Large time varying parameter VARs for macroeconomic forecasting

Gianni Amisano (FRB)
Domenico Giannone (NY Fed, CEPR and ECARES)
Michele Lenza (ECB and ECARES) [1]

Invited talk at the Cleveland Fed, May 31 2016

[1] Views expressed here are not those of the ECB or the FRS

1. Background
   - The TVP-BVAR model
   - Posterior simulation
   - Adding stochastic volatility

2. Large TVP VARs for macroeconomic forecasting
   - The challenge
   - The model
   - Stochastic volatility
   - Application results (so far)
   - Wrap up

Time varying parameter VARs, I/II

- Doan Litterman and Sims (1984), Econometric Reviews paper
- Litterman (1979, 1986)
- RATTS package
- Widespread use of TVP-VARs in US Fed system
- A convenient exploratory device in applied macro and for backing out effects of structural shocks
- (Primiceri, 2005, Cogley and Sargent, 2001, 2004)
- Also successful in a forecasting environment: Canova and Ciccarelli, 2004, D'Agostino, Gambetti Giannone (2012)
- Panel VAR dimension: Canova, Ciccarelli, Ortega (2007, 2012)

Time varying parameter VARs, II/II

- See the survey by Koop and Korobilis (2010)
- In spite of obvious overparameterization, these models are capable of reproducing salient features of the data in a very effective way.
- Interest re-ignited by the good-policy, good luck controversy.
- Cogley and Sargent 2001, 2005
- Primiceri 2005
- Koop and coauthors, many papers
- L. Benati, many papers

Specification of the model, I/II

- Suppose to consider a time varying parameters (TVP) VAR model potentially allowing for exogenous and deterministic regressors:

$$\mathbf{y}_t = \mathbf{x}_t^{'}\mathbf{B}_t^{'} + \varepsilon_t, , \varepsilon_t \backsim N(\mathbf{0}, \Sigma) \quad (1)$$

$$\Leftrightarrow \mathbf{y}_t = (\mathbf{I}_n \otimes \mathbf{x}_t^{'})vec(\mathbf{B}_t^{'}) + \varepsilon_t \quad (1)$$

Specification of the model, II/II

- Let us stick to (1) and call β_t the vector collecting all the first order parameters of the model.
- The state equation for them is:

$$\beta_t = vec(\mathbf{B}_t^{'}) \tag{2}$$
$$\beta_t = \beta_{t-1} + \eta_t \tag{3}$$
$$\eta_t \backsim NID(\mathbf{0}, \Omega)$$

- Assumption:

$$Var\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} = \begin{bmatrix} \Sigma & [0] \\ [0] & \Omega \end{bmatrix}$$

- The initialisation of the $\mathbf{B}_t$ sequence (prior distribution for β_0) is:

$$p(\beta_0) \backsim N(\underline{\beta}_0, \mathbf{Q}_0) \tag{4}$$

Posterior simulation

- Usually MCMC implemented by simulating latent states (betas) and fixed parameters in different blocs
- latent states by running Kalman filter and simulation smoother
- simulation of fixed parameters conceptually easy

Stochastic volatility

- In the presence of outliers, i.e. large shocks in the VAR, a TVP model with homoskedastic innovations might spuriously signal changes in the VAR coefficients.
- This motivated Primiceri and then Cogley and Sargent (2005) to extend the model to have VAR shocks with stochastic volatility.
- Now I will describe the setting by using a simplified framework

A simple SV model I/III

Suppose we have the following univariate time series

$$
\begin{aligned}
y_t &= \exp\left(h_t/2\right) \times e_t \\
h_t &= \rho_h h_{t-1} + \sigma_h v_t \\
\left[\begin{array}{c} e_t \\ v_t \end{array}\right] &\sim NID\left[\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right)\right]
\end{aligned}
$$

this is the simplest SV model
Gaussian non-linear state space model: bye bye to Kalman filter.

A simple SV model II/III

- But the model can be transformed into a linear state space with non-Gaussian innovations
- $$z_t = \ln(y_t^2) = h_t + \ln(e_t^2)$$
$$h_t = \rho_h h_{t-1} + \sigma_h v_t$$
$$\ln(e_t^2) \sim \ln \chi^2_{(1)}$$

  i.e. a linear state space system with Gaussian state equation shocks and with log $\chi^2_{(1)}$ measurement errors.

A simple SV model III/III

- Kim, Shephard and Chib (1998) showed how to deal with this model:
- approximate $\ln \chi^2_{(1)}$ with a 5 component mixture of Gaussian (known probabilities) and then use Mixture of Gaussian state space model simulation
- Easy to implement
- In Primiceri (2005) it was implemented not correctly in the VAR framework but Del Negro and Primiceri (2015) have shown how to implement it properly in a TVP-VAR context.

