Bayesian Vector Autoregressive Models

Time Series Modelling for Empirical Macroeconomics

- Vector Autoregressive (VAR) models popular way of summarizing inter-relationships between macroeconomic variables.
- Used for forecasting, impulse response analysis, etc.
- Economy is changing over time. Is model in 1970s same as now?
- Thus, time-varying parameter VARs (TVP-VARs) are of interest.
- Great Moderation of business cycle leads to interest in modelling error variances
- TVP-VARs with multivariate stochastic volatility is our end goal.
- Begin with Bayesian VARs
- A common theme: These models are over-parameterized so need shrinkage to get reasonable results (shrinkage = prior).

Bayesian VARs

- VAR(p) model:

$$y_t = a_0 + \sum_{j=1}^{p} A_j y_{t-j} + \varepsilon_t$$

- y_t is $M \times 1$ vector
- ε_t is $M \times 1$ vector of errors
- a_0 is $M \times 1$ vector of intercepts
- A_j is an $M \times M$ matrix of coefficients.
- ε_t is i.i.d. $N(0, \Sigma)$.
- Exogenous variables or more deterministic terms can be added (but we don't to keep notation simple).

- Another way of writing VAR:
- Let Y and E be $T \times M$ matrices placing the T observations on each variable in columns next to one another.
- Then can write VAR as

$$Y = XA + E$$

- Notation: we will let α be $KM \times 1$ vector of VAR coefficients, where A is $K \times M$ (i.e. $K = Mp + 1$ is number of explanatory variables in each equation)
- $\alpha = vec(A)$

Likelihood Function

- Likelihood function can be derived and shown to be of a form that breaks into two parts
- First of these parts α given Σ and another for Σ
- $$\alpha|\Sigma, y \sim N\left(\widehat{\alpha}, \Sigma \otimes \left(X'X\right)^{-1}\right)$$
- Σ^{-1} has Wishart form

$$\Sigma^{-1}|y \sim W\left(S^{-1}, T-K-M-1\right)$$

- where $\widehat{A} = \left(X'X\right)^{-1} X'Y$ is OLS estimate of A, $\widehat{\alpha} = vec\left(\widehat{A}\right)$ and

$$S = \left(Y - X\widehat{A}\right)'\left(Y - X\widehat{A}\right)$$

Digression

- Remember regression models had parameters β and σ^2
- There proved convenient to work with $h = \frac{1}{\sigma^2}$
- In VAR proves convenient to work with Σ^{-1}
- In regression h typically had Gamma distribution
- With VAR Σ^{-1} will typically have Wishart distribution
- Wishart is matrix generalization of Gamma
- Details see appendix to textbook.
- If Σ^{-1} is $W(C, c)$ then "Mean" is cC and c is degrees of freedom.
- Note: easy to take random draws from Wishart.

Prior Issue 1

- VARs are not parsimonious models: α contains KM parameters
- For a VAR(4) involving 5 dependent variables: 105 parameters
- Large VARs with 100+ dependent variable: thousands (or tens of thousands) of parameters
- Macro data sets: number of observations on each variable might be a few hundred.
- Without prior information, hard to obtain precise estimates.
- Features such as impulse responses and forecasts will tend to be imprecisely estimated.
- Desirable to "shrink" forecasts and prior information offers a sensible way of doing this shrinkage.
- Different priors do shrinkage in different ways.

Prior Issue 2

- Some priors lead to analytical results for the posterior and predictive densities.
- Other priors require MCMC methods (which raise computational burden).
- E.g. recursive forecasting exercise typically requires repeated calculation of posterior and predictive distributions
- In this case, MCMC methods can be very computationally demanding.
- May want to go with not-so-good prior which leads to analytical results, if ideal prior leads to slow computation.

Prior Issue 3

- Priors differ in how easily they can handle extensions of the VAR defined above.
- Restricted VARs: different equations have different explanatory variables.
- TVP-VARs: Allowing for VAR coefficients to change over time.
- Heteroskedasticity
- Such extensions typically require MCMC, so no need to restrict consideration to priors which lead to analytical results in basic VAR

The Minnesota Prior

- The classic shrinkage priors developed by researchers (Litterman, Sims, etc.) at the University of Minnesota and the Federal Reserve Bank of Minneapolis.
- They use an approximation which simplifies prior elicitation and computation: replace Σ with an estimate, $\widehat{\Sigma}$.
- Original Minnesota prior simplifies even further by assuming Σ to be a diagonal matrix with $\widehat{\sigma}_{ii} = s_i^2$
- s_i^2 is OLS estimate of the error variance in the i^{th} equation
- If Σ not diagonal, can use, e.g., $\widehat{\Sigma} = \frac{S}{T}$.

