Bayesian Methods for Empirical Macroeconomics

Gary Koop

September 14, 2012

Introduction

- State space methods are used for a wide variety of time series problems
- They are important in and of themselves
- Also time-varying parameter VARs (TVP-VARs) and stochastic volatility are state space models
- Advantage of state space models: well-developed set of MCMC algorithms for doing Bayesian inference

- Remember: our general notation for a VAR was:

$$y_t = Z_t \beta + \varepsilon$$

- In many macroeconomic applications, unrealistic to assume constant β
- This leads to TVP-VAR:

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

- where

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

- This is a state space model.
- In VAR assume ε_t to be i.i.d. $N(0, \Sigma)$
- In empirical macroeconomics, this is often unrealistic.
- Want to have $var(\varepsilon_t) = \Sigma_t$
- This also leads to state space models.

The Normal Linear State Space Model

- Fairly general version of Normal linear state space model:
- Measurement equation:

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

- State equation:

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

- y_t and ε_t defined as for VAR
- W_t is known $M \times p_0$ matrix (e.g. lagged dependent variables or explanatory variables with constant coefficients)
- Z_t is known $M \times K$ matrix (e.g. lagged dependent variables or explanatory variables with time varying coefficients)
- β_t is $k \times 1$ vector of states (e.g. VAR coefficients)
- ε_t ind $N(0, \Sigma_t)$
- u_t ind $N(0, Q_t)$.
- ε_t and u_s are independent for all s and t.
- T_t is a $k \times k$ matrix (usually fixed, but sometimes not).

- Key idea: for given values for δ, T_t, Σ_t and Q_t (called "system matrices") posterior simulators for β_t for $t = 1, .., T$ exist.
- E.g. Carter and Kohn (1994, Btka), Fruhwirth-Schnatter (1994, JTSA), DeJong and Shephard (1995, Btka) and Durbin and Koopman (2002, Btka).
- I will not present details of these (standard) algorithms
- Notation: $\beta^t = (\beta_1', .., \beta_t')'$ stacks all the states up to time t (and similar superscript t convention for other things)
- Gibbs sampler: $p\left(\beta^T | y^T, \delta, T^T, \Sigma^T, Q^T\right)$ drawn use such an algorithm
- $p\left(\delta | y^T, \beta^T, T^T, \Sigma^T, Q^T\right)$, $p\left(T^T | y^T, \beta^T, \delta, \Sigma^T, Q^T\right)$, $p\left(\Sigma^T | y^T, \beta^T, \delta, T^T, Q^T\right)$ and $p\left(Q^T | y^T, \beta^T, \delta, T^T, \Sigma^T\right)$ depend on precise form of model (typically simple since, conditional on β^T have a Normal linear model)
- Typically restricted versions of this general model used
- TVP-VAR of Primiceri (2005, ReStud) has $\delta = 0$, $T_t = I$ and $Q_t = Q$

Example of an MCMC Algorithm

- Special case $\delta = 0$, $T_t = I, \Sigma_t = \Sigma$ and $Q_t = Q$
- Homoskedastic TVP-VAR of Cogley and Sargent (2001, NBER)
- Need prior for all parameters
- But state equation implies hierarchical prior for β^T:

$$\beta_{t+1} | \beta_t, Q \sim N\left(\beta_t, Q\right)$$

- Formally:

$$p\left(\beta^T | Q\right) = \prod_{t=1}^{T} p\left(\beta_t | \beta_{t-1}, Q\right)$$

- Hierarchical: since it depends on Q which, in turn, requires its own prior.

