using either estimated hyperparameters or the values from Primiceri.

infinite before). Figure 7 shows that the results are indeed robust to these changes in the prior.

6 Multivariate Scaling Parameters

Like in the case of intercepts of our VAR or parameters associated with one given equation of the VAR, some groups of parameters or volatilities might vary at a different rate than other parameters. In this section we show how to incorporate this idea into our framework. Benati (2015) also estimates different scaling parameters for different equations in his VAR.

We denote by $\overline{k_x}$ vectors of scaling parameters of dimension d_x, where

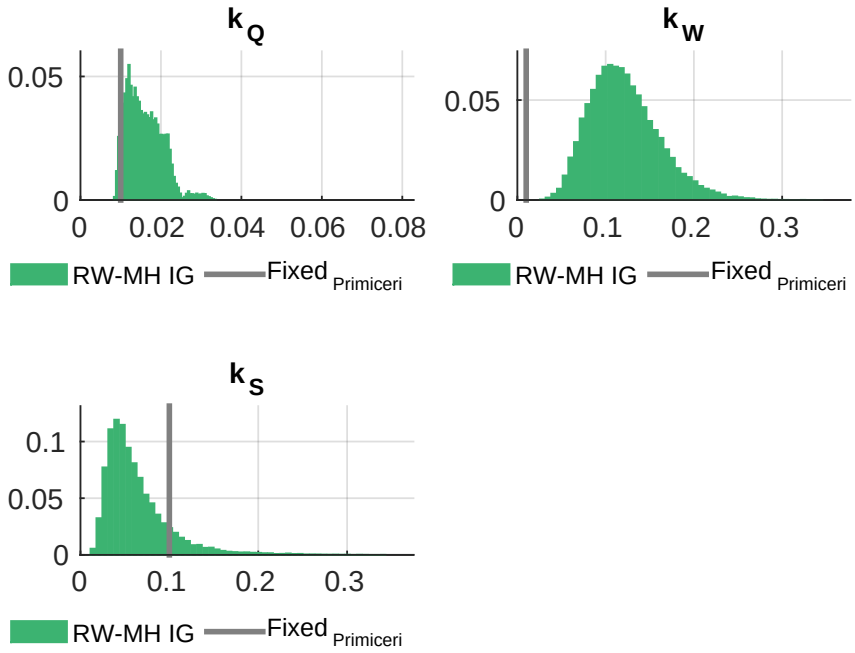

Figure 7: Marginal posterior distributions of the three hyperparameters along with common values used in the literature, prior with finite variance

matrix x is of dimension d_x by d_x. We then assume the following forms for the priors of the matrices Q, W, and S_j:

$$Q \sim IW(diag(\overline{k_Q})df_Q V_Q diag(\overline{k_Q}), df_Q) \tag{22}$$

$$W \sim IW(diag(\overline{k_W})df_W V_W diag(\overline{k_W}), df_W) \tag{23}$$

$$S_j \sim IW(diag(\overline{k_Sj})df_{Sj} V_{Sj} diag(\overline{k_Sj}), df_{Sj}) \tag{24}$$

where $diag$ is an operator that turns a d by 1 dimensional vector into a d by d dimensional diagonal matrix with the elements of the vector on the main diagonal. In practice, estimating one k scaling parameter per coefficient/volatility is not feasible for VARs of the size commonly used in applications because of the large number of coefficients that would have to be estimated. Instead, we propose to group parameters into a

relatively small number of groups and use one k scaling parameter per block of parameters. As mentioned before, natural choices for blocks in the case of the θ coefficients could be intercepts vs. all other parameters or a grouping of θ coefficients by equation. We would then augment our description of the algorithm with a deterministic mapping from the relatively small number of scaling parameters (which we call $\breve{k_x}$) to $\overline{k_x}$.

In terms of the estimation algorithm, nothing of substance changes:[20] in the proposal step, the proposal density is now multivariate normal and in the calculation of the acceptance probability we have to adjust the evaluation of $p(Q|k_Q)$ to take into account the updated form of the density (see equation (22)) and the fact that the prior $p(k_Q^*)$ is now multivariate. One could use independent priors for each element of $\breve{k_Q}$, for example. The rest of the Gibbs sampling steps for other parameters are unaffected, with the exception of the step where Q: the scaling matrix for the inverse-Wishart density needs to be updated as described in equation (22).

To give an example that this extension can make a substantial difference, we re-estimate our VAR described in the previous section, allowing for two different values of k_Q - one for the intercepts and one for all other parameters. We still use one hyperparameter each for W and $\{S_j\}_{j=1}^J$. For the sake of brevity, we focus here on results using our infinite variance specification for all hyperparameters (results using the finite variance prior are similar, just as before). Figure 8 plots the marginal posteriors of the hyperparameters. We can see an interesting pattern: The posterior distribution for the k_Q value associated with the intercepts is similar to that obtained for all parameters in the previous section, but the k_Q parameters for all other parameters is an order of magnitude larger,

[20] The changes are the same for all blocks. We describe it here for the case of Q.

