Bayesian Methods for Empirical Macroeconomics

Gary Koop

September 18, 2012

Introduction

- Why TVP-VARs?
- Example: U.S. monetary policy
- was the high inflation and slow growth of the 1970s were due to bad policy or bad luck?
- Some have argued that the way the Fed reacted to inflation has changed over time
- After 1980, Fed became more aggressive in fighting inflation pressures than before
- This is the "bad policy" story (change in the monetary policy transmission mechanism)
- This story depends on having VAR coefficients different in the 1970s than subsequently.

- Others think that variance of the exogenous shocks hitting economy has changed over time
- Perhaps this may explain apparent changes in monetary policy.
- This is the "bad luck" story (i.e. 1970s volatility was high, adverse shocks hit economy, whereas later policymakers had the good fortune of the Great Moderation of the business cycle – at least until 2008)
- This motivates need for multivariate stochastic volatility to VAR models
- Cannot check whether volatility has been changing with a homoskedastic model

- Most macroeconomic applications of interest involve several variables (so need multivariate model like VAR)
- Also need VAR coefficients changing
- Also need multivariate stochastic volatility
- TVP-VARs are most popular models with such features
- But other exist (Markov-switching VARs, Vector Floor and Ceiling Model, etc.)

Homoskedastic TVP-VARs

- Begin by assuming $\Sigma_t = \Sigma$
- Remember VAR notation: y_t is $M \times 1$ vector, Z_t is $M \times k$ matrix (defined so as to allow for a VAR with different lagged dependent and exogenous variables in each equation).
- TVP-VAR:

$$y_t = Z_t \beta_t + \varepsilon_t$$

$$\beta_{t+1} = \beta_t + u_t$$

- ε_t is i.i.d. $N(0, \Sigma)$ and u_t is i.i.d. $N(0, Q)$.
- ε_t and u_s are independent of one another for all s and t.

- Bayesian inference in this model?
- Already done: this is just the Normal linear state space model of the last lecture.
- MCMC algorithm of standard form (e.g. Carter and Kohn, 1994).
- But let us see how it works in practice in our empirical application
- Follow Primiceri (2005)

Illustration of Bayesian TVP-VAR Methods

- Same quarterly US data set from 1953Q1 to 2006Q3 as was used to illustrate VAR methods
- Three variables: Inflation rate $\Delta\pi_t$, the unemployment rate u_t and the interest rate r_t
- VAR lag length is 2.
- Training sample prior: prior hyperparameters are set to OLS quantities calculating using an initial part of the data
- Our training sample contains 40 observations.
- Data through 1962Q4 used to choose prior hyperparameter values, then Bayesian estimation uses data beginning in 1963Q1.

- β_{OLS} is OLS estimate of VAR coefficients in constant-coefficient VAR using training sample
- $V(\beta_{OLS})$ is estimated covariance of β_{OLS}.
- Prior for β_0:

$$\beta_0 \sim N(\beta_{OLS}, 4 \cdot V(\beta_{OLS}))$$

- Prior for Σ^{-1} Wishart prior with $\underline{\nu} = M + 1, \underline{S} = I$
- Prior for Q^{-1} Wishart prior with $\underline{\nu}_Q = 40, \underline{Q} = 0.0001 \cdot 40 \cdot V(\beta_{OLS})$

- With TVP-VAR we have different set of VAR coefficients in every time period
- So different impulse responses in every time period.
- Figure 1 presents impulse responses to a monetary policy shock in three time periods: 1975Q1, 1981Q3 and 1996Q1.
- Impulse responses defined in same way as we did for VAR
- Posterior median is solid line and dotted lines are 10^{th} and 90^{th} percentiles.