- Note β_0 enters prior for β_1.
- Need prior for β_0
- Standard treatments exist.
- E.g. assume $\beta_0 = 0$, then:

$$\beta_1 | Q \sim N(0, Q)$$

- Or Carter and Kohn (1994) simply assume β_0 has some prior that researcher chooses

- Convenient to use Wishart priors for Σ^{-1} and Q^{-1}
- $$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{\nu}\right)$$
- $$Q^{-1} \sim W\left(\underline{Q}^{-1}, \underline{\nu}_Q\right)$$

- Want MCMC algorithm which sequentially draws from $p\left(\Sigma^{-1}|y^T, \beta^T, Q\right)$, $p\left(Q^{-1}|y^T, \Sigma, \beta^T\right)$ and $p\left(\beta^T|y^T, \Sigma, Q\right)$.
- For $p\left(\beta^T|y^T, \Sigma, Q\right)$ use standard algorithm for state space models (e.g. Carter and Kohn, 1994)
- Can derive $p\left(\Sigma^{-1}|y^T, \beta^T, Q\right)$ and $p\left(Q^{-1}|y^T, \Sigma, \beta^T\right)$ using methods similar to those used in section on VAR independent Normal-Wishart model.

- Conditional on β^T, measurement equation is like a VAR with known coefficients.
- This leads to:

$$\Sigma^{-1}|y^T, \beta^T \sim W\left(\overline{S}^{-1}, \overline{\nu}\right)$$

- where

$$\overline{\nu} = T + \underline{\nu}$$

-

$$\overline{S} = \underline{S} + \sum_{t=1}^{T} \left(y_t - W_t\delta - Z_t\beta_t\right)\left(y_t - W_t\delta - Z_t\beta_t\right)'$$

- Conditional on β^T, state equation is also like a VAR with known coefficients.
- This leads to:

$$Q^{-1}|y^T, \beta^T \sim W\left(\overline{Q}^{-1}, \overline{\nu}_Q\right)$$

- where

$$\overline{\nu}_Q = T + \underline{\nu}_Q$$

- $$\overline{Q} = \underline{Q} + \sum_{t=1}^{T} \left(\beta_{t+1} - T_t\beta_t\right)\left(\beta_{t+1} - T_t\beta_t\right)'.$$

Optional Topic 2: DSGE Models

- DSGE = Dynamic, stochastic general equilibrium models popular in modern macroeconomics and commonly used in policy circles (e.g. central banks).
- I will not explain the macro theory, other than to note they are:
- Derived from microeconomic principles (based on agents and firms decision problems), dynamic (studying how economy evolves over time) and general equilibrium.
- Solution (using linear approximation methods) is a linear state space model
- Note: recent work with second order approximations yields nonlinear state space model
- Survey: An and Schorfheide (2007, Econometric Reviews)
- Computer code: http://www.dynare.org/ or some authors post code (e.g. code for Del Negro and Schorfheide 2008, JME on web)

Estimation Strategy for DSGE

- Most linearized DSGE models written as:

$$\Gamma_0(\theta) z_t = \Gamma_1(\theta) E_t(z_{t+1}) + \Gamma_2(\theta) z_{t-1} + \Gamma_3(\theta) u_t$$

- z_t is vector containing both observed variables (e.g. output growth, inflation, interest rates) and unobserved variables (e.g. technology shocks, monetary policy shocks).
- Note, theory usually written in terms of z_t as deviation of variable from steady state (an issue I will ignore here to keep exposition simple)
- θ are structural parameters (e.g. parameters for steady states, tastes, technology, policy and driving the exogenous shocks).
- u_t are structural shocks ($N(0, I)$).
- $\Gamma_j(\theta)$ are often highly nonlinear functions of θ

Solving the DSGE Model

- Methods exist to solve linear rational expectations models such as the DSGE
- If unique equilibrium exists can be written as:

$$z_t = A(\theta) z_{t-1} + B(\theta) u_t$$

- Looks like a VAR, but....
- Some elements of z_t typically unobserved
- and highly nonlinear restrictions involved in $A(\theta)$ and $B(\theta)$