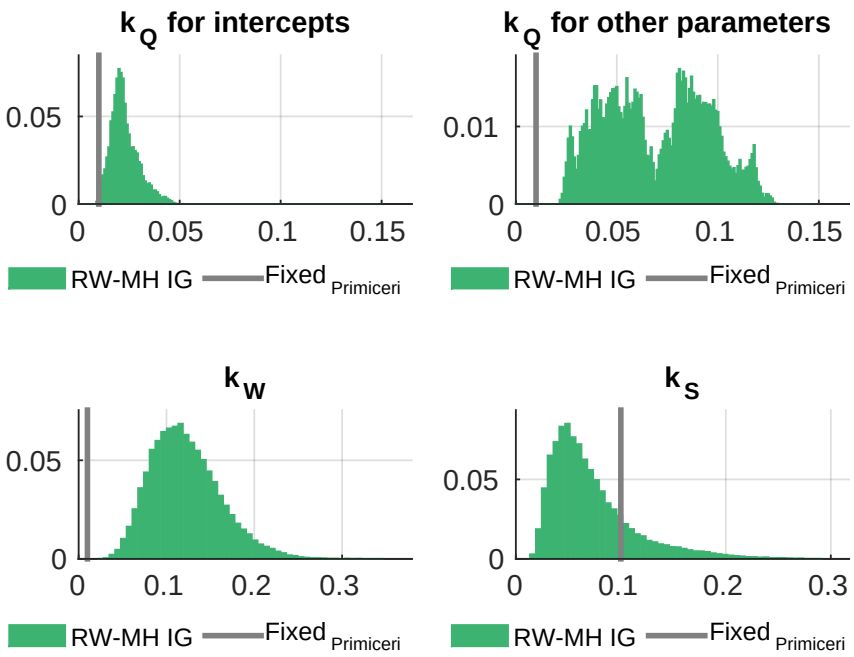

Figure 8: Marginal posterior distributions of the four hyperparameters along with common values used in the literature.

showing that substantial time variation can be hidden by assuming an overly tight prior structure. The posteriors for the hyperparameters are basically unchanged.

7 Conclusion

The choice of prior hyperparameters in large multivariate time series models is a daunting task. Because of the large dimensionality, using introspection to obtain a prior is very dificult. Thus, many researchers have turned to automated or semi-automated prior choices that depend only on few hyperparameters. Since those hyperparameters influence the prior distribution of large dimensional objects, their choice can be crucial. We argue that, considering the number of hyperparameters is usually relatively small, researchers should consider estimating them. This is

especially relevant because, as we show in this paper, this estimation can be carried out with only minor changes in existing codes and at negligible computational cost (because the densities that need to be evaluated in the additional estimation step are prior distributions that are usually fast to evaluate).

We show that estimating these hyperparameters in a popular class of models in empirical macroeconomics can drastically improve estimates.

A The Metropolis-within-Gibbs step implementation in more detail

In this section, we present details on the algorithm for the posterior sampling of the scaling parameters in the VAR with time-varying parameters and stochastic volatility. For the sake of brevity, we describe the sampling procedure for a generic scaling factor $k_X,\ X \in \{Q, S, W\}$. Given a draw for X, the conditional posterior $p(k_X|X) \propto p(X|k_X)p(k_X)$ can be obtained with a Metropolis-Hastings step. We use a version of the (Gaussian) random walk Metropolis-Hastings algorithm with an automatic tuning step for the proposal variance in a burn-in phase. The algorithm is initialized with values k_X^0 (which we choose to be the values from Primiceri (2005)) and $\sigma_{k_X}^2$, which we change in a preliminary burn-in phase to achieve a target acceptance rate.

1. At step i, take a candidate draw k_X^* from $N(k_X^{i-1}, \sigma_{k_X}^2)$
2. Calculate the acceptance probability $\alpha_{k_X}^i = \min\left(1, \frac{p(X|k_X^*)p(k_X^*)}{p(X|k_X^{i-1})p(k_X^{i-1})}\right)$
3. Accept the candidate draw by setting $k_X^i = k_X^*$ with probability $\alpha_{k_X}^i$. Otherwise set $k_X^i = k_X^{i-1}$.
4. Calculate the average acceptance ratio $\bar{\alpha}_{k_X}$. Adjust the increment standard deviation σ_{k_X} every qth iteration according to $\sigma_{k_X}^{New} = \sigma_{k_X}\frac{\bar{\alpha}_{k_X}}{\alpha^*}$, where α^* denotes the target average acceptance ratio. Do not adjust after the iteration i exceeds the burn-in threshold I.