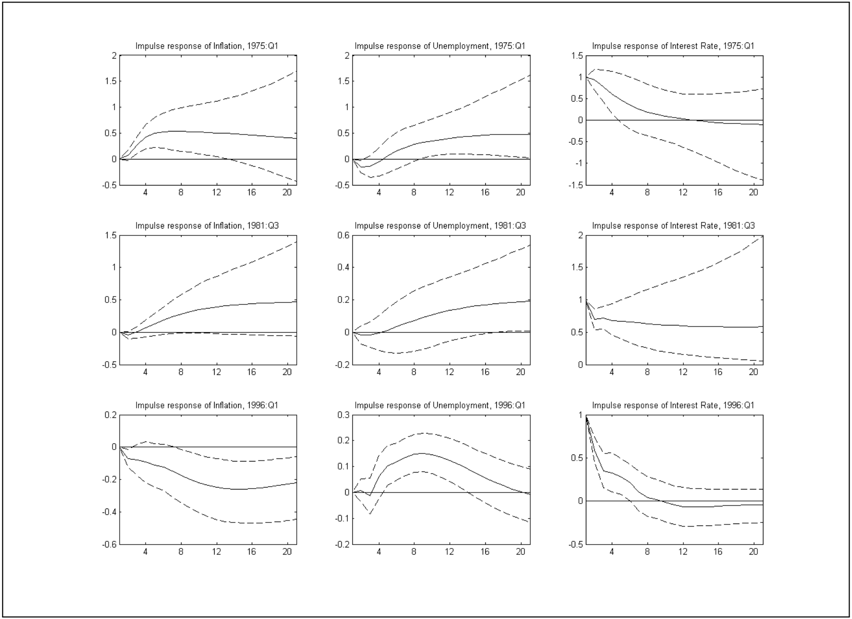

Figure 1: Impulse responses at different times

Optional Topic: Combining other Priors with the TVP Prior

- Often Bayesian TVP-VARs work very well in practice.
- In some case the basic TVP-VAR does not work as well, due to over-parameterization problems.
- Previously, we noted worries about proliferation of parameters in VARs, which led to use of priors such as the Minnesota prior or the SSVS prior.
- With many parameters and short macroeconomic time series, it can be hard to obtain precise estimates of coefficients.
- Risk of over-fitting
- Priors which exhibit shrinkage of various sorts can help mitigate these problems.
- With TVP-VAR proliferation of parameters problems is even more severe.
- Hierarchical prior of state equation is big help, but may want more in some cases.

Optional Topic (cont.): Combining TVP Prior with Minnesota Prior

- E.g. Ballabriga, Sebastian and Valles (1999, JIE), Canova and Ciccarelli (2004, JOE), and Canova (2007, book)
- Replace TVP-VAR state equation by

$$\beta_{t+1} = A_0 \beta_t + (I - A_0) \overline{\beta}_0 + u_t$$

- u_t is i.i.d. $N(0, Q)$
- A_0, $\overline{\beta}_0$ and Q can be unknown parameters or set to known values
- E.g. Canova (2007) sets $\overline{\beta}_0$ and Q to have forms based on the Minnesota prior and sets $A_0 = cI$ where c is a scalar.
- Note if $c = 1$, then $E(\beta_{t+1}) = E(\beta_t)$ (as in TVP-VAR)
- If $c = 0$ then $E(\beta_{t+1}) = \overline{\beta}_0$ (as in Minnesota prior)
- Q based on prior covariance of Minnesota prior
- c can either be treated as an unknown parameter or a value can be selected for it.

Optional Topic (cont.): Combining TVP Prior with SSVS Prior

- Same setup as preceding slide
- Set $\overline{\beta}_0 = 0$.
- Let $a_0 = vec\left(A_0\right)$
- Use SSVS prior for a_0
- a_{0j} (the j^{th} element of a_0) has prior:

$$a_{0j}|\gamma_j \sim \left(1-\gamma_j\right) N\left(0,\kappa_{0j}^2\right) + \gamma_j N\left(0,\kappa_{1j}^2\right)$$

- as before, γ_j is dummy variable
- κ_{0j}^2 is very small (so that a_{0j} is constrained to be virtually zero)
- κ_{1j}^2 is large (so that a_{0j} is relatively unconstrained).
- Property: with probability γ_j, a_{0j} is evolving according to a random walk in the usual TVP fashion
- With probability $\left(1-\gamma_j\right)$, $a_{0j} \approx 0$

Optional Topic (cont.): MCMC Algorithm

- I will not provide complete details, but note only:
- These are Normal linear state space models so standard algorithms (e.g. Carter and Kohn) can draw β^T
- For TVP+Minnesota prior this is enough (other parameters fixed)
- For TVP+SSVS simple to adapt MCMC algorithm for SSVS with VAR

Optional Topic (cont.): Adding Another Layer to the Prior Hierarchy

- Another approach used by Chib and Greenberg (1995, JOE) for SUR model
- Adapted for VARs by, e.g., Ciccarelli and Rebucci (2002)
-

$$
\begin{aligned}
y_t &= Z_t \beta_t + \varepsilon_t \\
\beta_{t+1} &= A_0 \theta_{t+1} + u_t \\
\theta_{t+1} &= \theta_t + \eta_t
\end{aligned}
$$

- all assumptions as for TVP-VAR, plus η_t is i.i.d. $N(0, R)$
- Slightly more general that previous Normal linear state space model, but very similar MCMC (so will not discuss MCMC)

