Bayesian Methods for Empirical Macroeconomics

Gary Koop

September 14, 2012

Introduction

- There are many popular time series models and all cannot be covered in a short course.
- In this and next lectures, we discuss models popular with empirical macroeconomists, characterized by:
- i) Multivariate in nature (macroeconomists interested in relationships between variables, not properties of a single variable).
- ii) Allow for parameters to change (e.g. over time, across business cycle, etc.).
- We will not cover univariate time series nor nonlinear time series models such as Markov switching, TAR, STAR, etc.
- See Bayesian Econometric Methods Chapters 17 and 18 for treatment of some of these models.
- We will discuss state space models (which can be used to model nonlinearities).

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

- One way of writing 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).

- Many alternative ways of writing the VAR (and we will use some alternatives below).
- One way: let y be $MT \times 1$ vector ($y = (y_1', .., y_T')$) and ε stacked conformably
- $x_t = \left(1, y_{t-1}', .., y_{t-p}'\right)$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}$$

- $K = 1 + Mp$ is number of coefficients in each equation of VAR and X is a $T \times K$ matrix.
- The VAR can be written as:

$$y = (I_M \otimes X)\,\alpha + \varepsilon$$

- $\varepsilon \sim N\left(0, \Sigma \otimes I_M\right)$.

- Second 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

$$Y = XA + E$$

- In first VAR, α is $KM \times 1$ vector of VAR coefficients, here A is $K \times M$
- Relationship between two: $\alpha = vec(A)$
- We will use both notations below (and later on, when working with restricted VAR need to introduce yet more notation).

Likelihood Function

- Likelihood function can be derived and shown to be of a form that breaks into two parts (see Bayesian Econometric Methods Exercise 17.6)
- First of these parts α given Σ and another for Σ
- $$\alpha|\Sigma, y \sim N\left(\widehat{\alpha}, \Sigma \otimes (X'X)^{-1}\right)$$
- Σ^{-1} has Wishart form

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

- where $\widehat{A} = (X'X)^{-1} X'Y$ 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.
- 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 to 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 growth) 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.
- See Litterman (1986) for motivation and discussion of these choices (e.g. explanation for how $\frac{\sigma_{ii}}{\sigma_{jj}}$ adjusts for differences in the units that the variables are measured in).
- Minnesota prior seems to work well in practice.
- E.g. Banbura, Giannone and Reichlin (JAE, 2010) use Minnesota prior in large VARs with over 100 dependent variables and find it forecasts very well (relative to factor methods).
- Recent working paper by Giannone, Lenza and Primiceri develops methods for estimating prior hyperparameters from the data

Posterior Inference with Minnesota Prior

- Simple analytical results involving only the Normal distribution.
- $$\alpha|y \sim N\left(\overline{\alpha}_{Min}, \overline{V}_{Min}\right)$$
- $$\overline{V}_{Min} = \left[\underline{V}_{Min}^{-1} + \left(\widehat{\Sigma}^{-1} \otimes (X'X)\right)\right]^{-1}$$
- $$\overline{\alpha}_{Min} = \overline{V}_{Min}\left[\underline{V}_{Min}^{-1}\underline{\alpha}_{Min} + \left(\widehat{\Sigma}^{-1} \otimes X\right)' y\right]$$

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 sell) has some drawbacks.
- In practice, noninformative limiting version of natural conjugate prior sometimes used (but noninformative prior does not do shrinkage)

- 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 $c \rightarrow 0$.

Posterior when using natural conjugate prior

- Posterior has analytical form:

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

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

- where

$$\overline{V} = \left[\underline{V}^{-1} + X'X\right]^{-1}$$

- $$\overline{A} = \overline{V}\left[\underline{V}^{-1}\underline{A} + X'X\widehat{A}\right]$$

- $$\overline{S} = S + \underline{S} + \widehat{A}'X'X\widehat{A} + \underline{A}'\underline{V}^{-1}\underline{A} - \overline{A}'\left(\underline{V}^{-1} + X'X\right)\overline{A}$$

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

- 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.

Problems with Natural Conjugate Prior

- Natural conjugate prior has great advantage of analytical results, but has some problems which make it rarely used in practice.
- 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 $(I_M \otimes X)$ form of the explanatory variables in VAR means every equation must have same set of explanatory variables.
- But if we do not maintain $(I_M \otimes X)$ form, don't get analytical conjugate prior (see Kadiyala and Karlsson, JAE, 1997 for details).

