Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian VARs?

Martin Feldkircher | OeNB
Florian Huber | WU
Gregor Kastner | WU

StatMath Brown Bag Seminar
March 14, 2018

Introduction

- Time-varying parameter VARs are rapidly gaining popularity in macroeconomics and finance (Primiceri, 2005; Cogley and Sargent, 2005)
- However, typical application is small dimensional, consisting of three to five endogenous variables in the VAR
- Model specification: High dimensional state space with a plethora of potential specification issues
  - Inclusion / Exclusion of a given covariate within a given equation
  - Constancy of regression parameters of a given covariate within a given equation
- Literature offers several solutions:
  1. Select whether a given coefficient is constant or time-varying over the full sample (Frühwirth-Schnatter and Wagner, 2010; Eisenstat et al., 2016; Bitto and Frühwirth-Schnatter, 2016; Huber and Feldkircher, forthcoming)
  2. Assume that parameters change only during specific points in time (Sims and Zha, 2006; Koop and Potter, 2007; Giordani and Kohn, 2008; Koop et al., 2009; Giordani and Villani, 2010; Maheu and Song, forthcoming)
  3. Hybrid approaches / time-varying shrinkage (Kalli and Griffin, 2014; Kalli and Griffin, 2018; Huber et al., 2018)

Motivation and research objectives

Typical problems:

		Information Set	
		small	sizeable
Model	simple	omitted variables + misspecification	misspecification
	sophisticated	omitted variables	overfitting

Agenda:

- Systematic assessment of model size and complexity for three macroeconomic datasets in three countries
- Show that in high-dimensional models, the detrimental impact of additional parameters in the model becomes too large $\rightarrow$ overfitting
- In small-dimensional models, sophisticated dynamics control for omitted variable issues, overruling the detrimental effect of the large number of additional parameters
- Provide a means that combines the "best of both worlds"

Econometric framework

- In this paper, the model of interest is a TVP-VAR with stochastic volatility (SV) in the spirit of Primiceri (2005)
- The model summarizes the joint dynamics of an M-dimensional zero-mean vector of macroeconomic time series $\{\boldsymbol{y}_t\}_{t=1}^{T}$ as follows:

$$\boldsymbol{y}_t = \boldsymbol{A}_{1t}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_{pt}\boldsymbol{y}_{t-p} + \varepsilon_t, \;\; \varepsilon_t \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_t).$$

- The $M \times M$ matrix $\boldsymbol{A}_{jt}$ $(j = 1, \ldots, p)$ contains time-varying autoregressive coefficients
- ε_t is a vector white noise error with zero mean and a time-varying variance-covariance matrix $\boldsymbol{\Sigma}_t = \boldsymbol{H}_t\boldsymbol{V}_t\boldsymbol{H}_t'$
- $\boldsymbol{H}_t$ is a lower unitriangular matrix and $\boldsymbol{V}_t = \text{diag}(e^{v_{1t}}, \ldots, e^{v_{Mt}})$ denotes a diagonal matrix with time-varying shock variances

Law of motion for the latent states

- We assume that coefficients in $\boldsymbol{a}_t = vec[(\boldsymbol{A}_{1t}, \ldots, \boldsymbol{A}_{pt})']$ follow a random walk

$$\boldsymbol{a}_t = \boldsymbol{a}_{t-1} + \boldsymbol{\xi}_t,$$
$$\boldsymbol{\xi}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_a),$$

$\boldsymbol{\Omega}_a = \text{diag}(\omega_1, \ldots, \omega_K)$ with $\dim(\boldsymbol{a}_t) = K = M^2 p$

- $M(M-1)/2$ free elements in $\boldsymbol{H}_t$ evolve according to

$$\boldsymbol{h}_t = \boldsymbol{h}_{t-1} + \boldsymbol{s}_t,$$
$$\boldsymbol{s}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_h)$$

- Log-volatilities follow

$$v_{jt} = \mu_j + \rho_j(v_{jt-1} - \mu_j) + w_{jt} \ \forall j$$

Issue #1: Overfitting

- Dimension of state vector:
  1. Dimension of $\boldsymbol{a}_t$ is $K = M^2 p$
  2. Dimension of free elements of $\boldsymbol{\Sigma}_t$ is $\frac{M(M-1)}{2} + M$

- BUT: we need to simulate $\{\boldsymbol{a}_t\}_{t=1}^T$ and $\{\boldsymbol{\Sigma}_t\}_{t=1}^T$

- T is typically small to moderate (i.e. around 200 for quarterly US time series, for EA data often much less)

- K can be large (i.e. $M = 15$ and $p = 2$ yields 302 coefficients per t)

- This calls for regularization: We follow a Bayesian approach based on two recent global-local shrinkage priors (Griffin and Brown, 2010; Bhattacharya et al., 2015)

Issue #2: Computational complexity

Rewrite the TVP-VAR-SV model as

$$\boldsymbol{y}_t = \underbrace{(\boldsymbol{I}_M \otimes \boldsymbol{x}_t')}_{\boldsymbol{Z}_t} \boldsymbol{a}_t + \varepsilon_t$$