Write DSGE Model as State Space Model

- Let y_t be elements of z_t which are observed.
- Measurement equation:

$$y_t = Cz_t$$

  where C is matrix which picks out observed elements of z_t
- Equation on previous slide is state equation in states z_t
- Thus we have state space model
- Special case since measurement equation has no errors (although measurement errors sometimes added) and state equation has some states which are observed.
- But state space algorithms described earlier in this lecture still work
- Remember, before I said: "for given values for system matrices, posterior simulators for the states exist"
- If θ were known, DSGE model provides system matrices in Normal linear state space model

Estimating the Structural Parameters

- If $A(\theta)$ and $B(\theta)$ involved simple linear restrictions, then methods similar to those for the restricted VAR (see Lecture 2) could be used to carry out inference on θ.
- Unfortunately, restrictions in $A(\theta)$ and $B(\theta)$ are typically nonlinear and complicated
- Parameters in θ are structural so we are likely to have prior information about them
- Example from Del Negro and Schorfheide (2008, JME):
- "Household-level data on wages and hours worked could be used to form a prior for a labor supply elasticity"
- "Product level data on price changes could be the basis for a price-stickiness prior"

Estimating the Structural Parameters (cont.)

- Prior for structural parameters, $p(\theta)$, can be formed from other sources of information (e.g. micro studies, economic theory, etc.)
- Here: prior times likelihood is a mess
- Thus, no analytical posterior for θ, no Gibbs sampler, etc...
- Solution: Metropolis-Hastings algorithm

Optional Topic 3: The Metropolis-Hastings Algorithm

- For now, I leave DSGE and state space models and return to our general notation:
- θ is a vector of parameters and $p(y|\theta)$, $p(\theta)$ and $p(\theta|y)$ are the likelihood, prior and posterior, respectively.
- Metropolis-Hastings algorithm takes draws from a convenient *candidate generating density*.
- Let θ^* indicate a draw taken from this density which we denote as $q\left(\theta^{(s-1)};\theta\right)$.
- Notation: θ^* is a draw taken of the random variable θ whose density depends on $\theta^{(s-1)}$.

- We are drawing the wrong distribution, $q\left(\theta^{(s-1)};\theta\right)$, instead of $p\left(\theta|y\right)$
- We have to correct for this somehow.
- Metropolis-Hastings algorithm corrects for this via an acceptance probability
- Takes candidate draws, but only some of these candidate draws are accepted.

- The Metropolis-Hastings algorithm takes following form:
- *Step 1:* Choose a starting value, $\theta^{(0)}$.
- *Step 2:* Take a candidate draw, θ^* from the candidate generating density, $q\left(\theta^{(s-1)};\theta\right)$.
- *Step 3:* Calculate an acceptance probability, $\alpha\left(\theta^{(s-1)},\theta^*\right)$.
- *Step 4:* Set $\theta^{(s)}=\theta^*$ with probability $\alpha\left(\theta^{(s-1)},\theta^*\right)$ and set $\theta^{(s)}=\theta^{(s-1)}$ with probability $1-\alpha\left(\theta^{(s-1)},\theta^*\right)$.
- *Step 5:* Repeat Steps 1, 2 and 3 S times.
- *Step 6:* Take the average of the S draws $g\left(\theta^{(1)}\right),...,g\left(\theta^{(S)}\right)$.

- These steps will yield an estimate of $E\left[g(\theta)|y\right]$ for any function of interest.
- Note: As with Gibbs sampling, Metropolis-Hastings algorithm requires the choice of a starting value, $\theta^{(0)}$. To make sure that the effect of this starting value has vanished, wise to discard S_0 initial draws.
- Intuition for acceptance probability, $\alpha\left(\theta^{(s-1)},\theta^{*}\right)$, given in textbook (pages 93-94).

$$\alpha\left(\theta^{(s-1)},\theta^{*}\right) = \min\left[\frac{p(\theta=\theta^{*}|y)q\left(\theta^{*};\theta=\theta^{(s-1)}\right)}{p\left(\theta=\theta^{(s-1)}|y\right)q\left(\theta^{(s-1)};\theta=\theta^{*}\right)},1\right]$$