- 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.

Some interesting approaches I will not discuss

- Choosing prior hyperparameters by using dummy observations (fictitious prior data set), see Sims and Zha (1998, IER).
- Using prior information from macro theory (e.g. DSGE models), see Ingram and Whiteman (1994, JME) and Del Negro and Schorfheide (2004, IER).
- Villani (2009, JAE): priors about means of dependent variables
- Useful since researchers often have prior information on these.
- Write VAR as:

$$\widetilde{A}(L)(y_t - \widetilde{a}_0) = \varepsilon_t$$

- $\widetilde{A}(L) = I - \widetilde{A}_1 L - .. - \widetilde{A}_p L^p$, L is the lag operator
- $\widetilde{a}_0$ are unconditional means of the dependent variables.
- Gibbs sampling required.

Optional Topic 1: A Macroeconomic Example

- Hybrid New Keynesian Phillips Curve (NKPC) Model
- Inflation (π_t) and y_t is output gap or unemployment rate

$$\pi_t = \beta_b \pi_{t-1} + \beta_f E_{t-1}(\pi_{t+1}) + \gamma y_t + \varepsilon_t.$$

- $E_{t-1}(\pi_{t+1})$ is expectation at $t-1$ of inflation at $t+1$
- Note: Adding equation for y_t will give a multivariate model.
- No feedback from π_t to y_t

Relating the NKPC to a VAR

- To take NKPC to data need to find rational expectations solution to get rid of $E_{t-1}(\pi_{t+1})$ term in NKPC
- Since no feedback from π_t to y_t can show solution is:

$$\pi_t = a_1 \pi_{t-1} + a_2 y_{t-1} + u_t$$

- where $a_1 = f_1(\beta_b, \beta_f)$ and $a_2 = f_2(\beta_b, \beta_f, \gamma, \rho)$ for functions f_1 and f_2
- Case 1: suppose y_t is

$$y_t = \rho y_{t-1} + v_t$$

- These 2 equations form a restricted VAR (reduced form model)
- Rational expectations macro models often lead to restricted VARs

- Problem: VAR has 3 parameters, a_1, a_2, and ρ, but structural model has 4 (β_f, β_b, γ, and ρ)
- Identification issues in rational expectations/DSGE models can be important.
- Case 2: suppose y_t is

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + v_t$$

- Identification problem is now solved since reduced form VAR now has 4 parameters a_1, a_2, ρ_1 and ρ_2
- But is this solution a good one? Identification depends on lag length. What if ρ_2 is near zero?
- Summary: macro theory can often lead to restricted VARs, but identification can be a worry

The Independent Normal-Wishart Prior

- Natural conjugate prior had $\alpha|\Sigma$ being Normal and Σ^{-1} being Wishart and VAR had same explanatory variables in every equation.
- Want more general setup without these restrictive features.
- Can do this with a prior for VAR coefficients and Σ^{-1} being independent (hence name "independent Normal-Wishart prior")
- And using a more general formulation for the VAR

- To allow for different equations in the VAR to have different explanatory variables, modify notation.
- To avoid, use "β" notation for VAR coefficients now instead of α.
- Each equation (for $m = 1, .., M$) of the VAR is:

$$y_{mt} = z'_{mt}\beta_m + \varepsilon_{mt},$$

- If we set $z_{mt} = \left(1, y'_{t-1}, .., y'_{t-p}\right)'$ for $m = 1, .., M$ then exactly same VAR as before.
- However, here z_{mt} can contain different lags of dependent variables, exogenous variables or deterministic terms.

- Vector/matrix notation:
- $y_t = (y_{1t}, .., y_{Mt})'$, $\varepsilon_t = (\varepsilon_{1t}, .., \varepsilon_{Mt})'$
-

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix}$$

-

$$Z_t = \begin{pmatrix} z'_{1t} & 0 & \cdots & 0 \\ 0 & z'_{2t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z'_{Mt} \end{pmatrix}$$

- β is $k \times 1$ vector, Z_t is $M \times k$ where $k = \sum_{j=1}^{M} k_j$.
- ε_t is i.i.d. $N(0, \Sigma)$.
- Can write VAR as:

$$y_t = Z_t \beta + \varepsilon_t$$

- Stacking:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}$$

- $$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

- $$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_T \end{pmatrix}$$

- VAR can be written as:

$$y = Z\beta + \varepsilon$$

- ε is $N(0, I \otimes \Sigma)$.

