TVP-VARs with Stochastic Volatility

TVP-VARs

- 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.)

State Space Models

- State space methods are used for a wide variety of time series problems, including TVP-VARs
- Provide an introduction to them before getting to TVP-VAR
- 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

- Remember: our general notation for a VAR was:

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

- In many macroeconomic applications, constant β is unrealistic
- 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).
- 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: $\beta^t = \left(\beta_1', .., \beta_t'\right)'$ 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

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

-

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

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, $s_t \in \{1, 2, .., 7\}$
- $$\begin{array}{l} \varepsilon_t^* | s_t = i \sim N\left(m_i, v_i^2\right) \\ \Pr\left(s_t = i\right) = q_i \end{array}$$
- 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
- $d_{it} = \exp(h_{it})$ and

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

- 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 $cov\left(\varepsilon_{1t}, \varepsilon_{2t}\right) = d_{1t}L_{21}$ 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)

More Uses of State Space Models

- 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.
- For state space models there are a standard set of algorithms which can be combined together in various ways to produce quite sophisticated models
- Our MCMC algorithms for complicated models all combine simpler algorithms.
- Now see how this works with TVP-VARs

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
- $V\left(\beta_{OLS}\right)$ is estimated covariance of β_{OLS}.
- Prior for β_0:
$$\beta_0 \sim N\left(\beta_{OLS}, 4 \cdot V\left(\beta_{OLS}\right)\right)$$
- Prior for Σ^{-1} Wishart prior with $\underline{\nu} = M + 1, \underline{S} = I$
- Prior for Q^{-1} Wishart prior with $\underline{\nu}_Q = 40, \underline{Q} = 0.0001 \cdot 40 \cdot V\left(\beta_{OLS}\right)$

- 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.

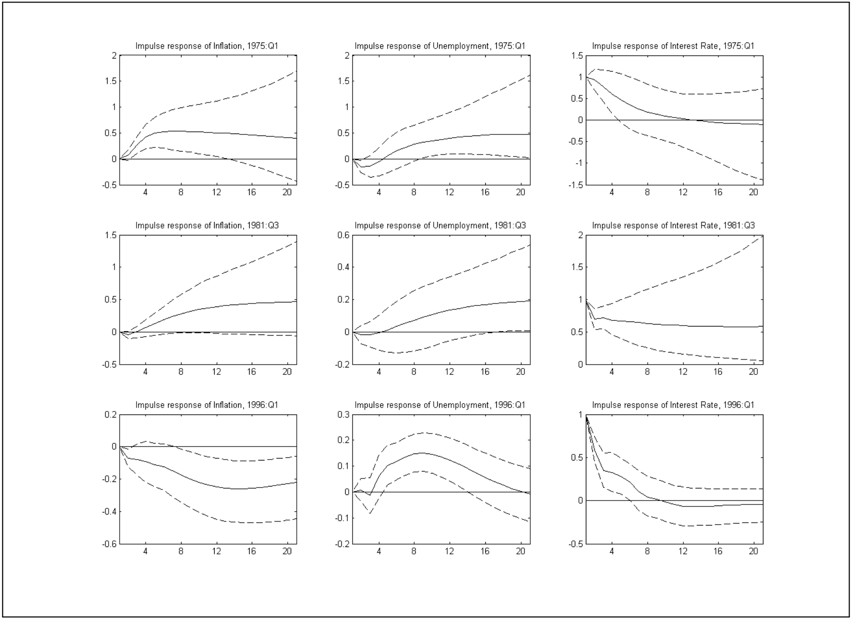

Figure 1: Impulse responses at different times

TVP-VARs with Stochastic Volatility

- In empirical work, you will usually want to add multivariate stochastic volatility to the TVP-VAR
- 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)

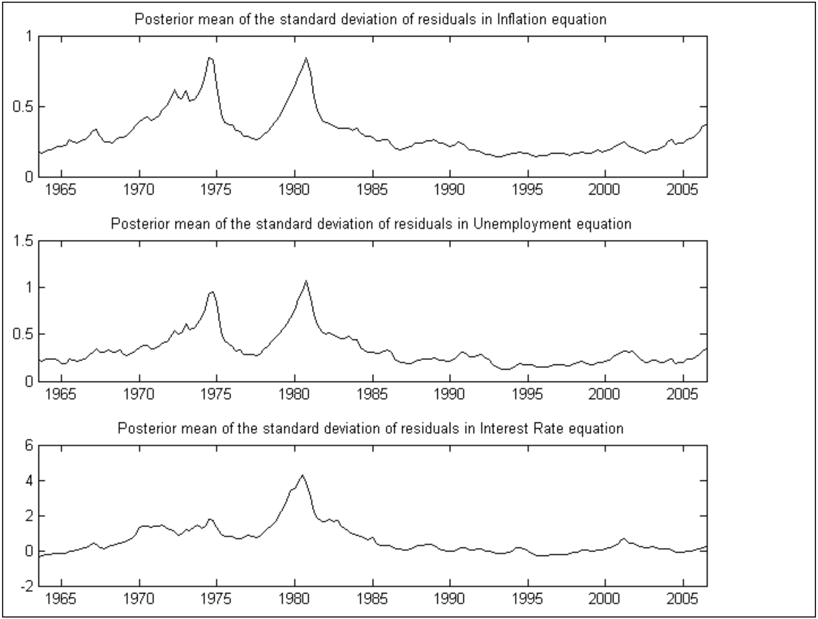

Figure 2: Volatilities in the 3 Equations

Summary of TVP-VARs

- 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 familiar MCMC algorithms developed for state space models.
- Much recent work on shrinkage priors for TVP-VARs to avoid over-parameterization concerns