Optional Topic (cont.): Adding Another Layer to the Prior Hierarchy

- Why might this generalization be useful?
- Can be written as:

$$\begin{aligned} y_t &= Z_t \beta_t + \varepsilon_t \\ \beta_{t+1} &= \beta_t + v_t \end{aligned}$$

- where $v_t = A_0 \eta_t + u_t - u_{t-1}$.
- So still a TVP-VAR with random walk state equation but state equation errors have different form (MA(1)).
- Also can see that:

$$E\left(\beta_{t+1} | \beta_t\right) = A_0 \beta_t$$

Optional Topic (cont.): Adding Another Layer to the Prior Hierarchy

- A_0 can be chosen to reflect some other prior information
- E.g. SSVS prior as above
- E.g. Ciccarelli and Rebucci (2002) is panel VAR application
- G countries and, for each country, k_G explanatory variables exist with time-varying coefficients.
- They set

$$A_0 = \iota_G \otimes I_{k_G}$$

- Implies time-varying component in each coefficient which is common to all countries
- Parsimony: θ_t is of dimension k_G whereas β_t is of dimension $k_G \times G$.

Imposing Inequality Restrictions on the VAR Coefficients

- Another way of ensuring shrinkage
- E.g. restrict β_t to be non-explosive (i.e. roots of the VAR polynomial defined by β_t lie outside the unit circle)
- Sometimes (given over-fitting and imprecise estimates) can get posterior weight in explosive region
- Even small amount of posterior probability in explosive regions for β_t can lead to impulse responses or forecasts which have counter-intuitively large posterior means or standard deviations.
- Koop and Potter (2009, on my website) discusses how to do this. I will not present details, but outline basic idea

- With unrestricted TVP-VAR, took draws $p\left(\beta^T | y^T, \Sigma, Q\right)$ using MCMC methods for Normal linear state space models
- One method to impose inequality restrictions involves:
- Draw β^T in the unrestricted VAR. If *any* drawn β_t violates the inequality restriction then the *entire* vector β^T is rejected.
- Problem: this algorithm can get stuck, rejecting virtually every β^T (all you need is a single drawn β_t to violate inequality and entire β^T is rejected)
- Note: algorithms like Carter and Kohn are "multi-move algorithms" (draw β^T all at same time).
- Alternative is "single move algorithm": drawing β_t for $t = 1, .., T$ one at a time from $p\left(\beta_t | y^T, \Sigma, Q, \beta_{-t}\right)$ where $\beta_{-t} = \left(\beta_1', .., \beta_{t-1}, \beta_{t+1}, .., \beta_T'\right)'$

- Koop and Potter (2009) suggest using single move algorithm
- Reject β_t only (not β^T) if it violates inequality restriction
- Usually multi-move algorithms are better than single-move algorithms since latter can be slow to mix.
- I.e. they produce highly correlated series of draws which means that, relative to multi-move algorithms, more draws must be taken to achieve a desired level of accuracy.
- But if multi-move algorithm gets stuck, single move might be better.

Dynamic Mixture Models

- Remember: Normal linear state space model depends on so-called system matrices, Z_t, Q_t, T_t, W_t and Σ_t.
- Suppose some or all of them depend on an $s \times 1$ vector $\widetilde{K}_t$
- Suppose $\widetilde{K}_t$ is Markov random variable (i.e. $p\left(\widetilde{K}_t | \widetilde{K}_{t-1}, .., \widetilde{K}_1\right) = p\left(\widetilde{K}_t | \widetilde{K}_{t-1}\right)$ or independent over t
- Particularly simple if $\widetilde{K}_t$ is a discrete random variable.
- Result is called a dynamic mixture model
- Gerlach, Carter and Kohn (2000, JASA) have an efficient MCMC algorithm

- Why are dynamic mixture models useful in empirical macroeconomics?
- E.g. TVP-VAR:

$$y_t = Z_t \beta_t + \varepsilon_t$$

$$\beta_{t+1} = \beta_t + u_t$$

- ε_t is i.i.d. $N(0, \Sigma)$
- BUT: u_t is i.i.d. $N\left(0, \widetilde{K}_t Q\right)$.
- Let $\widetilde{K}_t \in \{0, 1\}$ with hierarchical prior:

$$p\left(\widetilde{K}_t = 1\right) = q.$$
$$p\left(\widetilde{K}_t = 0\right) = 1 - q$$

- where q is an unknown parameter.