To retrieve estimates for the state vector we have to employ the Kalman Filter (irrespective of whether we are Frequentists or Bayesians). The key recursions are given by

$$\begin{aligned} \boldsymbol{a}_{t|t-1} &= \boldsymbol{a}_{t-1|t-1} \\ \boldsymbol{S}_{t|t-1} &= \boldsymbol{S}_{t-1|t-1} + \boldsymbol{\Omega} \\ \boldsymbol{K}_t &= \boldsymbol{S}_{t|t-1}\boldsymbol{Z}_t'(\boldsymbol{Z}_t\boldsymbol{S}_{t|t-1}\boldsymbol{Z}_t' + \Sigma_t)^{-1} \\ \boldsymbol{a}_{t|t} &= \boldsymbol{a}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{Z}_t\boldsymbol{a}_{t|t-1}), \\ \boldsymbol{S}_{t|t} &= \boldsymbol{S}_{t|t-1} - \boldsymbol{K}_t\boldsymbol{H}_t\boldsymbol{S}_{t|t-1} \end{aligned}$$

with $\boldsymbol{S}_{t|s}$ being the conditional variance of $\boldsymbol{a}_t$ up to time s

Issue #2: Computational complexity

Note that the joint density of the states can be factorized as

$$p(\boldsymbol{a}_{1:T}|\bullet) = p(\boldsymbol{a}_T|\bullet) \prod_{t=1}^{T-1} p(\boldsymbol{a}_t|\boldsymbol{a}_{t+1}, \bullet)$$

The KF yields $p(\boldsymbol{a}_T|\bullet) \sim \mathcal{N}(\boldsymbol{a}_{T|T}, \boldsymbol{S}_{T|T})$

Following Frühwirth-Schnatter (1994) and Carter and Kohn (1994), sample the states from

$$p(\boldsymbol{a}_t|\bullet) \sim \mathcal{N}(\boldsymbol{a}_{t|t+1}, \boldsymbol{S}_{t|t+1}),$$

with

$$\begin{aligned} \boldsymbol{a}_{t|t+1} &= \boldsymbol{a}_{t|t} + \boldsymbol{S}_{t|t}\boldsymbol{S}_{t+1|t}^{-1}(\boldsymbol{a}_{t+1} - \boldsymbol{a}_{t|t}), \\ \boldsymbol{S}_{t|t+1} &= \boldsymbol{S}_{t|t} - \boldsymbol{S}_{t|t}\boldsymbol{S}_{t+1|t}^{-1}\boldsymbol{S}_{t|t} \end{aligned}$$

Problem: Lots of matrix operations!

Remedy #2: Triangularization

Reduced-form errors are given by

$$\boldsymbol{y}_t - \boldsymbol{A}_t \boldsymbol{x}_t = \boldsymbol{H}_t \boldsymbol{\eta}_t, \text{ with } \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_t),$$

and multiplying from the left with $\tilde{\boldsymbol{H}}_t := \boldsymbol{H}_t^{-1}$ yields

$$\tilde{\boldsymbol{H}}_t \boldsymbol{\varepsilon}_t = \boldsymbol{\eta}_t.$$

For further illustration, note that the first two equations of the system are given by

$$\begin{aligned} \varepsilon_{1t} &= \eta_{1t}, \\ \tilde{h}_{21,t}\varepsilon_{1t} + \varepsilon_{2t} &= \eta_{2t}, \end{aligned}$$

with $\tilde{h}_{21,t}$ denoting the second element of the first column of $\tilde{\boldsymbol{H}}_t$. The last equation can be rewritten as

$$y_{2t} = \boldsymbol{A}_{2\bullet,t} \boldsymbol{x}_t - \tilde{h}_{21,t}\varepsilon_{1t} + \eta_{2t},$$

where $\boldsymbol{A}_{i\bullet,t}$ denotes the ith row of $\boldsymbol{A}_t$

Remedy #2: Triangularization

The ith equation is given by

$$y_{it} = \boldsymbol{A}_{i\bullet,t}\boldsymbol{x}_t - \sum_{s=1}^{i-1} \tilde{h}_{is,t}\varepsilon_{st} + \eta_{it} = \boldsymbol{B}_{it}'\boldsymbol{z}_{it} + \eta_{it}.$$

> $\boldsymbol{z}_{it} = (\boldsymbol{x}_t', -\varepsilon_{1t}, \ldots, -\varepsilon_{i-1,t})'$ is a K_i-dimensional vector of explanatory variables
> $\boldsymbol{B}_{it} = (\boldsymbol{A}_{i\bullet}, \tilde{h}_{i1,t}, \ldots, \tilde{h}_{ii-1,t})'$ is a K_i-dimensional vector of regression coefficients

Instead of estimating a single large model with a K-dimensional state vector we estimate M conditional models with K_i-dimensional state vectors

Remarks:

> Reduces to Cholesky SV (Lopes et al., 2016) when the VAR part is dropped
> In contrast to Cholesky SV not embarrassingly parallel (unless approximated)
> In contrast to Carriero et al. (2016) we sample $\tilde{\boldsymbol{H}}_t$ and $\boldsymbol{A}_t$ in one block