- Thus, VAR can be written as a Normal linear regression model with error covariance matrix of a particular form (SUR form).
- Independent Normal-Wishart prior:

$$p\left(\beta, \Sigma^{-1}\right) = p\left(\beta\right) p\left(\Sigma^{-1}\right)$$

- where

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

- and

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

- $\underline{V}_{\beta}$ can be anything the researcher chooses (not restrictive $\Sigma \otimes \underline{V}$ form of the natural conjugate prior).
- $\underline{\beta}$ and $\underline{V}_{\beta}$ could be set as in the Minnesota prior.
- A noninformative prior obtained by setting $\underline{\nu} = \underline{S} = \underline{V}_{\beta}^{-1} = 0$.

Posterior inference in the VAR with independent Normal-Wishart prior

- $p\left(\beta, \Sigma^{-1}|y\right)$ does not have a convenient form allowing for analytical results.
- But Gibbs sampler can be set up.
- Conditional posterior distributions $p\left(\beta|y, \Sigma^{-1}\right)$ and $p\left(\Sigma^{-1}|y, \beta\right)$ do have convenient forms
- $$\beta|y, \Sigma^{-1} \sim N\left(\overline{\beta}, \overline{V}_{\beta}\right)$$
- where
$$\overline{V}_{\beta} = \left(\underline{V}_{\beta}^{-1} + \sum_{t=1}^{T} Z_t' \Sigma^{-1} Z_t\right)^{-1}$$
- and
$$\overline{\beta} = \overline{V}_{\beta}\left(\underline{V}_{\beta}^{-1}\underline{\beta} + \sum_{i=1}^{T} Z_t' \Sigma^{-1} y_t\right)$$

- $$\Sigma^{-1}|y,\beta \sim W\left(\overline{S}^{-1},\overline{\nu},\right)$$
- where
$$\overline{\nu} = T + \underline{\nu}$$
- $$\overline{S} = \underline{S} + \sum_{t=1}^{T}\left(y_t - Z_t\beta\right)\left(y_t - Z_t\beta\right)'$$
- Remember: for any Gibbs sampler, the resulting draws can be used to calculate posterior properties of any function of the parameters (e.g. impulse responses), marginal likelihoods (for model comparison) and/or to do prediction.

Note on Prediction in VARs

- For the VAR, Z_τ contains information dated $\tau - 1$ or earlier.
- For predicting at time τ given information through $\tau - 1$, can use:

$$y_\tau | Z_\tau, \beta, \Sigma \sim N\left(Z_t \beta, \Sigma\right)$$

- This result and Gibbs draws $\beta^{(s)}, \Sigma^{(s)}$ for $s = 1, .., S$ allows for predictive inference.
- E.g. predictive mean (a popular point forecast) could be obtained as:

$$E\left(y_\tau | Z_\tau\right) = \frac{\sum_{s=1}^{S} Z_t \beta^{(s)}}{S}$$

- Other predictive moments can be calculated in a similar fashion.

Stochastic Search Variable Selection (SSVS) in VARs

- There are many approaches which seek parsimony/shrinkage in VARs, take SSVS as a representative example
- SSVS is usually done in VAR where every equation has same explanatory variables
- Hence, return to our initial notation for VARs where X contains lagged dependent variable, α are VAR coefficients, etc.
- SSVS can be interpreted as a prior shrinks some VAR coefficients to zero
- Or as a model selection device (select the model with explanatory variables with non-zero coefficients)
- Or as a model averaging device (which averages over models with different non-zero coefficients).
- Can be implemented in various ways, here we follow George, Sun and Ni (2008, JoE)

- Basic idea for a VAR coefficient, α_j
- Before we used conventional priors, but SSVS is a hierarchical prior
- Hierarchical prior $=$ prior expressed in terms of parameters which in turn have a prior of their own

- SSVS prior is mixture of two Normal distributions:

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

- γ_j is dummy variable.
- $\gamma_j = 1$ 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).