- Minnesota prior assumes

$$\alpha \sim N\left(\underline{\alpha}_{Min}, \underline{V}_{Min}\right)$$

- Minnesota prior is way of automatically choosing $\underline{\alpha}_{Min}$ and $\underline{V}_{Min}$
- Note: explanatory variables in any equation can be divided as:
- own lags of the dependent variable
- the lags of the other dependent variables
- exogenous or deterministic variables

- $\underline{\alpha}_{Min} = 0$ implies shrinkage towards zero (a nice way of avoiding over-fitting).
- When working with differenced data (e.g. GDP growth), Minnesota prior sets $\underline{\alpha}_{Min} = 0$
- When working with levels data (e.g. GDP) Minnesota prior sets element of $\underline{\alpha}_{Min}$ for first own lag of the dependent variable to 1.
- Idea: Centred over a random walk. Shrunk towards random walk (specification which often forecasts quite well)
- Other values of $\underline{\alpha}_{Min}$ also used, depending on application.

- Prior mean: "towards what should we shrink?"
- Prior variance: "by how much should we shrink?"
- Minnesota prior: $\underline{V}_{Min}$ is diagonal.
- Let $\underline{V}_i$ denote block of $\underline{V}_{Min}$ for coefficients in equation i
- $\underline{V}_{i,jj}$ are diagonal elements of $\underline{V}_i$
- A common implementation of Minnesota prior (for $r = 1, .., p$ lags):

$$\underline{V}_{i,jj} = \begin{cases} \frac{\underline{a}_1}{r^2} \text{ for coefficients on own lags} \\ \frac{\underline{a}_2 \sigma_{ii}}{r^2 \sigma_{jj}} \text{ for coefficients on lags of variable } j \neq i \\ \underline{a}_3 \sigma_{ii} \text{ for coefficients on exogenous variables} \end{cases}$$

- Typically, $\sigma_{ii} = s_i^2$.

- Problem of choosing $\frac{KM(KM+1)}{2}$ elements of $\underline{V}_{Min}$ reduced to simply choosing , $\underline{a}_1, \underline{a}_2, \underline{a}_3$.
- Property: as lag length increases, coefficients are increasingly shrunk towards zero
- Property: by setting $\underline{a}_1 > \underline{a}_2$ own lags are more likely to be important than lags of other variables.
- $\frac{\sigma_{ii}}{\sigma_{jj}}$ adjusts for differences in the units that the variables are measured in).
- Minnesota prior seems to work well in practice.
- Recent paper by Giannone, Lenza and Primiceri (in ReStat) develops methods for estimating prior hyperparameters from the data

Posterior Inference with Minnesota Prior

- Simple analytical results involving only the Normal distribution.
- $$\alpha | y \sim N\left(\overline{\alpha}_{Min}, \overline{V}_{Min}\right)$$
- Formula for $\overline{\alpha}_{Min}$ and $\overline{V}_{Min}$ can be obtained from standard sources (including my Bayesian Econometric Methods second edition)

Natural conjugate prior

- A drawback of Minnesota prior is its treatment of Σ.
- Ideally want to treat Σ as unknown parameter
- Natural conjugate prior allows us to do this in a way that yields analytical results.
- But (as we shall see) has some drawbacks.

- An examination of likelihood function (see also similar derivations for Normal linear regression model where Normal-Gamma prior was natural conjugate) suggests VAR natural conjugate prior:

$$\alpha|\Sigma \sim N\left(\underline{\alpha}, \Sigma \otimes \underline{V}\right)$$

- $$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{\nu}\right)$$

- $\underline{\alpha}$, $\underline{V}$, $\underline{\nu}$ and $\underline{S}$ are prior hyperparameters chosen by the researcher.
- Noninformative prior: $\underline{\nu} = 0$ and $\underline{S} = \underline{V}^{-1} = cI$ and let $c \to 0$.