The non-centered parameterization

> Following Frühwirth-Schnatter and Wagner (2010) we parameterize the dynamic regression model for equation i as follows

$$\begin{aligned} y_{it} &= \boldsymbol{B}_{i0}' \boldsymbol{z}_{it} + \tilde{\boldsymbol{B}}_{it}' \sqrt{\boldsymbol{\Omega}_i} \boldsymbol{z}_{it} + \eta_{it}, \\ \tilde{\boldsymbol{B}}_{it} &= \tilde{\boldsymbol{B}}_{it-1} + \boldsymbol{u}_{it}, \quad \text{with } \boldsymbol{u}_{it} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{K_i}) \end{aligned}$$

> $\boldsymbol{\Omega}_i$ is the $K_i \times K_i$ block of $\boldsymbol{\Omega} = diag(\boldsymbol{\Omega}_a, \boldsymbol{\Omega}_h)$ associated with the ith equation

> $\sqrt{\boldsymbol{\Omega}_i}$ denotes the matrix square root such that $\boldsymbol{\Omega}_i = \sqrt{\boldsymbol{\Omega}_i}\sqrt{\boldsymbol{\Omega}_i}$

> $\tilde{\boldsymbol{B}}_{it}$ has typical element j given by $\tilde{b}_{ij,t} = \frac{b_{ij,t} - b_{ij,0}}{\sqrt{\omega_{ij}}}$ and $\sqrt{\omega_{ij}} := [\sqrt{\boldsymbol{\Omega}_i}]_{jj}$

Remedy #1: Shrinkage

In what follows we let $\boldsymbol{a}_0 = \text{vec}(\boldsymbol{B}_0)$ denote the time-invariant part of the VAR coefficients with typical element a_{0j} for $j = 1, \ldots, K$

The corresponding signed squared root of the state innovation variance is consequently denoted by $\pm\sqrt{\omega}_j$ or simply $\sqrt{\omega}_j$

We introduce shrinkage by using two shrinkage priors, the Normal-Gamma shrinkage prior (Griffin and Brown, 2010; Bitto and Frühwirth-Schnatter, 2016) and the Dirichlet-Laplace shrinkage prior (Bhattacharya et al., 2015)

Normal-Gamma shrinkage prior

Our prior specification is a scale mixture of Gaussians,

$$a_{0j}|\tau_{aj}^2, \lambda_l \sim \mathcal{N}(0, 2/\lambda_l\ \tau_{aj}^2), \quad \tau_{aj}^2 \sim \mathcal{G}(\vartheta_l, \vartheta_l)$$
$$\sqrt{\omega}_j|\tau_{\omega j}^2, \lambda_l \sim \mathcal{N}(0, 2/\lambda_l\ \tau_{\omega j}^2), \quad \tau_{\omega j}^2 \sim \mathcal{G}(\vartheta_l, \vartheta_l)$$
$$\lambda_l = \prod_{s=1}^{l} \nu_s,\ \nu_s \sim \mathcal{G}(c_\lambda, d_\lambda)$$

- τ_{aj}^2 and $\tau_{\omega j}^2$ denote a set of local scaling parameters that follow a Gamma distribution
- λ_l is a lag-specific shrinkage parameter

Qualitatively same prior setup on the covariance parameters $\tilde{h}_{ij,0}$ and their process standard deviations

The Dirichlet-Laplace shrinkage prior

The DL prior is, again, based on a scale mixture of Gaussians,

$$a_{0j}|\psi_{aj}, \xi_{aj}^2, \tilde{\lambda}_l \sim \mathcal{N}(0, \psi_{aj}\xi_{aj}^2/\tilde{\lambda}_l^2), \quad \psi_{aj} \sim Exp(1/2), \quad \xi_j \sim Dir(n_a, \dots, n_a)$$
$$\sqrt{\omega}_j|\psi_{\omega j}, \xi_{\omega j}^2, \tilde{\lambda}_l \sim \mathcal{N}(0, \psi_{\omega j}\xi_{\omega j}^2/\tilde{\lambda}_l^2), \quad \psi_{\omega j} \sim Exp(1/2), \quad \xi_j \sim Dir(n_a, \dots, n_a)$$
$$\tilde{\lambda}_l = \prod_{s=1}^{l} \tilde{\nu}_s, \;\; \tilde{\nu}_s \sim \mathcal{G}(c_\lambda, d_\lambda)$$

- $s \in \{a, \omega\}$, ψ_{sj} is again a set of local scaling parameters
- ξ_{sj} constitutes an auxiliary scaling parameter defined on the $(K-1)$-dimensional unit simplex $\mathcal{S}^{K-1} = \{\boldsymbol{x} = (x_1, \dots, x_K)' : x_j \geq 0, \sum_{j=1}^{K} x_j = 1\}$ with $\boldsymbol{\xi}_s = (\xi_{s1}, \dots, \xi_{sK})'$
- The lag-specific shrinkage parameter $\tilde{\lambda}_l$ is defined analogously to the NG prior