Optional Topic 4: Choosing a Candidate Generating Density

- Independence Chain Metropolis-Hastings Algorithm
- Uses a candidate generating density which is independent across draws.
- That is, $q\left(\theta^{(s-1)};\theta\right)=q^{*}(\theta)$ and the candidate generating density does not depend on $\theta^{(s-1)}$.
- Useful in cases where a convenient approximation exists to the posterior. This convenient approximation can be used as a candidate generating density.
- Acceptance probability simplifies to:

$$\alpha\left(\theta^{(s-1)},\theta^{*}\right)=\min\left[\frac{p\left(\theta=\theta^{*}|y\right)q^{*}\left(\theta=\theta^{(s-1)}\right)}{p\left(\theta=\theta^{(s-1)}|y\right)q^{*}\left(\theta=\theta^{*}\right)},1\right].$$

- Not popular with DSGE models since convenient approximation unlikely to exist

- Random Walk Chain Metropolis-Hastings Algorithm
- Popular with DSGE – useful when you cannot find a good approximating density for the posterior.
- No attempt made to approximate posterior, rather candidate generating density is chosen to wander widely, taking draws proportionately in various regions of the posterior.
- Generates candidate draws according to:

$$\theta^* = \theta^{(s-1)} + w$$

where w is called the *increment random variable*.

- Acceptance probability simplifies to:

$$\alpha\left(\theta^{(s-1)}, \theta^{*}\right) = \min\left[\frac{p\left(\theta = \theta^{*}|y\right)}{p\left(\theta = \theta^{(s-1)}|y\right)}, 1\right]$$

- Choice of density for w determines form of candidate generating density.
- Common choice is Normal:

$$q\left(\theta^{(s-1)}; \theta\right) = f_N(\theta|\theta^{(s-1)}, \Sigma).$$

- Researcher must select Σ. Should be selected so that the acceptance probability tends to be neither too high nor too low.
- There is no general rule which gives the optimal acceptance rate. A rule of thumb is that the acceptance probability should be roughly 0.5.
- A common approach sets $\Sigma = c\Omega$ where c is a scalar and Ω is an estimate of posterior covariance matrix of θ (e.g. the inverse of the Hessian evaluated at the posterior mode)

- Popular (e.g. DYNARE) to use random walk Metropolis-Hastings with DSGE models.
- Note acceptance probability depends only on posterior = prior times likelihood
- DSGE Prior chosen as discussed above
- Algorithms for Normal linear state space models evaluate likelihood function

Metropolis-within-Gibbs

- Remember: the Gibbs sampler involved sequentially drawing from $p\left(\theta_{(1)}|y,\theta_{(2)}\right)$ and $p\left(\theta_{(2)}|y,\theta_{(1)}\right)$.
- Using a Metropolis-Hastings algorithm for either (or both) of the posterior conditionals used in the Gibbs sampler, $p\left(\theta_{(1)}|y,\theta_{(2)}\right)$ and $p\left(\theta_{(2)}|y,\theta_{(1)}\right)$, is perfectly acceptable.
- This statement is also true if the Gibbs sampler involves more than two blocks.
- Such *Metropolis-within-Gibbs* algorithms are common since many models have posteriors where most of the conditionals are easy to draw from, but one or two conditionals do not have convenient form.