Alternative route: single move algorithm, I/III

- Jacquier, Polson and Rossi (1994) algorithm.
- Problem: want to draw $\underline{\mathbf{h}}_T$, whole sequence of sample volatilities, conditioned on other parameters/state variables and data.
- Single move approach: draw $h_t^{(i)}$ from its distribution conditional on $h_1^{(i-1)}, h_2^{(i-1)}, ..., h_{t-1}^{(i-1)}, h_{t+1}^{(i-1)}, ..., h_T^{(i-1)}$

$$p(h_t | h_1^{(i-1)}, h_2^{(i-1)}, ..., h_{t-1}^{(i-1)}, h_{t+1}^{(i-1)}, ..., h_T^{(i-1)}, \underline{\beta}_T^{(i-1)}, \theta^{(i-1)}, \underline{\mathbf{y}}_t^*)$$

Alternative route: single move algorithm, II/III

- It turns out that, given Markovian property of the volatility process, the above expression simplifies to

$$
\begin{aligned}
p(h_t | h_{t-1}^{(i-1)}, h_{t+1}^{(i-1)}, \theta^{(i-1)}, \varepsilon_t) \propto \\
& p(h_t | h_{t-1}^{(i-1)}, h_{t+1}^{(i-1)}, \theta^{(i-1)}) \times \\
& p(h_{t+1}^{(i-1)} | h_t, \theta^{(i-1)}) \times \\
& p(\varepsilon_t | h_t, \theta^{(i-1)})
\end{aligned}
$$

Alternative route: single move algorithm, III/III

- The first two terms are Gaussian distributions in terms of h_t (it enters quadratically in the exponential part) but the last bit is non-standard

$$p(\varepsilon_t|h_t, \theta^{(i-1)}) = \frac{1}{\sqrt{2\pi}\exp(h_t/2)} \times \exp\left(-\frac{1}{2\exp(h_t)}\varepsilon_t^2\right)$$

- Hence JPR (1994) suggest to use Metropolis-Hastings with candidate

$$p(h_t|h_{t-1}^{(i-1)}, h_{t+1}^{(i-1)}, \theta^{(i-1)}) \times p(h_{t+1}^{(i-1)}|h_t, \theta^{(i-1)})$$

and then accept-reject using

$$\alpha(h_t, h_t) = \min\left\{\frac{p(\varepsilon_t|h_t, \theta^{(i-1)})}{p(\varepsilon_t|h_t^{(i-1)}, \theta^{(i-1)})}, 1\right\}$$

Curse of dimensionality in computation and inference, I

- Based on Amisano, Giannone and Lenza (2016)
- The challenge is to use them with many variables:
- $n^2 \times p + n = n \times k$, the number of state variables one for each VAR coefficient
- a covariance matrix of shocks in the state equation with the same number of rows and columns

Curse of dimensionality in computation and inference, II

- Example 1: Primiceri, 18 elements in the state vector and a 18×18 state equation shocks covariance matrix, i.e. 171 free parameters to estimate.
- Example 2: in the application conducted in this project (7 series, 4 lags + constant), we have 203 elements in the state vector and and a 203×203 state equation shocks covariance matrix (20706 free coefficients).

Prior and initialization

- To contain effects of random walk dynamics on coefficients and large state space it is very important to
  - specify a prior that greatly limits the amount of time variation per period
  - specify a sensible initialization of coefficients
- Usually both tasks are performed by OLS estimation over a training sample but this is problematic to do on a large dimension, since OLS will not work well under these circumstances.

Some solutions

- Doan Litterman and Sims (1984): based on uni-equational analysis and Minnesota prior: each coefficient drifts with independent shock
- Canova and Ciccarelli (2004, 2007), based on imposing factor structure on coefficients. Very useful for panel VARs.
- Koop and Korobilis (2014): use forgetting factors and approximation that allows to avoid completely MCMC.
  - How large is approximation error?
  - How about effects of parameter uncertainty?

Our approach

- Use Bayesian VAR on the pre-sample to obtain initialization and calibrate prior in a conservative way.
- Use VAR Kronecker structure to reduce number of free parameters in state equation.
- Augment the model with a stochastic volatility specification that preserves the Kronecker structure.