- Property:
- If $\widetilde{K}_t = 1$ then usual TVP-VAR:

$$\beta_{t+1} = \beta_t + u_t$$

- If $\widetilde{K}_t = 0$ then VAR coefficients do not change at time t:

$$\beta_{t+1} = \beta_t$$

- Parsimony.
- This model can have flexibility of TVP-VAR if the data warrant it (i.e. can select $\widetilde{K}_t = 1$ for $t = 1, .., T$).
- But can also select a much more parsimonious representation.
- An extreme case: if $\widetilde{K}_t = 0$ for $t = 1, .., T$ then back to VAR without time-varying parameters.

- I will not present details of MCMC algorithm since it is becoming a standard one
- See also the Matlab code on my website
- Dynamic mixture models used to model structural breaks, outliers, nonlinearities, etc.
- E.g. Giordani, Kohn and van Dijk (2007, JoE).

TVP-VARs with Stochastic Volatility

- In empirical work, you will usually want to add multivariate stochastic volatility to the TVP-VAR
- But this can be dealt with quickly, since the appropriate algorithms were described in the lecture on State Space Modelling
- Remember, in particular, the approaches of Cogley and Sargent (2005) and Primiceri (2005).
- MCMC: need only add another block to our algorithm to draw Σ_t for $t = 1, .., T$.
- Homoskedastic TVP-VAR MCMC: $p\left(Q^{-1}|y^T, \beta^T\right)$, $p\left(\beta^T|y^T, \Sigma, Q\right)$ and $p\left(\Sigma^{-1}|y^T, \beta^T\right)$
- Heteroskedastic TVP-VAR MCMC: $p\left(Q^{-1}|y^T, \beta^T\right)$, $p\left(\beta^T|y^T, \Sigma_1, .., \Sigma_T, Q\right)$ and $p\left(\Sigma_1^{-1}, .., \Sigma_T^{-1}|y^T, \beta^T\right)$

Empirical Illustration of Bayesian Inference in TVP-VARs with Stochastic Volatility

- Continue same illustration as before.
- All details as for homoskedastic TVP–VAR
- Plus allow for multivariate stochastic volatility as in Primiceri (2005).
- Priors as in Primiceri
- Can present empirical features of interest such as impulse responses
- But (for brevity) just present volatility information
- Figure 2: time-varying standard deviations of the errors in the three equations (i.e. the posterior means of the square roots of the diagonal element of Σ_t)
- If time permits, I will show empirical results from dynamic mixture version of model from my paper with Leon-Gonzalez and Strachan (working paper version available on my website)

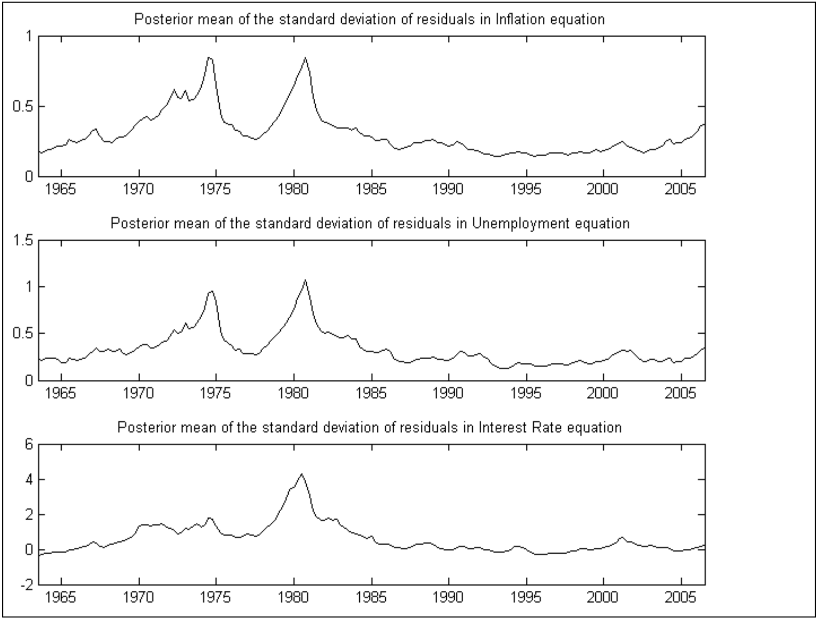

Figure 2: Volatilities in the 3 Equations

Summary

- TVP-VARs are useful for the empirical macroeconomists since they:
- are multivariate
- allow for VAR coefficients to change
- allow for error variances to change
- They are state space models so Bayesian inference can use familar MCMC algorithms developed for state space models.
- They can be over-parameterized so care should be taken with priors.
- I think this is enough material to be digested in a short course, however....
- If there is extra time I will give a brief introduction to Bayesian analysis of factor models