- Below we describe a Gibbs sampler for this model which provides draws of γ and other parameters
- SSVS can select a single restricted model.
- Run Gibbs sampler and calculate $\Pr\left(\gamma_j = 1|y\right)$ for $j = 1, .., KM$
- Set to zero all coefficients with $\Pr\left(\gamma_j = 1|y\right) < a$ (e.g. $a = 0.5$).
- Then re-run Gibbs sampler using this restricted model
- Alternatively, if the Gibbs sampler for unrestricted VAR is used to produce posterior results for the VAR coefficients, result will be Bayesian model averaging (BMA).

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 $p\left(\Sigma^{-1}|y,\gamma,\alpha\right)$
- $$\alpha|y,\gamma,\Sigma \sim N(\overline{\alpha}_{\alpha},\overline{V}_{\alpha})$$
- where
$$\overline{V}_{\alpha} = [\Sigma^{-1} \otimes (X'X) + (DD)^{-1}]^{-1}$$
- $$\overline{\alpha}_{\alpha} = \overline{V}_{\alpha}[(\Psi\Psi') \otimes (X'X)\hat{\alpha}]$$
- $$\hat{A} = (X'X)^{-1}X'Y$$
- $$\hat{\alpha} = vec(\hat{A})$$

- $p\left(\gamma|y,\alpha,\Sigma\right)$ has γ_j being independent Bernoulli random variables:
- $$\Pr\left(\gamma_j = 1|y,\alpha,\Sigma\right) = \overline{q}_j$$
  $$\Pr\left(\gamma_j = 0|y,\alpha,\Sigma\right) = 1-\overline{q}_j$$
- where

$$\overline{q}_j = \frac{\frac{1}{\kappa_{1j}}\exp\left(-\frac{\alpha_j^2}{2\kappa_{1j}^2}\right)\underline{q}_j}{\frac{1}{\kappa_{1j}}\exp\left(-\frac{\alpha_j^2}{2\kappa_{1j}^2}\right)\underline{q}_j + \frac{1}{\kappa_{0j}}\exp\left(-\frac{\alpha_j^2}{2\kappa_{0j}^2}\right)\left(1-\underline{q}_j\right)}$$

- $p\left(\Sigma^{-1}|y,\gamma,\alpha\right)$ has similar Wishart form as previously, so I will not repeat here.

Illustration of Bayesian VAR Methods

- 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.
- The data are plotted in Figure 1.
- We use unrestricted VAR with intercept and 4 lags

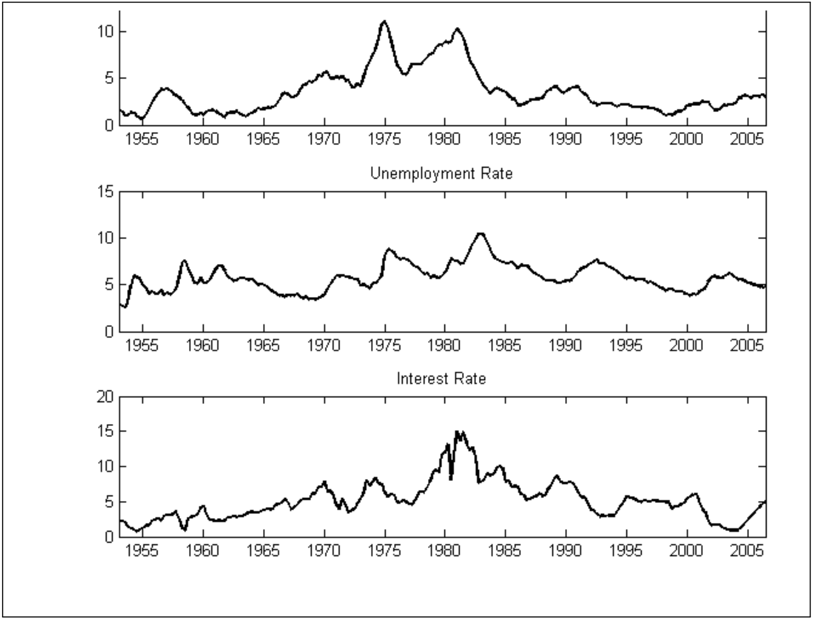

Figure 1: The Data

- 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, $\Pr\left(\gamma_j = 1|y\right)$ is the "posterior inclusion probability" for each coefficient, see Table 2
- Model selection using $\Pr\left(\gamma_j = 1|y\right) > \frac{1}{2}$ restricts 25 of 39 coefficients to zero.
- Table 3, prediction: $p\left(y_{T+1}|y_1.., y_T\right)$ where $T = 2006Q3$.