Results so far

- Experiment with 7 variable VAR, 4 lags and constant, i.e. 29 x 7 coefficients
- Experiment with 20 variable VAR, 2 lags and constant, i.e. 41 x 20 coefficients
- It can be done
- It works well in prediction

A simple TVP-VAR model

- State space form (k parameters for each of n equations):

$$\mathbf{y}_t^{'} = \mathbf{x}_t^{'}\mathbf{B}_t + \varepsilon_t^{'} \Leftrightarrow \mathbf{y}_t = (\mathbf{I}_n \otimes \mathbf{x}_t)\,\beta_t + \varepsilon_t, \beta_t = vec(\underset{(k\times n)}{\mathbf{B}_t})$$

$$\beta_t = \beta_{t-1} + \eta_t \Leftrightarrow \mathbf{B}_t = \mathbf{B}_{t-1} + \mathbf{H}_t, \eta_t = vec(\mathbf{H}_t)$$

$$\varepsilon_t \backsim NID(\mathbf{0}, \Sigma)$$

$$\eta_t \backsim NID(\mathbf{0}, \Sigma \otimes \Omega)$$

$$cov(\eta_t, \varepsilon_s) = [\mathbf{0}]$$

What does this Kronecker structure imply?

- Bayesian notion of exchangeability
- Example with $n = 2$:

$$\begin{bmatrix} \sigma_{11}\Omega & \sigma_{12}\Omega \\ \sigma_{12}\Omega & \sigma_{22}\Omega \end{bmatrix}$$

- Ratio of correlations across shocks associated to parameters in the same position in two sets of equations will be driven by ratios of covariances of VAR shocks

$$\frac{\sigma_{ij}\omega_{hk}}{\sigma_{lm}\omega_{hk}} = \frac{\sigma_{ij}}{\sigma_{lm}}$$

Kalman filtering and simulation smoother

$$\widehat{\mathbf{B}}_{t|t-1} = \widehat{\mathbf{B}}_{t-1|t-1}, \mathbf{Q}_{t|t-1} = \Sigma \otimes \mathbf{P}_{t|t-1}, \mathbf{P}_{t|t-1} = \mathbf{P}_{t-1|t-1} + \Omega$$

$$\widehat{\beta}_{t|t} = \widehat{\beta}_{t|t-1} + (\mathbf{I}_n \otimes \mathbf{K}_t) \times \varepsilon_{t|t-1} \Longleftrightarrow \widehat{\mathbf{B}}_{t|t} = \widehat{\mathbf{B}}_{t|t-1} + \mathbf{K}_t \varepsilon'_{t|t-1}$$

$$\mathbf{K}_t = \frac{\mathbf{P}_{t|t-1}\mathbf{x}_t}{(\mathbf{x}'_t \mathbf{P}_{t|t-1}\mathbf{x}_t + 1)}, \mathbf{Q}_{t|t} = \Sigma \otimes \mathbf{P}_{t|t},$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \frac{\mathbf{P}_{t|t-1}\mathbf{x}_t \mathbf{x}'_t \mathbf{P}_{t|t-1}}{(\mathbf{x}'_t \mathbf{P}_{t|t-1}\mathbf{x}_t + 1)}$$

- It involves updating with respect to $\mathbf{P}_{t|t}$, a $(k \times k)$ matrix, i.e. the same dimension as the number of regressors for each equation
- Smoother has same qualitative properties

Scalar SV factor

- Clark, Carriero and Marcellino (2015): use scalar factor to drive volatility in a set of macro aggregates in a fixed coefficients VAR.
- We adopt the same approach to our TV VAR modifying the covariance matrix of VAR shocks as follows:

$$\begin{aligned} \Sigma_t &= \exp(h_t/2) \times \Sigma \\ \Delta h_t &= \sigma_h e_t, \\ h_1 &= 0 \text{ (normalization)} \end{aligned}$$

- It is possible to scale each observation by $\exp(-h_t/2)$ and use the same computations as in no SV case.
- Allowing for multiple volatility factors would destroy the Kronecker structure we use. It would not be crucial in terms of number of additional parameters to be estimated, but it would slow down computations considerably

Priors and initialization

- Priors based on estimating a Bayesian fixed coefficient VAR with Minnesota prior on a training sample (here 60 observations).
- Minnesota prior: shring all parameters to zero. Shrink first own lag to one.
- Wishart priors on Ω^{-1} and Σ^{-1} such that

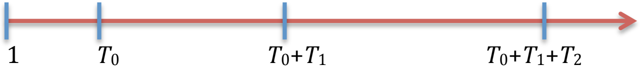
$$E(\Omega^{-1}) = \kappa \times (\mathbf{X}_0^{*'}\mathbf{X}_0^{*}), df = 40, \kappa = 10000$$
$$E(\Sigma^{-1}) = \widehat{\Sigma}_0^{-1}, df = 15$$