Nonlinear State Space Models

- Normal linear state space model useful for empirical macroeconomists
- E.g. trend-cycle decompositions, TVP-VARs, linearized DSGE models, etc.
- Some models have y_t being a nonlinear function of the states (e.g. DSGE models which have not been linearized)
- Increasing number of Bayesian tools for nonlinear state space models (e.g. the particle filter)
- Here we will focus on stochastic volatility

Univariate Stochastic Volatility

- Begin with y_t being a scalar (common in finance)
- Stochastic volatility model:

$$y_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t$$

- $$h_{t+1} = \mu + \phi\left(h_t - \mu\right) + \eta_t$$

- ε_t is i.i.d. $N\left(0, 1\right)$ and η_t is i.i.d. $N\left(0, \sigma_\eta^2\right)$. ε_t and η_s are independent.
- This is state space model with states being h_t, but measurement equation is not a linear function of h_t

- h_t is log of the variance of y_t (log volatility)
- Since variances must be positive, common to work with log-variances
- Note μ is the unconditional mean of h_t.
- Initial conditions: if $|\phi| < 1$ (stationary) then:

$$h_0 \sim N\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right)$$

- if $\phi = 1$, μ drops out of the model and However, when $\phi = 1$, need a prior such as $h_0 \sim N\left(\underline{h}, \underline{V}_h\right)$
- e.g. Primiceri (2005) chooses $\underline{V}_h$ using training sample

MCMC Algorithm for Stochastic Volatility Model

- MCMC algorithm involves sequentially drawing from $p\left(h^T|y^T,\mu,\phi,\sigma_\eta^2\right)$, $p\left(\phi|y^T,\mu,\sigma_\eta^2,h^T\right)$, $p\left(\mu|y^T,\phi,\sigma_\eta^2,h^T\right)$ and $p\left(\sigma_\eta^2|y^T,\mu,\phi,h^T\right)$
- Last three standard forms based on results from Normal linear regression model and will not present here.
- Several algorithms exist for $p\left(h^T|y^T,\mu,\phi,\sigma_\eta^2\right)$
- Here we describe a popular one from Kim, Shephard and Chib (1998, ReStud)
- For complete details, see their paper. Here we outline ideas.

- Square and log the measurement equation:

$$y_t^* = h_t + \varepsilon_t^*$$

- where $y_t^* = \ln\left(y_t^2\right)$ and $\varepsilon_t^* = \ln\left(\varepsilon_t^2\right)$.
- Now the measurement equation is linear so maybe we can use algorithm for Normal linear state space model?
- No, since error is no longer Normal (i.e. $\varepsilon_t^* = \ln\left(\varepsilon_t^2\right)$)
- Idea: use mixture of different Normal distributions to approximate distribution of ε_t^*.

- Mixtures of Normal distributions are very flexible and have been used widely in many fields to approximate unknown or inconvenient distributions.
- $$p\left(\varepsilon_t^*\right) \approx \sum_{i=1}^{7} q_i f_N\left(\varepsilon_t^* | m_i, v_i^2\right)$$
- where $f_N\left(\varepsilon_t^* | m_i, v_i^2\right)$ is the p.d.f. of a $N\left(m_i, v_i^2\right)$
- since ε_t is $N(0,1)$, ε_t^* involves no unknown parameters
- Thus, q_i, m_i, v_i^2 for $i = 1, .., 7$ are not parameters, but numbers (see Table 4 of Kim, Shephard and Chib, 1998).

- Mixture of Normals can also be written in terms of component indicator variables, $s_t \in \{1, 2, .., 7\}$
- $$\begin{aligned} \varepsilon_t^* | s_t = i &\sim N\left(m_i, v_i^2\right) \\ \Pr\left(s_t = i\right) &= q_i \end{aligned}$$
- MCMC algorithm does not draw from $p\left(h^T | y^T, \mu, \phi, \sigma_\eta^2\right)$, but from $p\left(h^T | y^T, \mu, \phi, \sigma_\eta^2, s^T\right)$.
- But, conditional on s^T, knows which of the Normals ε_t^* comes from.
- Result is a Normal linear state space model and familiar algorithm can be used.
- Finally, need $p\left(s^T | y^T, \mu, \phi, \sigma_\eta^2, h^T\right)$ but this has simple form (see Kim, Shephard and Chib , 1998)