Estimation and inference

1. Draw $(\boldsymbol{B}_{i0}', \omega_{i1}, \ldots, \omega_{iK_i})'$ for $i = 1, \ldots, M$ from $\mathcal{N}(\boldsymbol{\mu}_{Bi}, \boldsymbol{V}_i)$ with $\boldsymbol{V}_i = (\boldsymbol{Z}_i'\boldsymbol{Z}_i + \underline{\boldsymbol{V}}_i^{-1})^{-1}$ and $\boldsymbol{\mu}_{Bi} = \boldsymbol{V}_i(\boldsymbol{Z}_i\boldsymbol{Y}_i)$. We let $\boldsymbol{Z}_i$ be a $T \times (2K_i)$ matrix with typical tth row $[\boldsymbol{z}_{it}', (\boldsymbol{B}_{it} \odot \boldsymbol{z}_{it})'] \ e^{-(v_{it}/2)}$, $\boldsymbol{Y}_i$ is a T-dimensional vector with element $y_{it} \ e^{-(v_{it}/2)}$, and $\underline{\boldsymbol{V}}_i$ is a prior covariance matrix that depends on the prior specification adopted

2. Simulate the full history of $\{\tilde{\boldsymbol{B}}_{it}\}_{t=1}^{T}$ by means of a forward filtering backward sampling algorithm (see Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) per equation.

3. The log-volatilities and the corresponding parameters of the state equation are simulated using the algorithm put forward in Kastner and Frühwirth-Schnatter (2014) via the R package stochvol (Kastner, 2016)

4. Depending on the prior specification adopted, draw the parameters used to construct $\underline{\boldsymbol{V}}_i$ using the conditional posterior distributions (next two slides)

Full conditionals: NG prior

- Scaling parameters of time-invariant VAR coefficients and process innovation variances,

$$\tau_{aj}^2|\bullet \sim \mathcal{GIG}(\vartheta_l - 1/2, a_{0j}^2, \vartheta_l\lambda_l), \quad \tau_{\omega j}^2|\bullet \sim \mathcal{GIG}(\vartheta_l - 1/2, \omega_j^2, \vartheta_l\lambda_l),$$

- Sampling ν_l by combining each component of $p(\tau_{aj}^2, \tau_{\omega j}^2|\nu_l, \lambda_{l-1}) = p(\tau_{aj}^2|\nu_l, \lambda_{l-1}) \times p(\tau_{\omega j}^2|\nu_l, \lambda_{l-1})$ with Gamma prior

$$\nu_l|\lambda_{l-1}, \bullet \sim \mathcal{G}\left\{c_\lambda + 2\vartheta_l M^2, d_\lambda + \lambda_{l-1}\frac{\vartheta_l}{2}\sum_{j\in\mathcal{A}_l}(\tau_{aj}^2 + \tau_{\omega j}^2)\right\}.$$

  with $\lambda_0 = 1$
- Full conditionals of local and global scalings for elements of the VC matrix look similar

Full conditionals: DL prior

- Local scalings again follow a GIG distribution,

$$\psi_{aj}|\bullet \sim \mathcal{GIG}(1/2, |a_{j0}|\tilde{\lambda}_l/\xi_{aj}, 1), \quad \psi_{\omega j}|\bullet \sim \mathcal{GIG}(1/2, |\sqrt{\omega}_j|\tilde{\lambda}_l/\xi_{\omega j}, 1).$$

- For the Dirichlet components, the conditional posterior distribution is obtained by sampling a set of K auxiliary variables $N_{aj}, N_{\omega j}$ $(j = 1, \ldots, K)$,

$$N_{aj}|\bullet \sim \mathcal{GIG}(n_a - 1, 2|a_{j0}|, 1), \quad N_{\omega j}|\bullet \sim \mathcal{GIG}(n_a - 1, 2|\sqrt{\omega}_{j0}|, 1),$$

  We set $\xi_{aj} = N_{aj}/N_a$ and $\xi_{\omega j} = N_{\omega j}/N_\omega$ with $N_a = \sum_{j=1}^{K} N_{aj}$ and $N_\omega = \sum_{j=1}^{K} N_{\omega j}$

- Lag-wise shrinkage parameters come from a Gamma distribution

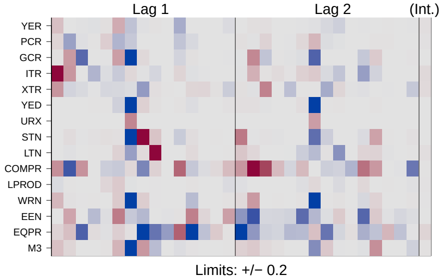
$$p(\tilde{\nu}_l|\tilde{\lambda}_{l-1}, \bullet) \sim \mathcal{G}\left\{c_\lambda + 2M^2, d_\lambda + \tilde{\lambda}_{l-1} \sum_{j \in \mathcal{A}_l} \left(\frac{|a_{0j}|}{\xi_{aj}} + \frac{|\sqrt{\omega}_j|}{\xi_{\omega j}} \right) \right\}$$

