Introduction to Bayesian State Space Models

Introduction

- State space methods are used for a wide variety of time series problems
- They are important in and of themselves in economics (e.g. trend-cycle decompositions, structural time series models, dealing with missing observations, etc.)
- Also time-varying parameter VARs (TVP-VARs) and stochastic volatility are state space models
- DSGE models are state space models (DYNARE popular Bayesian code for estimation)
- Advantage of state space models: well-developed set of MCMC algorithms for doing Bayesian inference

- Normal linear state space model:

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

- where

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

- TVP-VAR has Z_t containing lags of dependent variables and β_t being VAR coefficients
- But unlike VAR of previous lecture, VAR coeffs are varying over time
- 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).
- Precision based sampler of Joshua Chan (http://joshuachan.org/)
- I will not present details of these (standard) algorithms
- These algorithms involve use of methods called Kalman filtering and smoothing
- Filtering = estimating a state at time t using data up to time t
- Smoothing = estimating a state at time t using data up to time T

- 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:

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

- 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

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

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

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, $s_t \in \{1, 2, .., 7\}$
- $$\varepsilon_t^* | s_t = i \sim N\left(m_i, v_i^2\right)$$
  $$\Pr\left(s_t = i\right) = q_i$$
- 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)