Timevarying VARs

Wouter J. Den Haan
London School of Economics

Time-Varying VARs

- Gibbs-Sampler
  - general idea
  - probit regression application
- (Inverted) Wishart distribution
- Drawing from a multi-variate Normal in Matlab
- Time-varying VAR
  - model specification
  - Gibbs sampler

Gibbs Sampler

Suppose

- x, y, and z are distributed according to $f(x, y, z)$
- Suppose that drawing x, y, and z from $f(x, y, z)$ is difficult
- *but*
  you can draw x from $f(x|y, z)$ and
  you can draw y from $f(y|x, z)$ and
  you can draw y from $f(z|x, y)$

Gibbs Sampler - how it works

- Start with y_0, z_0
- Draw x_1 from $f(x|y_0, z_0)$,
- Draw y_1 from $f(y|x_1, z_0)$,
- Draw z_1 from $f(z|x_1, y_1)$,
- Draw x_2 from $f(x|y_1, z_1)$

- (x_i, y_i, z_i) is one draw from the joint density $f(x, y, z)$
- Although series are constructed recursively, they are *not* time series

Gibbs Sampler - convergence

- The idea is that this sequence converges to a sequence drawn from $f(x, y, z)$.
- Since convergence is not immediate, you have to discard beginning of sequence (burn-in period).
- See Casella and George (1992) for a discussion on why and when this works.

Gibbs Sampler - probit regression

This example is from Lancaster (2004)

- y_i is the i^{th} observation of a binary variable, i.e., $y_i \in \{0, 1\}$
- y_i^* is an unobservable and given by

$$y_i^* = x_i \beta + \varepsilon_i, \;\; \varepsilon_i \sim N(0, 1)$$

- $$y = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{o.w.} \end{cases}$$

Probit regression

- Parameters: β and $y^* = [y_1^*, y_2^*, \cdots, y_n^*]'$
- Data: $X = [x_1, x_2, \cdots, x_n]', Y = \{y_i, x_i\}_{i=1}^n$
- Objective: get $p\left(\widehat{\beta}|Y\right)$, i.e., the distribution of $\widehat{\beta}$ given Y.
- With the Gibbs sample we can get a sequence of obervations for $\left(\widehat{\beta}, \widehat{y^*}\right)$ distributed according to $p\left(\widehat{\beta}, \widehat{y}^*|Y\right)$, from which we can get $p\left(\widehat{\beta}|Y\right)$

Probit - Gibb sampler step 1

We need to draw from $p\left(\widehat{\beta}|y^*, Y\right)$

- Given y^* and X

$$\widehat{\beta} \sim N\left(\left(X'X\right)^{-1} X'y^*, \left(X'X\right)^{-1}\right),$$

since the standard deviation of ε_i is known and equal to 1.

Probit - Gibb sampler step 2

We need to draw from $p\left(y^{*}|\beta,\Upsilon\right)$

- Since the y_is are independent, we can do this separately for each i

$$\begin{array}{ll} y_i^{*} \sim N_{>0}\left(x_i\beta,1\right) & \text{if } y_i = 1 \\ y_i^{*} \sim N_{<0}\left(x_i\beta,1\right) & \text{if } y_i = 0 \end{array} ,$$

where
$N_{>0}\left(\cdot\right)$ is a Normal distribution truncated on the left at 0
$N_{<0}\left(\cdot\right)$ is a Normal distribution truncated on the right at 0

Wishart distribution

- generalization of Chi-square distribution to more variables
- $X : n \times p$ matrix; each row drawn from $N_p(0, \Sigma)$, where Σ is the $p \times p$ variance-covariance matrix
- $W = X'X \sim W_p(\Sigma, n)$, i.e., the p-dimensional Wishart with scale matrix Σ and degrees of freedom n
- You get the Chi-square if $p = 1$ and $\Sigma = 1$

Inverse Wishart distribution

- If W has a Wishart distribution with parameters Σ and n, then W^{-1} has an inverse Wishart with scale matrix Σ^{-1} and degrees of freedom n
- !!! In the assignment, the input of the Matlab Inverse Wishart function is Σ not Σ^{-1}.