  with $\tilde{\lambda}_0 = 1$

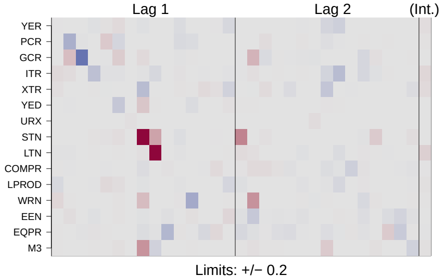
Data and model specification

- We use prominent macroeconomic data sets for the EA, the UK and the US
  - For the euro area we take data from the area wide model (Fagan et al., 2001) and additionally include equity prices available from 1987Q1 to 2015Q4
  - UK data stem from the Bank of England's "A millenium of macroeconomic data" (Thomas et al., 2010) and covers the period from 1982Q2 to 2016Q4
  - For the US, we use a subset from the FRED QD data base (McCracken and Ng, 2016) which covers the period from 1959Q1 to 2015Q1
- For each country, we consider three model sizes
  - small: real activity, prices and short-term interest rates ($M = 3$)
  - medium: small + investment, consumption, unemployment rate and either nominal or real effective exchange rates ($M = 7$)
  - large: medium+ wages, money (M2 or M3), government consumption, exports, equity prices and 10-year government bond yields + various indicators ($M = 15$)
- NG prior: $\vartheta = \vartheta_h = 0.1$, $c_\lambda = 1.5, d_\lambda = 1$ and $c_\varpi = d_\varpi = 0.01$ to induce heavy shrinkage on the covariance parameters
- DL prior: c_λ and d_λ are specified as in the NG case and $n_a = 1/K, n_h = 1/v$