Posterior when using natural conjugate prior

- Posterior has analytical form:

$$\alpha|\Sigma, y \sim N\left(\overline{\alpha}, \Sigma \otimes \overline{V}\right)$$

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

- Formulae for $\overline{S}$ and $\overline{\nu}$ available in standard sources

- Remember: in regression model joint posterior for (β, h) was Normal-Gamma, but marginal posterior for β had t-distribution
- Same thing happens with VAR coefficients.
- Marginal posterior for α is a multivariate t-distribution.
- Posterior mean is $\overline{\alpha}$
- Degrees of freedom parameter is $\overline{\nu}$
- Posterior covariance matrix:

$$var(\alpha|y) = \frac{1}{\overline{\nu} - M - 1} \overline{S} \otimes \overline{V}$$

- Posterior inference can be done using (analytical) properties of t-distribution.
- Predictive inference can also be done analytically (for one-step ahead forecasts)

Problems with Natural Conjugate Prior

- Natural conjugate prior has great advantage of analytical results, but has some restrictive properties that can cause problems in some applications.
- Just out in 2022: "Asymmetric conjugate priors for large Bayesian VARs" in Quantitative Economics by Joshua Chan
- New version of a conjugate prior which surmounts some of the problems I am about to list
- To make problems concrete consider a macro example:
- The VAR involves variables such as output growth and the growth in the money supply
- Researcher wants to impose the neutrality of money.
- Implies: coefficients on the lagged money growth variables in the output growth equation are zero (but coefficients of lagged money growth in other equations would not be zero).

- Problem 1: Cannot simply impose neutrality of money restriction.
- The unrestricted VAR means each equation has the same explanatory variables (p lags of all of the dependent variables)
- But can show that, if we relax this assumption, and allow for different equations to have different explanatory variables, analytical results are not available

- Problem 2: Cannot "almost impose" neutrality of money restriction through the prior.
- Cannot set prior mean over neutrality of money restriction and set prior variance to very small value.
- To see why, let individual elements of Σ be σ_{ij}.
- Prior covariance matrix has form $\Sigma \otimes \underline{V}$
- This implies prior covariance of coefficients in equation i is $\sigma_{ii}\underline{V}$.
- Thus prior covariance of the coefficients in any two equations must be proportional to one another.
- So can "almost impose" coefficients on lagged money growth to be zero in ALL equations, but cannot do it in a single equation.
- Note also that Minnesota prior form $\underline{V}_{Min}$ is not consistent with natural conjugate prior.

Other Bayesian VAR Priors

- Many other Bayesian VAR priors proposed (not time to cover here)
- Independent Normal-Wishart prior, steady state VAR, priors based on macro theory (e.g. DSGE prior)
- Lots of machine learning VAR priors (e.g. Bayesian Lasso VAR)
- BEAR Toolbox (available on course website) provides details of some of them

Stochastic Search Variable Selection (SSVS) in VARs

- There are many approaches (often global-local shrinkage priors) which seek parsimony/shrinkage in VARs, take SSVS as an example
- Remember: basic idea for a VAR coefficient, α_j
- SSVS is hierarchical prior, mixture of two Normal distributions:

$$\alpha_j | \gamma_j \sim (1 - \gamma_j) N\left(0, \kappa_{0j}^2\right) + \gamma_j N\left(0, \kappa_{1j}^2\right)$$

- γ_j is 0 or 1
- $\gamma_j = 1$ then α_j has prior $N\left(0, \kappa_{1j}^2\right)$
- $\gamma_j = 0$ then α_j has prior $N\left(0, \kappa_{0j}^2\right)$
- Prior is hierarchical since γ_j is unknown parameter and estimated in a data-based fashion.
- κ_{0j}^2 is "small" (so coefficient is shrunk to be virtually zero)
- κ_{1j}^2 is "large" (implying a relatively noninformative prior for α_j).

Gibbs Sampling with the SSVS Prior

- SSVS prior for VAR coefficients, α, can be written as:

$$\alpha|\gamma \sim N(0, DD)$$

- γ is a vector with elements $\gamma_j \in \{0, 1\}$,
- D is diagonal matrix with $(j, j)^{th}$ element d_j:

$$d_j = \begin{cases} \kappa_{0j} \text{ if } \gamma_j = 0 \\ \kappa_{1j} \text{ if } \gamma_j = 1 \end{cases}$$

- "default semi-automatic approach" to selecting κ_{0j} and κ_{1j}
- Set $\kappa_{0j} = c_0\sqrt{\widehat{var}(\alpha_j)}$ and $\kappa_{1j} = c_1\sqrt{\widehat{var}(\alpha_j)}$
- $\widehat{var}(\alpha_j)$ is estimate from an unrestricted VAR
- E.g. OLS or a preliminary Bayesian estimate from a VAR with noninformative prior
- Constants c_0 and c_1 must have $c_0 \ll c_1$ (e.g. $c_0 = 0.1$ and $c_1 = 10$).