Inverse Wishart in Bayesian statistics

- Data: x_t is a $p \times 1$ vector with i.i.d. random observations with distribution $N(0, V)$
- prior of V :

$$p(V) = IW\left(\overline{V}^{-1}, n\right)$$

- posterior of V :

$$\begin{aligned} p\left(V|X^T\right) &= IW\left(W^{-1}, n+T\right) \\ W &= \overline{V} + \widehat{V}_T \\ \widehat{V}_T &= \sum_{t=1}^{T} x_t' x_t \end{aligned}$$

Note that $\widehat{V}_T$ is like a sum of squares

Multivariate normal in Matlab

- x_t is a $p \times 1$ vector and we want $x_t \sim N(0, \Sigma)$
- `C=chol(Σ)`
  Thus C is an upper-triangular matrix and $C'C = \Sigma$
- e_t is a $p \times 1$ vector with draws from $N(0, I_p)$

$$\mathbb{E}\left[C'e_te_t'C\right] = \Sigma$$

- Thus, $C'e_t$ is a $p \times 1$ vector with draws from $N(0, \Sigma)$

Time-varying VARs - intro

- Idea: capture changes in model specification in a flexible way
- The analysis here is based on Cogley and Sargent (2002), CS

Model specification

$$
\begin{aligned}
y_t &= X_t'\theta_t + \varepsilon_t \\
X_t' &= \left[1, y_{t-1}, y_{t-2}, \cdots, y_{t-p}\right] \\
\theta_t &= \theta_{t-1} + v_t
\end{aligned}
$$

$$
\begin{aligned}
\varepsilon_t &\sim N(0, R) \\
v_t &\sim N(0, Q)
\end{aligned}
$$

Model specification

$$\mathbb{E}_t \begin{bmatrix} \varepsilon_t \\ v_t \end{bmatrix} \begin{bmatrix} \varepsilon_t & v_t \end{bmatrix} = V = \begin{pmatrix} R & C' \\ C & Q \end{pmatrix}$$

- θ_t : "parameters"
- R, C, and Q are the "hyperparameters"

Model specification - details

Simplifying assumptions:

- CS impose that θ_t is such that y_t would be stationary if $\theta_{t+\tau} = \theta_t$ for all $\tau \geq 0$. This stationarity requirement is left out for transparency.
- $C = 0$.

Notation

$$
\begin{aligned}
Y^T &= [y_1', \cdots, y_T'] \\
\theta^T &= [\theta_0', \theta_1', \cdots, \theta_T']
\end{aligned}
$$

Priors

- Prior for initial condition:

$$\theta_0 \sim N\left(\overline{\theta}, \overline{P}\right)$$

- Prior for hyperparameters:

$$p(V) = IW\left(\overline{V}^{-1}, T_0\right)$$

- $\overline{\theta}, \overline{P}, \overline{V}, T_0$ are taken as given

Posterior

- The posterior is given by

$$p\left(\theta^T, V|Y^T\right)$$

- We can use the Gibbs sampler if we can draw from

$$P\left(\theta^T|Y^T, V\right)$$

and from

$$P\left(V|Y^T, \theta^T\right)$$

Stationarity

- CS exclude draws of θ_t for which the *dgp* of y_t is nonstationary:
  - $p\left(\theta_t|\cdot\right)$ is density without imposing stationarity and
  - $f\left(\theta_t|\cdot\right)$ is density with imposing stationarity
- This restriction is ignored in these slides

Gibbs part I: Posterior of theta given V

- Since $f(A,B) = f(A|B) \times f(B)$, we have

$$p\left(\theta^T|Y^T,V\right) = f\left(\theta^T|Y^T,V\right)$$

$$\begin{aligned}
&= f\left(\theta^{T-1}|\theta_T,Y^T,V\right) \times f\left(\theta_T|Y^T,V\right) \\
&= f\left(\theta^{T-2}|\theta_T,\theta_{T-1},Y^T,V\right) \times f\left(\theta_{T-1}|\theta_T,Y^T,V\right) \\
&\quad \times f\left(\theta_T|Y^T,V\right) \\
&= f\left(\theta^{T-3}|\theta_T,\theta_{T-1},\theta_{T-2},Y^T,V\right) \times f\left(\theta_{t-2}|\theta_T,\theta_{T-1},Y^T,V\right) \\
&\quad \times f\left(\theta_{T-1}|\theta_T,Y^T,V\right) \times f\left(\theta_T|Y^T,V\right)
\end{aligned}$$

Posterior of theta given V

- Since

$$\theta_t = \theta_{t-1} + v_t,$$

  $\theta_{t+\tau}$ has no predictive power for θ_{t-1} for all $\tau \geq 1$ given Y^T and θ_t,

- Thus

$$\begin{aligned} f\left(\theta_{T-2}|\theta_T, \theta_{T-1}, Y^T, V\right) &= f\left(\theta_{T-2}|\theta_{T-1}, Y^T, V\right) \\ f\left(\theta_{T-3}|\theta_T, \theta_{T-1}, \theta_{T-2}, Y^T, V\right) &= f\left(\theta_{T-3}|\theta_{T-2}, Y^T, V\right) \\ &\text{etc.} \end{aligned}$$