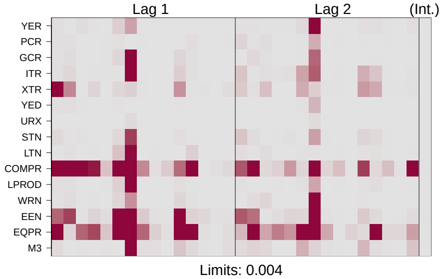
Posterior sparsity: Euro area



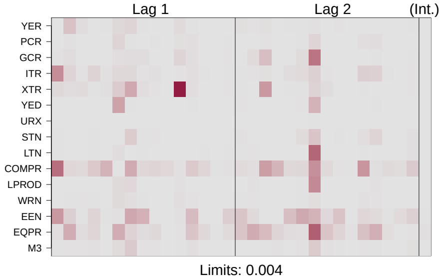
VAR coefficients: DL prior



VAR coefficients: NG prior

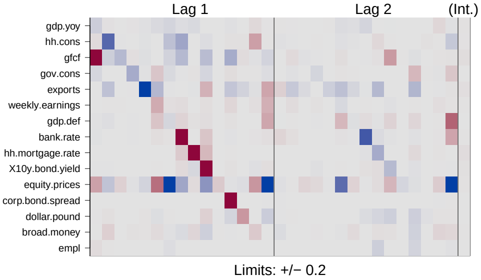


State innovation sds: DL prior

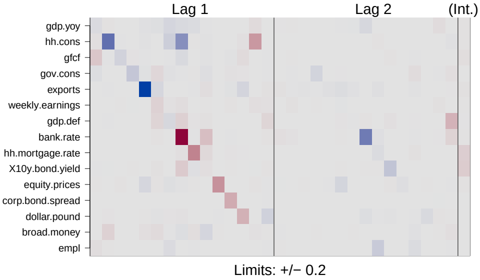


State innovation sds: NG prior

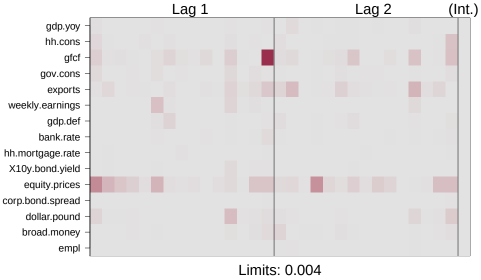
Posterior sparsity: United Kingdom



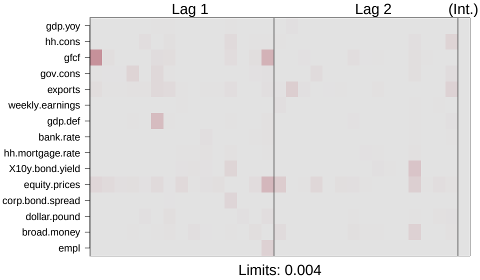
VAR coefficients: DL prior



VAR coefficients: NG prior

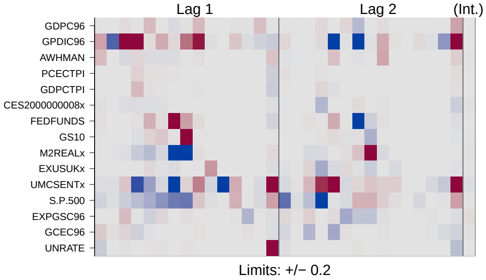


State innovation sds: DL prior

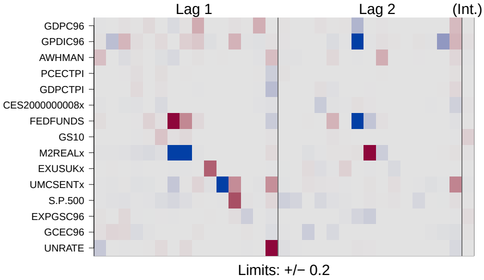


State innovation sds: NG prior

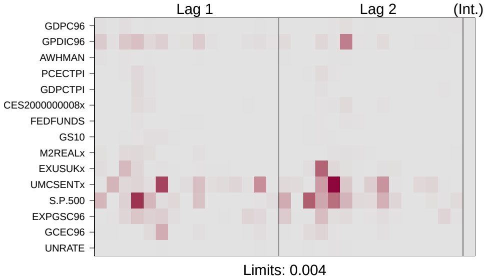
Posterior sparsity: United States



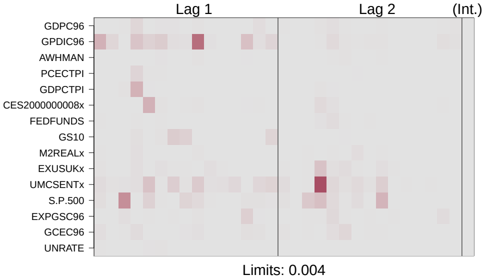
VAR coefficients: DL prior



VAR coefficients: NG prior



State innovation sds: DL prior



State innovation sds: NG prior

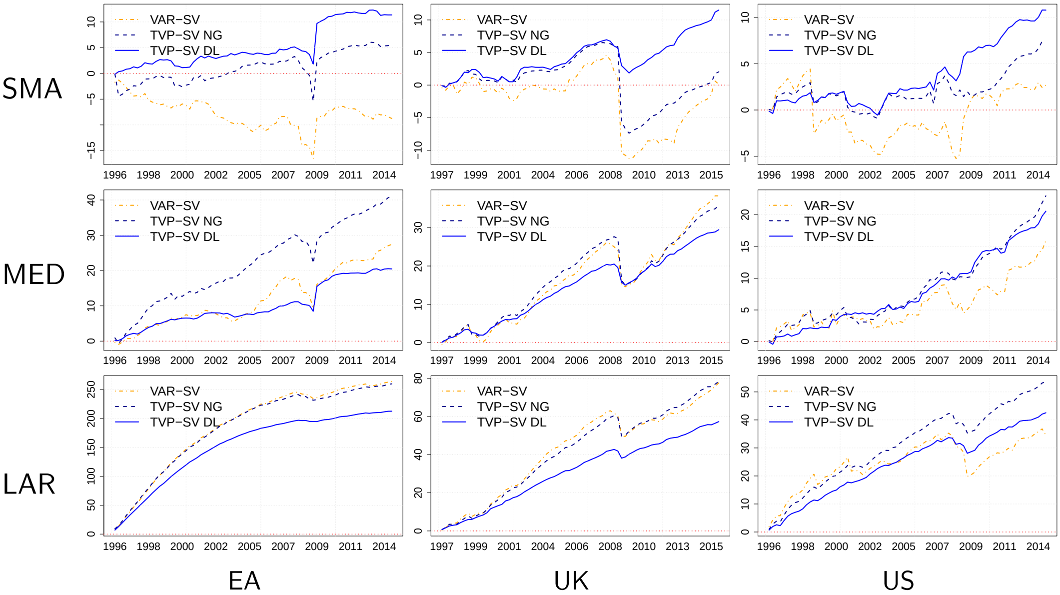
Forecasting exercise: Setup

- We use an expanding window and a hold-out sample of 20 years (80 quarters) which results into the following hold out samples
  - 1995Q4 to 2015Q3 for the EA
  - 1997Q1 to 2016Q4 for the UK
  - 1995Q4 to 2015Q3 for the US
- Forecasts are evaluated using log predictive scores (LPSs), a widely used metric to measure density forecast accuracy (see e.g. Geweke and Amisano, 2010)
- Benchmarks: constant parameter Bayesian VAR (BVAR-SV) and a time-varying parameter VAR with a loose prior setting (TVP-SV) as a general benchmark.
- TVP-SV features a prior on $\sqrt{\omega_j}$ is given by

$$\omega_j \sim \mathcal{G}(1/2, 1/2) \Leftrightarrow \pm\sqrt{\omega_j} \sim \mathcal{N}(0, 1).$$

  On $\boldsymbol{a}_0$ and for the BVAR-SV we use the NG shrinkage prior
- We use the joint density over inflation, output and interest rates for all model sizes