In practice, we set $\alpha^* = .5$ and the burn-in threshold I equal to one-half of the total repetition number.

B The algorithm for a fixed-coefficient VAR

Fixed coefficient VARs are often estimated using the Gibbs sampler (see Koop & Korobilis (2010)). A fixed coefficient Gaussian VAR is of the form:

$$y_t = \mu + \sum_{j=1}^{L} A_j y_{t-j} + e_t \tag{25}$$

with $e_t \sim_{iid} N(0, \Sigma)$.

If we define $\beta \equiv [\mu' \; vec(A_1)' \ldots vec(A_L)']'$, the most commonly used Gibbs sampler assumes that

$$\beta \sim N(\overline{\beta}(\phi), V_\beta(\phi)) \tag{26}$$

$$\Sigma \sim IW(S, df) \tag{27}$$

where we have made the dependence of the prior for β on hyperparameters ϕ explicit. Note that the priors on β and Σ are assumed independent and are thus not natural conjugate priors. We could also introduce additional hyperparameters for the prior on Σ, but since popular priors such as the Minnesota prior focus on β, we will do the same here. A Gibbs sampler for this model consists of the following three steps:

1. Draw $\beta|\Sigma, \phi$
2. Draw $\Sigma|\beta, \phi$ - since this step conditions on β, it simplifies to drawing Σ conditional only on β since ϕ does not carry any additional information about Σ once we condition on β
3. Draw $\phi|\beta, \Sigma$. As discussed in this paper, this simplifies to drawing $\phi|\beta$

The first two steps of the Gibbs sampler are standard in the literature (see again Koop & Korobilis (2010)), except that we have to possibly change ϕ at every iteration when drawing β. The last step is described in detail in this paper.

C The complete algorithm

In this section, we describe the complete algorithm to estimate the TVP-VAR model with stochastic volatility described in the main text. We modify the algorithm described in Del Negro & Primiceri (2015) to include additional steps for the drawing of the hyperparameters.
The algorithm proceeds as follows[21]

1. Draw Σ^T from $p(\Sigma^T|y^T, \theta^T, \Lambda^T, V, s^T, k_Q, k_S, k_W)$. This step requires us to generate draws from a nonlinear state space system. We use the approach by Kim et al. (1998) to approximate draws from the desired distribution. For a correct posterior sampling of the stochastic volatilities, we follow the corrigendum in Del Negro and Primiceri (2013) and the modified steps therein.

2. Draw θ^T from $p(\theta^T|y^T, \Lambda^T, \Sigma^T, V, k_Q, k_S, k_W)$. Conditional on all other parameter blocks, equations (4) and (5) from the main text form a linear Gaussian state space system. This step can be carried out using the simulation smoother detailed in Carter & Kohn (1994).

3. Draw Λ^T from $p(\Lambda^T|y^T, \theta^T, \Sigma^T, V, k_Q, k_S, k_W)$. Again, we draw these covariance states based on the simulation smoother of the previous step, exploiting our assumption that the covariance matrix of

[21] A superscript T denotes a sample of the relevant variable from $t = 1$ to T.

the innovations in the law of motion for the λ coefficients is block diagonal. This assumption follows Primiceri (2005), where further details on this step can be found.

4. Draw W, Q and S_j. Given our distributional assumptions, these conditional posteriors of the time-invariant variances follow inverse-Wishart distributions (which are functions of k_Q, k_S, k_W).

5. Draw s^T, the sequence of indicators for the mixture of normals needed for the Kim et al. (1998) stochastic volatility algorithm.

6. Draw k_Q, k_S, k_W. Each of these scaling parameters is drawn via the algorithm described in section B of the appendix.

References

Ahmadi, P. A., Matthes, C. & Wang, M.-C. (2016), 'Drifts and volatilities under measurement error: Assessing monetary policy shocks over the last century', *Quantitative Economics (forthcoming)* .

Banbura, M., Giannone, D. & Reichlin, L. (2010), 'Large Bayesian vector auto regressions', *Journal of Applied Econometrics* **25**(1), 71–92.

Baumeister, C., Liu, P. & Mumtaz, H. (2013), 'Changes in the effects of monetary policy on disaggregate price dynamics', *Journal of Economic Dynamics and Control* **37**(3), 543–560.

Benati, L. (2015), How Fast Can Advanced Economies Grow?, Technical report.

Benati, L. & Lubik, T. (2014), 'Sales, inventories, and real interest rates: A century of stylized facts', *Journal of Applied Econometrics* p. forthcoming.