- We need prior for γ and a simple one is:

$$\Pr\left(\gamma_j = 1\right) = \underline{q}_j$$
$$\Pr\left(\gamma_j = 0\right) = 1 - \underline{q}_j$$

- $\underline{q}_j = \frac{1}{2}$ for all j implies each coefficient is *a priori* equally likely to be included as excluded.
- Can use same Wishart prior for Σ^{-1}
- Note: George, Sun and Ni also show how to do SSVS on off-diagonal elements of Σ

- Gibbs sampler sequentially draws from $p\left(\alpha|y,\gamma,\Sigma\right)$, $p\left(\gamma|y,\alpha,\Sigma\right)$ and $p\left(\Sigma^{-1}|y,\gamma,\alpha\right)$
- $$\alpha|y,\gamma,\Sigma \sim N(\overline{\alpha}_{\alpha},\overline{V}_{\alpha})$$
- $$\begin{array}{c} \Pr\left(\gamma_j = 1|y,\alpha,\Sigma\right) = \overline{q}_j \\ \Pr\left(\gamma_j = 0|y,\alpha,\Sigma\right) = 1 - \overline{q}_j \end{array}$$
- $p\left(\Sigma^{-1}|y,\gamma,\alpha\right)$ is Wishart
- I won't write out forumulae for all arguments in posterior (e.g. $\overline{q}_j$), but they have simple forms

Illustration of Bayesian VAR Methods in a Small VAR

- Data set: standard quarterly US data set from 1953Q1 to 2006Q3.
- Inflation rate $\Delta\pi_t$, the unemployment rate u_t and the interest rate r_t
- $y_t = (\Delta\pi_t, u_t, r_t)'$.
- These three variables are commonly used in New Keynesian VARs.
- We use unrestricted VAR with intercept and 4 lags

- We consider 6 priors:
- Noninformative: Noninformative version of natural conjugate prior
- Natural conjugate: Informative natural conjugate prior with subjectively chosen prior hyperparameters
- Minnesota: Minnesota prior
- Independent Normal-Wishart: Independent Normal-Wishart prior with subjectively chosen prior hyperparameters
- SSVS-VAR: SSVS prior for VAR coefficients and Wishart prior for Σ^{-1}
- SSVS: SSVS on both VAR coefficients and error covariance

- Point estimates for VAR coefficients often are not that interesting, but Table 1 presents them for 2 priors
- With SSVS priors, $\Pr(\gamma_j = 1|y)$ is the "posterior inclusion probability" for each coefficient, see Table 2
- Model selection using $\Pr(\gamma_j = 1|y) > \frac{1}{2}$ restricts 25 of 39 coefficients to zero.

Table 1. Posterior mean of VAR Coefficients for Two Priors						
	Noninformative			SSVS - VAR		
	$\Delta\pi_t$	u_t	r_t	$\Delta\pi_t$	u_t	r_t
Intercept	0.2920	0.3222	-0.0138	0.2053	0.3168	0.0143
$\Delta\pi_{t-1}$	1.5087	0.0040	0.5493	1.5041	0.0044	0.3950
u_{t-1}	-0.2664	1.2727	-0.7192	-0.142	1.2564	-0.5648
r_{t-1}	-0.0570	-0.0211	0.7746	-0.0009	-0.0092	0.7859
$\Delta\pi_{t-2}$	-0.4678	0.1005	-0.7745	-0.5051	0.0064	-0.226
u_{t-2}	0.1967	-0.3102	0.7883	0.0739	-0.3251	0.5368
r_{t-2}	0.0626	-0.0229	-0.0288	0.0017	-0.0075	-0.0004
$\Delta\pi_{t-3}$	-0.0774	-0.1879	0.8170	-0.0074	0.0047	0.0017
u_{t-3}	-0.0142	-0.1293	-0.3547	0.0229	-0.0443	-0.0076
r_{t-3}	-0.0073	0.0967	0.0996	-0.0002	0.0562	0.1119
$\Delta\pi_{t-4}$	0.0369	0.1150	-0.4851	-0.0005	0.0028	-0.0575
u_{t-4}	0.0372	0.0669	0.3108	0.0160	0.0140	0.0563
r_{t-4}	-0.0013	-0.0254	0.0591	-0.0011	-0.0030	0.0007

