Bayesian Estimation of a TVP-VAR Model

Thomas Stockwell

5/30/2018

Normal VAR Model

- The normal VAR is expressed in this way:

$$Y_t = Y_{t-1}\beta + \epsilon_t$$

- where:
  - Y_t is a Mx1 vector of our data
  - Y_{t-1} is a Mxk matrix containing exogenous explanatory variables (contains lagged values of Y and the constant term)
  - β is a Mx1 vector of parameters to be estimated
    - Assumed to be constant over time
  - ϵ_t is a Mx1 vector of disturbance terms
    - Assume: $\epsilon_t \sim N(0, \Sigma)$
    - Σ is assumed constant over time

Shortcoming of a VAR Model

- In practice, the VAR model may not be very practical
- Its restrictive in two main ways:
  - The parameters of the model may vary over time
  - The variance of the exogenous shocks may vary over time
- Example: US Monetary Policy and the high inflation and low growth in the 1970's
  - Changing parameters: The Fed may have changed its response to inflation during this time
  - Changing variance: The '70's had high volatility whereas later policy makers experienced the Great Moderation

The TVP-VAR Model

- We can use the TVP-VAR to address both of these issues
- TVP: Time Varying Parameters
- We can allow the parameters of the model to change over time or the variance of the disturbance term (or both)
- We will assume that errors are homoskedastic for simplicity and only focus on the time varying parameters

The TVP-VAR Model

- The TVP-VAR is expressed in this way:

$$Y_t = Y_{t-1}\beta_t + \epsilon_t$$

- We let the parameters evolve according to an AR(1) process:

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

- All the variables have the same definitions that we saw in the normal VAR case
- The new variable is $u_t \sim N(0, Q)$
- Do you notice anything familiar about the setup of this model?

The TVP-VAR Model

- This is a state-space model:
  - The measurement equation is: $Y_t = Y_{t-1}\beta_t + \epsilon_t$
  - The transition equation is: $\beta_t = \beta_{t-1} + u_t$
- When we think of Bayesian estimation of this model, we can use the same techniques that we learned in class when estimating state-space models

Bayesian Estimation

- Our parameters for this model: Σ, Q
  - Well call these θ
- Unknown objects of interest:
  - θ
  - β_t
- Our goal is to sample from $P(\theta, \tilde{\beta}_T | \tilde{Y}_T)$

MCMC Algorithm

- We will use a 3-step Gibbs sampler to accomplish our goal:
  - Step 1: Draw Σ from $P(\Sigma|\tilde{\beta}_T^{g-1}, \tilde{Y}_T)$, gives us Σ^g
  - Step 2: Draw Q from $P(Q|\tilde{\beta}_T^{g-1}, \tilde{Y}_T)$, gives us Q^g
  - Step 3: Draw $\tilde{\beta}_T$ from $P(\tilde{\beta}_T|\theta^g, \tilde{Y}_T)$, gives us $\tilde{\beta}_T^g$
- Step 3 is exactly what Jeremy did in the state-space model notes
- Steps 1 and 2 are the ones we did not do in class, but the Koop and Korobilis book goes over this
  - Start with step 1:

Sampling $P(\theta|\tilde{\beta}_T^{g-1}, \tilde{Y}_T)$

- For Σ^{-1}, our prior is a $Wishart(S, v)$
- After combining this with the marginal likelihood, the conditional posterior is:

$$P(\Sigma^{-1}|\tilde{\beta}_T, \tilde{Y}_T) \sim Wishart(\bar{S}^{-1}, \bar{v})$$

$$\bar{v} = v + T$$

$$\bar{S}^{-1} = S + \sum_{t=2}^{T}(Y_t - Y_{t-1}^{'}\beta_t)(Y_t - Y_{t-1}^{'}\beta_t)'$$

  - "v" and "S" are hyperparameters

Sampling $P(\theta|\tilde{\beta}_T^{g-1}, \tilde{Y}_T)$

- For Q^{-1}, our prior is also a $Wishart(S_Q, v_Q)$
- After combining this with the marginal likelihood, the conditional posterior is:

$$P(Q^{-1}|\tilde{\beta}_T, \tilde{Y}_T) \sim Wishart(\bar{S}_Q^{-1}, \bar{v}_Q)$$

$$\bar{v}_Q = v_Q + T$$

$$\bar{S}_Q^{-1} = S_Q + \sum_{t=1}^{T} (\beta_{t+1} - I\beta_t)(\beta_{t+1} - I\beta_t)'$$