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 Probabilites 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
- The VAR so far is a reduced form model:

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

- where $var(\varepsilon_t) = \Sigma$
- Macroeconomists often work with structural VARs:

$$C_0 y_t = c_0 + \sum_{j=1}^{p} C_j y_{t-j} + u_t$$

- where $var(u_t) = I$
- u_t are shocks which have an economic interpretation (e.g. monetary policy shock)

- Macroeconomist interested in effect of (e.g.) monetary policy shock now on all dependent variables in future = impulse response analysis
- Need to restrict C_0 to identify model.
- We assume C_0 lower triangular
- This is a standard identifying assumption used, among many others, by Bernanke and Mihov (1998), Christiano, Eichanbaum and Evans (1999) and Primiceri (2005).
- Allows for the interpretation of interest rate shock as monetary policy shock.
- Aside: sign-restricted impulse responses of Uhlig (2005) are increasingly popular

- 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.

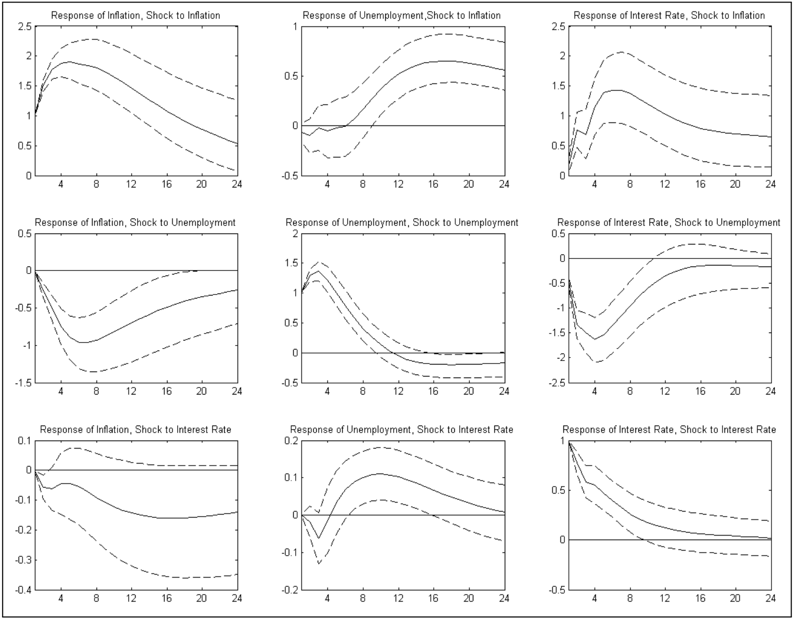

Figure 2: Impulse Responses for Noninformative Prior

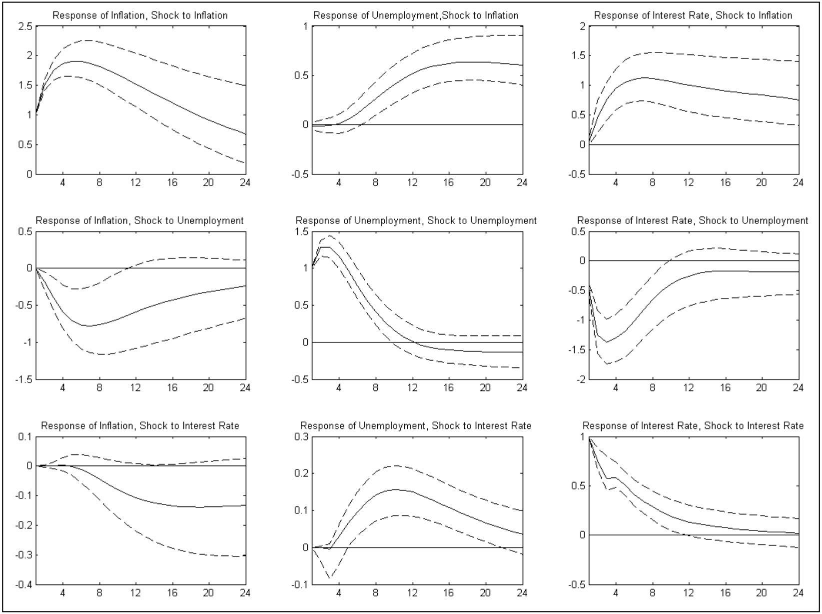

Figure 3: Impulse Responses for SSVS