Table 2. Posterior Inclusion Probabilities for VAR Coefficients: SSVS-VAR Prior

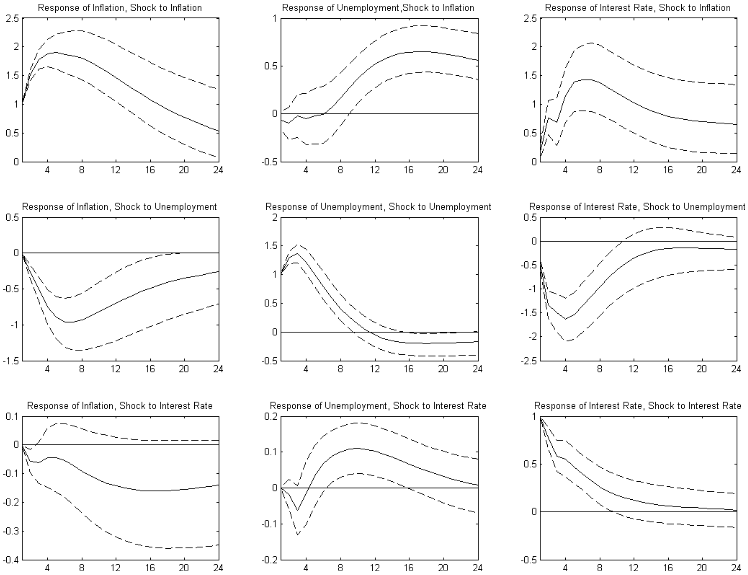
	$\Delta\pi_t$	u_t	r_t
Intercept	0.7262	0.9674	0.1029
$\Delta\pi_{t-1}$	1	0.0651	0.9532
u_{t-1}	0.7928	1	0.8746
r_{t-1}	0.0612	0.2392	1
$\Delta\pi_{t-2}$	0.9936	0.0344	0.5129
u_{t-2}	0.4288	0.9049	0.7808
r_{t-2}	0.0580	0.2061	0.1038
$\Delta\pi_{t-3}$	0.0806	0.0296	0.1284
u_{t-3}	0.2230	0.2159	0.1024
r_{t-3}	0.0416	0.8586	0.6619
$\Delta\pi_{t-4}$	0.0645	0.0507	0.2783
u_{t-4}	0.2125	0.1412	0.2370
r_{t-4}	0.0556	0.1724	0.1097

Impulse Response Analysis

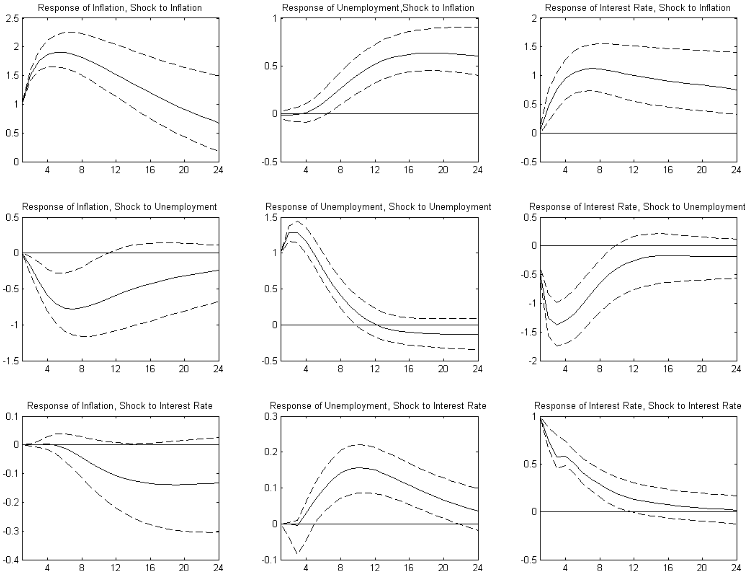
- Impulse response analysis is commonly done with VARs
- Given my focus on the Bayesian econometrics, as opposed to macroeconomics, I will not explain in detail
- Make standard identifying assumption which allows for the interpretation of interest rate shock as monetary policy shock.

- Figures 2 and 3 present impulse responses of all variables to shocks
- Use two priors: the noninformative one and the SSVS prior
- Posterior median is solid line and dotted lines are 10^{th} and 90^{th} percentiles.
- Priors give similar results, but a careful examination reveals SSVS leads to slightly more precise inferences (evidenced by a narrower band between the 10^{th} and 90^{th} percentiles) due to the shrinkage it provides.

Impulse Responses for Noninformative Prior



Impulse Responses for SSVS Prior



Summary

- Lecture began with summary of basic methods and issues which arise with Bayesian VAR modelling and addressed questions such as:
- Why is shrinkage necessary?
- How should shrinkage be done?
- With recent explosion of interest in large VARs, need for answers for such questions is greatly increased
- Many researchers now developing models/methods to address them