Posterior of theta given V

- Combining gives

$$p\left(\theta^T | Y^T, V\right) = f\left(\theta_T | Y^T, V\right) \prod_{t=1}^{T-1} f\left(\theta_t | \theta_{t+1}, Y^T, V\right)$$

- All the densities are Gaussian $\Longrightarrow$ if we know the means and the variances, then we can draw from $p\left(\theta^T | Y^T, V\right)$

Posterior of theta given V

We need to find the means and variances of

$$f\left(\theta_T|Y^T,V\right)\ \& f\left(\theta_t|\theta_{t+1},Y^T,V\right)$$

Notation

$$\begin{aligned}
\theta_{t|t} &= \mathbb{E}\left(\theta_t|Y^t,V\right) \\
P_{t|t-1} &= VAR\left(\theta_t|Y^{t-1},V\right) \\
P_{t|t} &= VAR\left(\theta_t|Y^t,V\right) \\
\theta_{t|t+1} &= \mathbb{E}\left(\theta_t|\theta_{t+1},Y^t,V\right) = \mathbb{E}\left(\theta_t|\theta_{t+1},Y^T,V\right) \\
P_{t|t+1} &= VAR\left(\theta_t|\theta_{t+1},Y^t,V\right) = VAR\left(\theta_t|\theta_{t+1},Y^T,V\right)
\end{aligned}$$

Posterior of theta given V

- First, use Kalman filter to go forward
  - start with θ_0 and $P_{0|0}$
- Next, go backwards to get draws for θ_t given θ_{t+1}

Posterior of theta given V

- Kalman filter part:

$$
\begin{aligned}
y_t &= X_t'\theta_t + \varepsilon_t \\
X_t' &= \left[1, y_{t-1}, y_{t-2}, \cdots, y_{t-p}\right] \\
\theta_t &= \theta_{t-1} + v_t
\end{aligned}
$$

$$
\begin{aligned}
\varepsilon_t &\sim N(0, R) \\
v_t &\sim N(0, Q)
\end{aligned}
$$

- Here:
  - the $p+1$ elements of X_t are the known (time-varying) *coefficients* of the state-space represenation
  - the elements of θ_t are the unobserved underlying state variables

Posterior of theta given V

- Kalman filter in the first period:

$$
\begin{aligned}
P_{1|0} &= P_{0|0} + Q \\
K_1 &= P_{1|0} X_1 \left(X_1' P_{1|0} X_1 + R \right)^{-1} \\
\theta_{1|1} &= \theta_{0|0} + K_1 \left(y_1 - X_1' \theta_{0|0} \right)
\end{aligned}
$$

- and then iterate

$$
\begin{aligned}
P_{t|t-1} &= P_{t-1|t-1} + Q \\
K_t &= P_{t|t-1} X_t \left(X_t' P_{t|t-1} X_t + R \right)^{-1} \\
\theta_{t|t} &= \theta_{t-1|t-1} + K_t \left(y_t - X_t' \theta_{t-1|t-1} \right) \\
P_{t|t} &= P_{t|t-1} - K_t X_t' P_{t|t-1}
\end{aligned}
$$

Posterior of theta given V

- In the Kalman filter part of the assignment:

$$
\begin{aligned}
\text{TH(:,1)} &= \theta_0 \\
\text{TH(:,t+1)} &= \theta_{t|t} \\
\text{Pe(:,:,t)} &= P_{t-1|t-1} \\
\text{Po(:,:,t)} &= P_{t|t-1}
\end{aligned}
$$

- and we go up to

$$
\begin{aligned}
\text{TH(:,T+1)} &= \theta_{T|T} \\
\text{Pe(:,:,T+1)} &= P_{T|T} \\
\text{Po(:,:,T)} &= P_{T|T-1}
\end{aligned}
$$

Posterior of theta given V

- Distribution terminal state:

$$f\left(\theta_T | Y^T, V\right) = N\left(\theta_{T|T}, P_{T|T}\right)$$

- From this we get a draw θ_T

Posterior of theta given V

- Draws for $\theta_{T-1}, \theta_{T-2}, \cdots, \theta_1$ are obtained recursively from

$$
\begin{aligned}
f\left(\theta_t | \theta_{t+1}, Y^T, V\right) &= N\left(\theta_{t|t+1}, P_{t|t+1}\right) \\
\theta_{t|t+1} &= \theta_{t|t} + P_{t|t} P_{t+1|t}^{-1} \left(\theta_{t+1} - \theta_{t|t}\right) \\
P_{t|t+1} &= P_{t|t} - P_{t|t} P_{t+1|t}^{-1} P_{t|t}
\end{aligned}
$$

- The terms needed to calculate $\theta_{t|t+1}$ and $P_{t|t+1}$ are generated by the Kalman filter (that is, from going forward) and the standard projection formulas (and note that the covariance of θ_{t+1} and θ_t is the variance of θ_t)