- X_0^* =matrix of training sample observations for regressors + dummy observations for Minnesota Prior
- $\widehat{\Sigma}_0^{-1}$ =posterior mean of Σ^{-1} from training sample.
- Initialize $\mathbf{B}_{t|t}$ at posterior mean of $\mathbf{B}$ from training sample and $\mathbf{P}_{t|t}$ at $\mathbf{X}_0^{*'}\mathbf{X}_0^{*}$

Data and computations, I/III

- US quarterly, 1948Q1-2011Q4 , revisions as of Feb 16th 2012. Same data set as in Amisano and Geweke (2013), extension of Smets and Wouters (2007) data set.
- 7 Series:real per capita consumption, investment, GDP and wage, hours worked, GDP inflation, Fed funds rate.

Data and computations, II/III



- $[1 : T_0]$: 1948Q1-1963Q4. Training sample (VAR with Minnesota prior and sum of coefficients prior)
- $[T_0 + 1 : T_0 + T_1]$: 1964Q1-1973Q4.Initial estimation period.
- $[T_0 + T_1 + : T_0 + T_1 + T_2]$: 1974Q1-2011Q4. Evaluation period, during which $p(y_{T_0+T_1+\tau+h}|y_{T_0+1:T_0+T_1+\tau}), \tau = 1, 2, .., T_2 - h$ is recursively computed (out of sample behaviour)

Data and computations, III/III

- Straightforward Gibbs + data augmentation for Σ, Ω, σ_h. Easy Metropolis step for drawing volatilities.
- full recursive estimation via MCMC on subsamples, and out of sample evaluation.
- Easy to distribute on a cluster: using 12 nodes it takes 1-2 days to perform the analysis of all subsamples (D' Agostino et al. 2012 employed 1 month for a 3 equation model.

Predictive performance (I)

- Evaluations and comparisons based on log scores

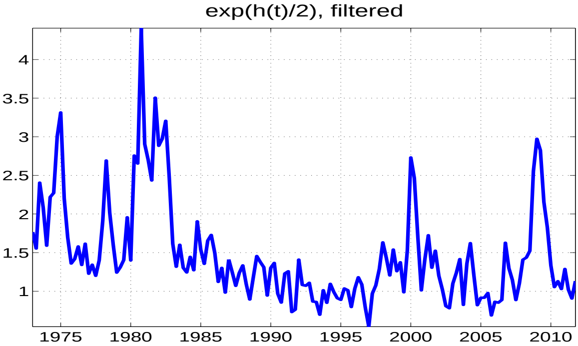
$$\ln p(\mathbf{y}_{T_0+t+1}|\mathbf{y}_{T_0+1:T_0 t}, M_j) - \ln p(\mathbf{y}_{T_0+t+1}|\mathbf{y}_{T_0+1:T0+t}, RW)$$

Comparison across models and subperiods

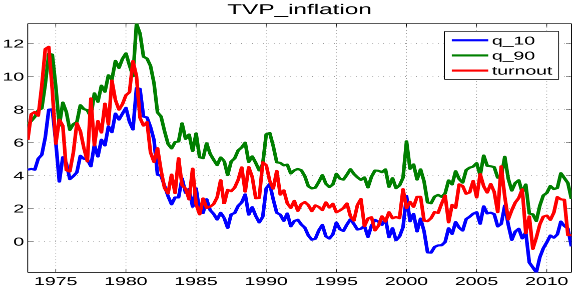
- Differences with respect to multivariate random walk with estimated covariance matrix.

	VAR	TVP	TVP SV
all subsamples	24.32	53.13	59.73
mean all subsamples	0.16	0.34	0.38
mean recessions	0.18	0.59	0.68
mean expansions	0.15	0.28	0.31

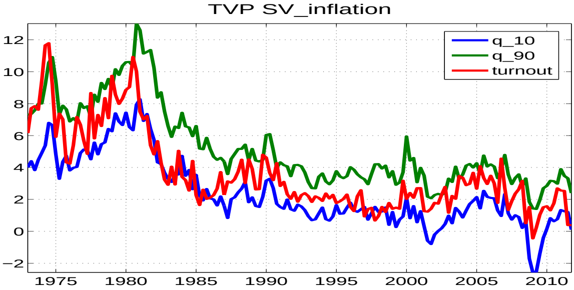
- Time varying parameter specifications beat fixed parameter VAR.
- SV version slightly superior.
- TV specifications seems more useful in recessions.