  - "v_Q" and "S_Q" are hyperparameters

Sampling $P(\tilde{\beta}_T|\theta^g, \tilde{Y}_T)$

- To sample this distribution, we will follow the Carter and Kohn (1994) that we did in the state-space model notes

$$P(\tilde{\beta}_T|\theta, \tilde{Y}_T) = P(\beta_1, ..., \beta_T|\theta, \tilde{Y}_T)$$

$$= P(\beta_1|\beta_2, ..., \beta_T, \theta, \tilde{Y}_T)...P(\beta_{T-1}|\beta_T, \theta, \tilde{Y}_T)P(\beta_T|\theta, \tilde{Y}_T)$$

Sampling $P(\tilde{\beta}_T|\theta^g, \tilde{Y}_T)$

- Recall that β_t is an AR(1) process with an iid disturbance vector, therefore:
  - Given β_{t+1}, the pdf of β_t does not depend on $\beta_{t+2}, ..., \beta_T$
  - Conditional on $\tilde{Y}_T$, then $Y_{t+1}, ..., Y_T$ contains no additional information about β_t
- This means that:

$$P(\tilde{\beta}_T|\theta, \tilde{Y}_T) = P(\beta_1|\beta_2, \theta, \tilde{Y}_1)...P(\beta_{T-1}|\beta_T, \theta, \tilde{Y}_{T-1})P(\beta_T|\theta, \tilde{Y}_T)$$

Sampling $P(\tilde{\beta}_T|\theta^g, \tilde{Y}_T)$

- This means we can draw from $P(\tilde{\beta}_T|\theta, \tilde{Y}_T)$ as follows:
  - Step 1: Draw β_T from $P(\beta_T|\theta, \tilde{Y}_T)$
  - Step 2: Draw β_{T-1} from $P(\beta_{T-1}|\beta_T, \theta, \tilde{Y}_{T-1})$

$$\vdots$$

  - Step T: Draw β_1 from $P(\beta_1|\beta_2, \theta, \tilde{Y}_1)$
- If we need it, we can also draw β_0 from $P(\beta_0|\beta_1, \theta, \tilde{Y}_0)$

Sampling $P(\tilde{\beta}_T|\theta^g, \tilde{Y}_T)$

- How do we take these draws?
- From our time-series class, we know that:

$$P(\beta_t|\theta, \tilde{Y}_t) = N(\beta_{tt}, V_{tt})$$

$$\beta_{tt} = E(\beta_t|\theta, \tilde{Y}_t)$$

$$V_{tt} = Var(\beta_t|\theta, \tilde{Y}_t)$$

- β_{tt} and V_{tt} we get from the Kalman Filter
  - Therefore, we can get β_{TT} and V_{TT} from the final step of the Kalman Filter and then draw X_T from $N(\beta_{TT}, V_{TT})$

Sampling $P(\tilde{\beta}_T|\theta^g, \tilde{Y}_T)$

- What about the rest of the steps 2-T?

$$P(\beta_t|\beta_{t+1}, \theta, \tilde{Y}_t) = N(E(\beta_t|\beta_{t+1}, \theta, \tilde{Y}_t), Var(\beta_t|\beta_{t+1}, \theta, \tilde{Y}_t))$$

$$E(\beta_t|\beta_{t+1}, \theta, \tilde{Y}_t) = \beta_{tt} + V_{tt}(V_{tt} + Q)^{-1}(\beta_{t+1} - \beta_{tt})$$

$$Var(\beta_t|\beta_{t+1}, \theta, \tilde{Y}_t) = V_{tt} - V_{tt}(V_{tt} + Q)^{-1}$$

Other Items of Note

- Training Sample
  - Designating an early part of your data to estimate what your priors should be
- Minnesota Prior
  - Rather than have the TE be AR(1), we can combine it with the Minnesota Prior:

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

  - "A" is something we can set or we could add another block into our Gibbs sampler