Multivariate Stochastic Volatility

- y_t is now $M \times 1$ vector and ε_t is i.i.d. $N(0, \Sigma_t)$.
- Many ways of allowing Σ_t to be time-varying
- But must worry about overparameterization problems
- Σ_t for $t = 1, .., T$ contains $\frac{TM(M+1)}{2}$ unknown parameters
- Here we discuss three particular approaches popular in macroeconomics
- To focus on multivariate stochastic volatility, use model:

$$y_t = \varepsilon_t$$

Multivariate Stochastic Volatility Model 1

- $$\Sigma_t = D_t$$
- where D_t is a diagonal matrix with diagonal elements d_{it}
- d_{it} has standard univariate stochastic volatility specification
- $d_{it} = \exp\left(h_{it}\right)$ and

$$h_{i,t+1} = \mu_i + \phi_i \left(h_{it} - \mu_i\right) + \eta_{it}$$

- if η_{it} are independent (across both i and t) then Kim, Shephard and Chib (1998) MCMC algorithm can be used one equation at a time.
- But many interesting macroeconomic features (e.g. impulse responses) depend on error covariances so assuming Σ_t to be diagonal often will be a bad idea.

Multivariate Stochastic Volatility Model 2

- Cogley and Sargent (2005, RED)
- $$\Sigma_t = L^{-1} D_t L^{-1\prime}$$
- D_t is as in Model 1 (diagonal matrix with diagonal elements being variances)
- L is a lower triangular matrix with ones on the diagonal.
- E.g. $M = 3$
$$L = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix}$$

- We can transform model as:

$$Ly_t = L\varepsilon_t$$

- $\varepsilon_t^* = L\varepsilon_t$ will now have a diagonal covariance matrix – can use algorithm for Model 1.
- MCMC algorithm: $p\left(h^T|y^T, L\right)$ can use Kim, Shephard and Chib (1998) algorithm one equation at a time.
- $p\left(L|y^T, h^T\right)$ results similar to those from a series of M regression equations with independent Normal errors.
- See Cogley and Sargent (2005) for details.

- Cogley-Sargent model allows the covariance between errors to change over time, but in restricted fashion.
- E.g. $M = 2$ then $cov(\varepsilon_{1t}, \varepsilon_{2t}) = d_{1t}L_{21}$ which varies proportionally with the error variance of the first equation.
- Impulse response analysis: a shock to i^{th} variable has an effect on j^{th} variable which is constant over time
- In many macroeconomic applications this is too restrictive.

Multivariate Stochastic Volatility Model 3

- Primiceri (2005, ReStud):

$$\Sigma_t = L_t^{-1} D_t L_t^{-1\prime}$$

- L_t is same as Cogley-Sargent's L but is now time varying.
- Does not restrict Σ_t in any way.
- MCMC algorithm same as for Cogley-Sargent except for L_t

- How does L_t evolve?
- Stack unrestricted elements by rows into a $\frac{M(M-1)}{2}$ vector as $l_t = \left(L_{21,t}, L_{31,t}, L_{32,t}, .., L_{p(p-1),t}\right)'$.
-
$$l_{t+1} = l_t + \zeta_t$$
- ζ_t is i.i.d. $N\left(0, D_\zeta\right)$ and D_ζ is a diagonal matrix.
- Can transform model so that algorithm for Normal linear state space model can draw l_t
- See Primiceri (2005) for details
- Note: if D_ζ is not diagonal have to be careful (no longer Normal state space model)

Summary

- MCMC algorithms such as the Gibbs sampler are modular in nature (sequentially draw from blocks)
- By combining simple blocks together you can end up with very flexible models
- This is strategy pursued here.
- Our MCMC algorithms for complicated models all combine simpler algorithms.
- E.g. Primiceri's complicated model involves blocks which use Carter and Kohn's algorithm and blocks which use Kim, Shephard and Chib's algorithm (and even the latter relies upon Carter and Kohn)