Cumulative log predictive Bayes factors over the loose TVP-VAR-SV



Dynamically selecting models

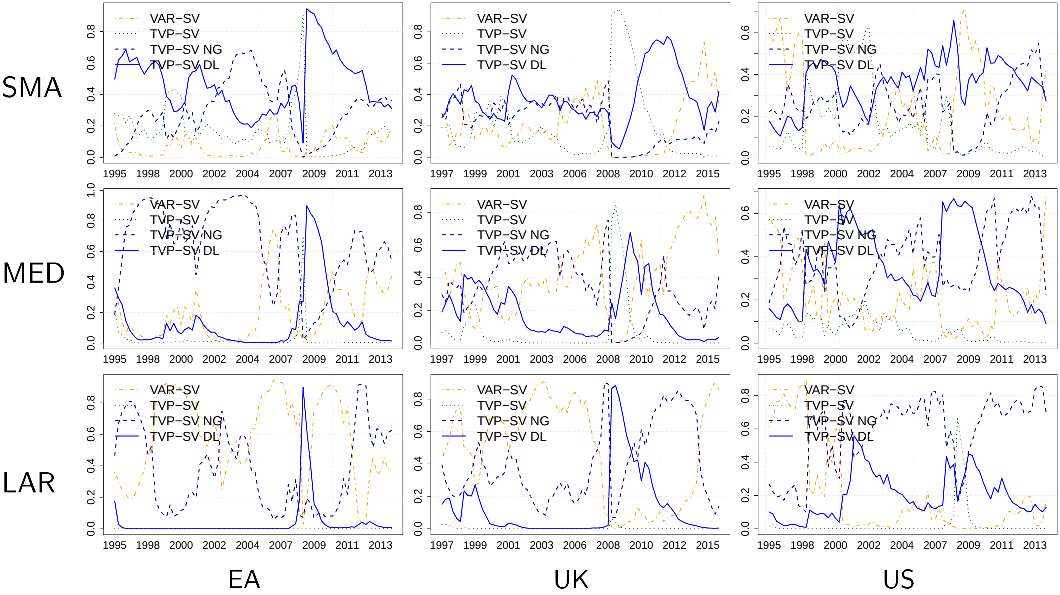
- We perform dynamic model selection by computing a set of weights for each model within a given model size and pick the model that yields the highest weight across all models (Raftery et al., 2010; Koop and Korobilis, 2012)
- The predicted weight associated with model i is computed as follows

$$\mathfrak{w}_{t|t-1,i} := \frac{\mathfrak{w}^{\alpha}_{t-1|t-1,i}}{\sum_{i\in\mathcal{M}} \mathfrak{w}^{\alpha}_{t-1|t-1,i}},$$

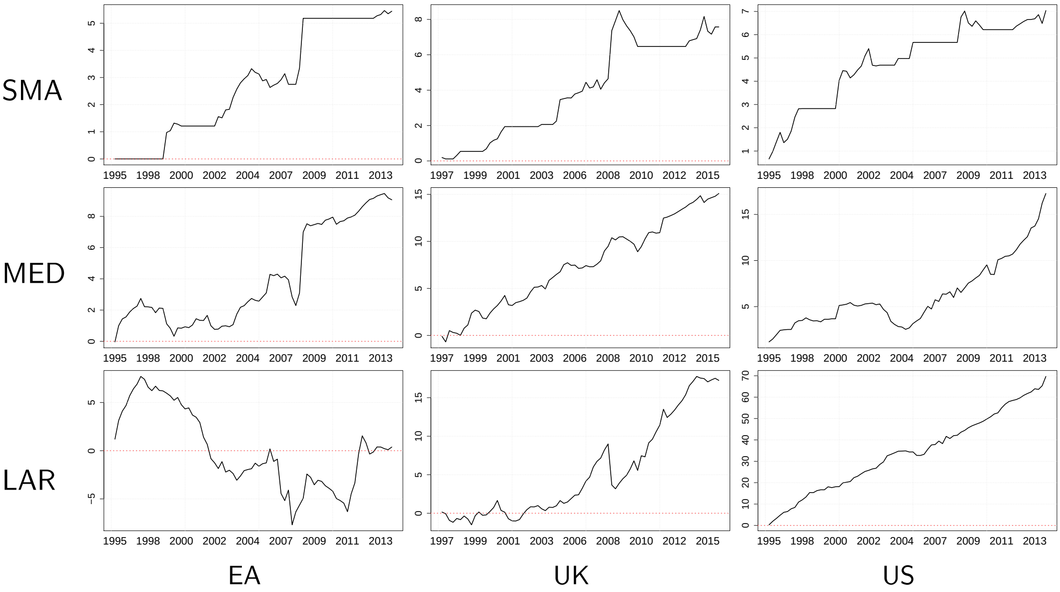
$$\mathfrak{w}_{t-1|t-1,i} = \frac{\mathfrak{w}_{t-1|t-2,i}\, p_{t-1|t-2,i}}{\sum_{i\in\mathcal{M}} \mathfrak{w}_{t-1|t-2,i}\, p_{t-1|t-2,i}}.$$

- $\alpha = 0.99$ denotes a forgetting factor close to unity
- $p_{t-1|t-2,i}$ denotes the one-step-ahead predictive likelihood for the three focus variables in $t-1$ for model i
- The initial weights $\mathfrak{w}_{t_0+1|t_0,i}$ are assumed to be equal for each model.

Model weights over time



Cumulative log predictive Bayes factors over the best individual model



Conclusive remarks

- Two global-local shrinkage priors aim to facilitate the usage of large TVP-VAR models, effectively controlling for overparameterization
- Empirical evidence that in small models, allowing for time-varying parameters pays off significantly
- In larger models, this effect is not as pronounced (especially without significant shrinkage)
- Strong heterogeneity with respect to model choice over time (i.e. crisis/non-crisis episodes)
- Dynamic model selection further improves predictions

References I

Bhattacharya, A. et al. (2015). "Dirichlet–Laplace priors for optimal shrinkage". *Journal of the American Statistical Association* 110, pp. 1479–1490 (cit. on pp. 6, 12).

Bitto, A. and S. Frühwirth-Schnatter (2016). *Achieving shrinkage in a time-varying parameter model framework*. Tech. rep. arXiv:1611.01310 (cit. on pp. 2, 12).

Carriero, A., T. E. Clark, and M. G. Marcellino (2016). "Large vector autoregressions with stochastic volatility and flexible priors". *FRB of Cleveland Working Paper* (cit. on p. 10).