Canova, F. & Gambetti, L. (2009), 'Structural changes in the US economy: Is there a role for monetary policy?', *Journal of Economic Dynamics and Control* **33**(2), 477–490.

Carter, C. K. & Kohn, R. (1994), 'On gibbs sampling for state space models', *Biometrika* **81**(3), 541–553.

Chib, S., Nardari, F. & Shephard, N. (2006), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **134**(2), 341–371.

Clark, T. E. & Terry, S. J. (2010), 'Time Variation in the Inflation Passthrough of Energy Prices', *Journal of Money, Credit and Banking* **42**(7), 1419–1433.

Cogley, T. & Sargent, T. A. (2015), 'Measuring price-level uncertainty and instability in the us, 1850-2012', *Review of Economics & Statistics* **97**(4), 827–838.

Cogley, T. & Sargent, T. J. (2005), 'Drift and volatilities: Monetary policies and outcomes in the post WWII U.S.', *Review of Economic Dynamics* **8**(2), 262–302.

Cooley, T. F. & Prescott, E. C. (1976), 'Estimation in the Presence of Stochastic Parameter Variation', *Econometrica* **44**(1), 167–84.

D'Agostino, A. & Surico, P. (2011), A century of inflation forecasts, CEPR discussion papers, C.E.P.R. Discussion Papers.

Del Negro, M. & Primiceri, G. (2015), 'Time-varying structural vector autoregressions and monetary policy: a corrigendum', *Review of Economic Studies, forthcoming* .

Diebold, F. X., Schorfheide, F. & Shin, M. (2016), Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility, Technical report.

Doan, T., Litterman, R. B. & Sims, C. A. (1984), 'Forecasting and conditional projection using realistic prior distribution', *Econometric Reviews* **31**(3), 1–100.

Ellis, C., Mumtaz, H. & Zabczyk, P. (2014), 'What Lies Beneath? A Time-varying FAVAR Model for the UK Transmission Mechanism', *Economic Journal* (576), 668–699.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian Data Analysis*, Chapman & Hall.

Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*, Wiley.

Giannone, D., Lenza, M. & Primiceri, G. E. (2013), 'Prior Selection for Vector Autoregressions', *Review of Economics and Statistics, forthcoming* .

Justiniano, A. & Primiceri, G. E. (2008), 'The Time-Varying Volatility of Macroeconomic Fluctuations', *American Economic Review* **98**(3), 604–41.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**(3), 361–93.

Kliem, M., Kriwoluzky, A. & Sarferaz, S. (2013), On the low-frequency relationship between public deficits and inflation, Technical report.

Koop, G. & Korobilis, D. (2010), Bayesian multivariate time series methods for empirical macroeconomics, Working paper, University of Strathclyde.

Koop, G., Leon-Gonzalez, R. & Strachan, R. W. (2011), 'Bayesian inference in a time varying cointegration model', *Journal of Econometrics* **165**(2), 210–220.

Korobilis, D. (2014), Data-based priors for vector autoregressions with drifting coefficients, MPRA Paper 53772, University Library of Munich, Germany.

Lopes, H. F., Moreira, A. R. B. & Schmidt, A. M. (1999), 'Hyperparameter estimation in forecast models', *Computational Statistics & Data Analysis* **29**(4), 387–410.

Lubik, T. A., Matthes, C. & Owens, A. (2014), Beveridge Curve Shifts and Time-Varying Parameter VARs, Technical report.

Primiceri, G. (2005), 'Time varying structural vector autoregressions and monetary policy', *Review of Economic Studies* **72**(3), 821–852.

Reusens, P. & Croux, C. (2015), Detecting time variation in the price puzzle: An improved prior choice for time varying parameter VAR models, Technical report, KU Leuven.

Sargent, T. J. & Surico, P. (2011), 'Two illustrations of the quantity theory of money: Breakdowns and revivals', *American Economic Review* **101**(1), 109–28.

Schorfheide, F., Song, D. & Yaron, A. (2014), Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach, Nber working papers, National Bureau of Economic Research, Inc.

Sims, C. A. (1980), 'Macroeconomics and Reality', *Econometrica* **48**(1), 1–48.

Sims, C. A. & Zha, T. (2006), 'Were there regime switches in U.S. monetary policy?', *American Economic Review* **96**(1), 54–81.

Smets, F. & Wouters, R. (2007), 'Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach', *American Economic Review* **97**(3), 586–606.

Stock, J. H. & Watson, M. W. (1996), 'Evidence on Structural Instability in Macroeconomic Time Series Relations', *Journal of Business & Economic Statistics* **14**(1), 11–30.

Stock, J. & Watson, M. (2002), 'Forecasting Using Principal Components From a Large Number of Predictors', *Journal of the American Statistical Association* **97**, 1167–1179.