Details for previous slide

$$
\begin{aligned}
\mathbb{E}\left[y|x\right] &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}\left(x-\mu_x\right) \\
&\Longrightarrow \\
\mathbb{E}\left[\theta_t|\theta_{t+1};\cdot\right] &= \mathbb{E}\left[\theta_t|\cdot\right] + \Sigma_{\theta_t\theta_{t+1}}\Sigma_{\theta_{t+1}\theta_{t+1}}^{-1}\left(\theta_{t+1}-\mathbb{E}\left[\theta_{t+1}|\cdot\right]\right) \\
&= \mathbb{E}\left[\theta_t|\cdot\right] + \Sigma_{\theta_t\theta_t}\Sigma_{\theta_{t+1}\theta_{t+1}}^{-1}\left(\theta_{t+1}-\mathbb{E}\left[\theta_{t+1}|\cdot\right]\right) \\
&\Longrightarrow \\
\theta_{t|t+1} &= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}\left(\theta_{t+1}-\mathbb{E}\left[\theta_t+v_t|\cdot\right]\right) \\
&= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}\left(\theta_{t+1}-\mathbb{E}\left[\theta_t|\cdot\right]\right) \\
&= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}\left(\theta_{t+1}-\theta_{t|t}\right)
\end{aligned}
$$

Suppressing the dependence on Y^t and V to simplify notation

Posterior of theta given V

In the backward part of the assignment:

```
Draw from TH(t-1|t)
In the for loop below t goes from high to low.
At a particular t:
```

1. `TH(:,t+1) it is a random draw from a normal that has already been determined (either in this loop or for T above)`
2. `TH(:,t) on the RHS of the mean equation is equal to theta_(t-1)|(t-1)`
3. `TH(:,t) what we end up with is a random draw for theta(t-1) conditional on  knowning theta in the next period`

Why go forward & backward?

- The Kalman filter gives us $\mathbb{E}\left(\theta_t|Y^t, V\right)$ and $VAR(\theta_t|Y^t, V)$
- With this information, we can also obtain draws for θ_t
- However, we need draws from $f\left(\theta^T|Y^T, V\right)$ not from $f\left(\theta^T|Y^t, V\right)$. The analysis above showed how to get draws from $f\left(\theta^T|Y^T, V\right)$ recursively by going backward.

Relation to Kalman Smoother

- The Kalman smoother also goes backwards and resembles the procedure here.
- However, there is a difference.
  - The Kalman smoother computes the mean and variance for $f\left(\theta_t | Y^T, V\right)$
  - We need the mean and variance for $f\left(\theta_t | \theta_{t+1}, Y^T, V\right)$
  - Since

$$f\left(\theta_t | \theta_{t+1}, Y^T, V\right) = f\left(\theta_t | \theta_{t+1}, Y^t, V\right),$$

    we can calculate these from Kalman filter without using the Kalman smoother

Gibbs part II: Posterior of V given theta

- Next step is to draw *from the posterior* given Y^T and θ^T, that is get a draw from $p\left(V|Y^T,\theta^T\right)$

- The posterior combines the prior and information from the data $\Longrightarrow$ in each Gibbs iteration the prior is the same but the data set (i.e., θ^T) is different

Gibbs part II: Posterior of V given theta

- Given Y^T and θ^T, we can calcluate ε_t and ν_t.
- Both have mean zero and a Normal distribution
  - Thus

$$\begin{aligned} p\left(V|Y^T,\theta^T\right) &= IW(V_1^{-1},T_1) \\ T_1 &= T_0 + T \\ V_1 &= \overline{V} + \overline{V}_T \\ \overline{V}_T &= \sum_{t=1}^{T} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \left(\varepsilon_t' \; v_t'\right) \end{aligned}$$

!!! Note that $\overline{V}, \overline{V}_T, \& \overline{V}_1$ are like a sum of squares, whereas V (and $R\&Q$) are like a sum of squares divided by number of observations (same notation as in CS)

References

- Casella, G., and E.I. George, 1992, Explaining the Gibbs Sampler, The American Statistician, 46 167-74.
- Cogley, T. and T.J. Sargent, 2001, Evolving Post-World War II U.S. Inflation Dynamics, NBER Macroeconomics Annual 2001, volume 16, 331-338.
- De Wind, J, 2014, Accounting for time-varying and nonlinear relationships in macroeconomic models, dissertation, University of Amsterdam. Available at http://dare.uva.nl/document/523663.
- Lancaster, T., 2004, An introduction to modern Bayesian econometrics, Blackwell Publishing.
  - very nice textbook covering lots of stuff in an understandable way