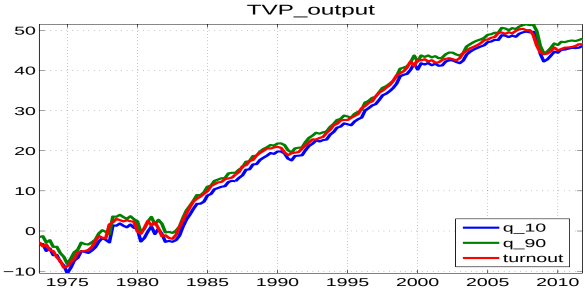

exp(h(t)/2), filtered
4
3.5
3
2.5
2
1.5
1
1975
1980
1985
1990
1995
2000
2005
2010

A look at individual variable forecasts: inflation, TVP

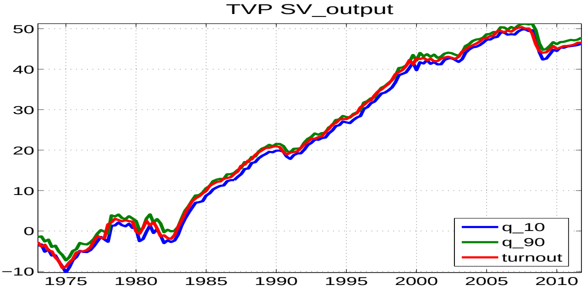


A look at individual variable forecasts: inflation, TVP-SV

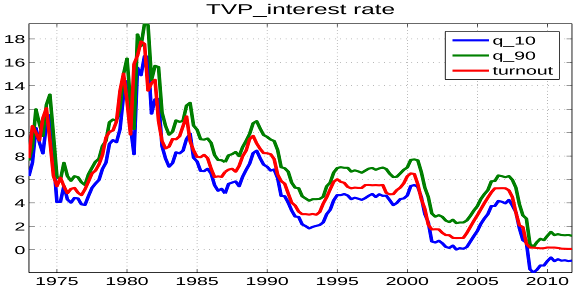



TVP_output
50
40
30
20
10
0
−10
1975
1980
1985
1990
1995
2000
2005
2010
q_10
q_90
turnout

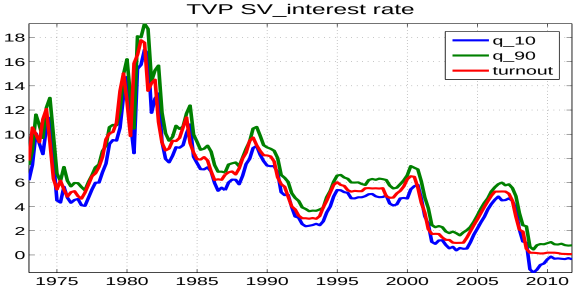
A look at individual variable forecasts: output, TVP-SV



A look at individual variable forecasts: short rate, TVP



A look at individual variable forecasts: short rate, TVP-SV



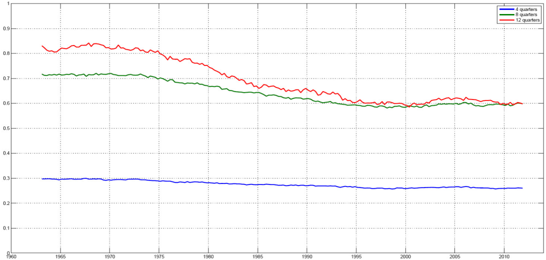
Coverage, One minus coverage rate

		TVP	TVP-SV
10-90 range	inflation	0.14	0.22
	output	0.22	0.28
	short rate	0.14	0.19
25-75 range	inflation	0.44	0.50
	output	0.49	0.53
	short rate	0.28	0.40

- No formal analysis of PITs yet. On the agenda (Rossi and Sekhposyan, 2014, Amisano and Geweke, 2015)

- Full sample estimation (excluding training sample: 1964Q1-2011Q4)
- Estimate "at each quarter", implied generalized impulse
- response distributions to a 0.1% in GDP
- Report median ratio of Δ Infl. to Δ GDP at 4, 8 and 12 quarters after the shock (Measure of Phillips correlation over time)

The Phillips correlation, II



What we have found so far

- We use Bayesian estimation of a fixed coefficient VAR on the training sample to regularize initialization and prior setting for TVP model
- We show that it is feasible and reasonable to Kronecker structure for innovations in coefficients
- We can extend to scalar SV using same algorithms. This is currently important for speed of computations but not important per se for forecasting

Ways ahead

- Investigate factor structure for coefficients, following Canova and Ciccarelli (2004), but also with other schemes
- Use sequential Monte Carlo (Durham and Geweke, 2015)
- Focus identification of structural shocks
- Focus more on interesting macro features: eg does Phillips curve slope change over time? (Stock and Watson, 2012)