Carter, C. K. and R. Kohn (1994). "On Gibbs sampling for state space models". *Biometrika* 81, pp. 541–553 (cit. on pp. 8, 15).

Cogley, T. and T. J. Sargent (2005). "Drifts and volatilities: monetary policies and outcomes in the post WWII US". *Review of Economic dynamics* 8, pp. 262–302 (cit. on p. 2).

Eisenstat, E., J. C. C. Chan, and R. W. Strachan (2016). "Stochastic Model Specification Search for Time-Varying Parameter VARs". *Econometric Reviews* 35, pp. 1638–1665 (cit. on p. 2).

Fagan, G., J. Henry, and R. Mestre (2001). *An Area Wide Model (AWM) for the Euro Area*. ECB Working Paper. European Central Bank. URL: `https://ssrn.com/abstract=356003` (cit. on p. 18).

Frühwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models". *Journal of Time Series Analysis* 15, pp. 183–202 (cit. on pp. 8, 15).

Frühwirth-Schnatter, S. and H. Wagner (2010). "Stochastic model specification search for Gaussian and partial non-Gaussian state space models". *Journal of Econometrics* 154, pp. 85–100 (cit. on pp. 2, 11).

Geweke, J. and G. Amisano (2010). "Comparing and evaluating Bayesian predictive distributions of asset returns". *International Journal of Forecasting* 26, pp. 216–230 (cit. on p. 22).

Giordani, P. and R. Kohn (2008). "Efficient Bayesian inference for multiple change-point and mixture innovation models". *Journal of Business & Economic Statistics* 26, pp. 66–77 (cit. on p. 2).

References II

Giordani, P. and M. Villani (2010). "Forecasting macroeconomic time series with locally adaptive signal extraction". *International Journal of Forecasting* 26, pp. 312–325 (cit. on p. 2).

Griffin, J. E. and P. J. Brown (2010). "Inference with normal-gamma prior distributions in regression problems". *Bayesian Analysis* 5, pp. 171–188 (cit. on pp. 6, 12).

Huber, F. and M. Feldkircher (forthcoming). "Adaptive Shrinkage in Bayesian Vector Autoregressive Models". *Journal of Business and Economic Statistics* (cit. on p. 2).

Huber, F., G. Kastner, and M. Feldkircher (2018). *A New Approach To Time Varying Parameters in Vector Autoregressive Models*. arXiv pre-print 1607.04532. URL: https://arxiv.org/abs/1607.04532 (cit. on p. 2).

Kalli, M. and J. E. Griffin (2014). "Time-varying sparsity in dynamic regression models". *Journal of Econometrics* 178, pp. 779–793 (cit. on p. 2).

Kalli, M. and J. E. Griffin (2018). "Bayesian nonparametric vector autoregressive models". *Journal of Econometrics* forthcoming (cit. on p. 2).

Kastner, G. (2016). "Dealing with Stochastic Volatility in Time Series Using the R Package stochvol". *Journal of Statistical Software* 69, pp. 1–30 (cit. on p. 15).

Kastner, G. and S. Frühwirth-Schnatter (2014). "Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models". *Computational Statistics & Data Analysis* 76, pp. 408–423 (cit. on p. 15).

Koop, G. and D. Korobilis (2012). "Forecasting Inflation Using Dynamic Model Averaging". *International Economic Review* 53, pp. 867–886 (cit. on p. 24).

Koop, G. and S. M. Potter (2007). "Estimation and forecasting in models with multiple breaks". *The Review of Economic Studies* 74, pp. 763–789 (cit. on p. 2).

References III

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009). "On the evolution of the monetary policy transmission mechanism". *Journal of Economic Dynamics and Control* 33, pp. 997–1017 (cit. on p. 2).

Lopes, H. F., R. E. McCulloch, and R. S. Tsay (2016). *Parsimony Inducing Priors for Large Scale State-Space Models*. Tech. rep. (cit. on p. 10).

Maheu, J. M. and Y. Song (forthcoming). "An efficient Bayesian approach to multiple structural change in multivariate time series". *Journal of Applied Econometrics* (cit. on p. 2).

McCracken, M. W. and S. Ng (2016). "FRED-MD: A monthly database for macroeconomic research". *Journal of Business & Economic Statistics* 34, pp. 574–589 (cit. on p. 18).

Primiceri, G. E. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy". *Review of Economic Studies* 72, pp. 821–852. URL: https://ideas.repec.org/a/oup/restud/v72y2005i3p821-852.html (cit. on pp. 2, 4).

Raftery, A., M. Karny, and P. Ettler (2010). "Online Prediction Under Model Uncertainty Via Dynamic Model Averaging: Application to a Cold Rolling Mill". *Technometrics* 52, pp. 52–66 (cit. on p. 24).

Sims, C. A. and T. Zha (2006). "Were there regime switches in US monetary policy?" *The American Economic Review* 96, pp. 54–81 (cit. on p. 2).

Thomas, R., S. Hills, and N. Dimsdale (Dec. 2010). "The UK Recession in Context – What do Three Centuries of Data Tell Us?" *Bank of England Quarterly Bulletin* 50, pp. 277–291 (cit. on p. 18).