Bayesian Vector Autoregressive Models

Introduction

- In this course, will focus on 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.)
- iii) Large number of variables (Big Data)
- 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.

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
- First half of course will build towards TVP-VARs with multivariate stochastic volatility is our end goal
- Begin with Bayesian VARs (with constant coefficients and homoskedasticity)
- 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).

- Several 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:

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

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

- 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

$$Y = XA + E$$

- 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)
- Bayesians combine likelihood function with prior to produce posterior
- Let us begin with likelihood

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
- $$\alpha|\Sigma, y \sim N\left(\widehat{\alpha}, \Sigma \otimes \left(X'X\right)^{-1}\right)$$
- 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

- Regression models have parameters β and σ^2
- Bayesian analysis often works with $h = \frac{1}{\sigma^2}$
- Prior/posterior for h has Gamma distribution
- Or you can work with σ^2 and have inverse Gamma
- Same issues arise with Σ
- With VAR Σ^{-1} will have Wishart distribution
- With Σ have inverse Wishart
- Wishart is matrix generalization of Gamma
- 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.
- With large VARs have thousands (or more) 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 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.
- Giannone, Lenza and Primiceri (ReStat, 2015) develops methods for estimating prior hyperparameters from the data

Posterior Inference with Minnesota Prior

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

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 $c \longrightarrow 0$.

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

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

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

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

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

- 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:

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

- 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 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.

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.

Prediction in VARs

- I will use prediction and forecasting to mean the same thing
- Goal predict y_τ for some period τ using data available at time $\tau - 1$
- 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

Prediction in VARs

- Or can do predictive simulation:
- For each Gibbs draw $\beta^{(s)}, \Sigma^{(s)}$ simulate one (or more) $y_\tau^{(s)}$
- Result will be $y_\tau^{(s)}$ for $s = 1, .., S$ draws
- Plot them to produce predictive density
- Average them to produce predictive mean
- Take their standard deviation to produce predictive standard deviation
- etc.

Prediction in VARs

- Preceding material was about predicting y_τ using data available at time $\tau - 1$
- This is one-period ahead forecasting
- But what about h-period ahead forecast
- h is the forecast horizon
- E.g. with quarterly data forecasting a year ahead $h = 4$
- Can do direct or iterated forecasting

Direct Forecasting in VARs

- Direct forecasting is straightforward: simply redefine Z_τ
- Above defined each equation using $z_{m\tau} = \left(1, y'_{\tau-1}, .., y'_{\tau-p}\right)'$
- Replace this by $z_{m\tau} = \left(1, y'_{\tau-h}, .., y'_{\tau-p-h+1}\right)'$
- Then your model is always predicting y_τ using data available at time $\tau - h$
- All posterior and predictive formulae are as above
- If forecasting (e.g.) for $h = 1, 2, 3, 4$ must re-estimate model for each h

Iterated Forecasting in VARs

- Estimate the model once using $z_{m\tau} = \left(1, y'_{\tau-1}, .., y'_{\tau-p}\right)'$
- Remember result that

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

- When forecasting y_τ using information available at time $\tau - h$ for $h > 1$ you face a problem using (**)
- Use $h = 2$ and $p = 2$ to illustrate
- In the model, y_τ depends on $y_{\tau-1}$ and $y_{\tau-2}$
- But as a forecaster, you do not know $y_{\tau-1}$ yet
- E.g. suppose you have data through 2015Q4
- When forecasting 2016Q1 ($h = 1$) will have data for 2015Q4 and 2015Q3
- So Z_t is known for $h = 1$
- But when forecasting 2016Q2 ($h = 2$) will not have data for 2016Q1 and not know Z_t

Iterated Forecasting in VARs

- Solution to problem:
- Do predictive simulation beginning with $h = 1$
- Use draw of $y_{\tau-1}^{(s)}$ (along with y_{t-2}, $\beta^{(s)}, \Sigma^{(s)}$) to plug into (**)
- This is called iteration
- For $h > 2$ just keep on iterating
- Strategy above will provide you with draws $y_{\tau-1}^{(s)}$ and $y_{\tau-2}^{(s)}$
- For $h = 3$ can use these to define appropriate Z_t for use in (**)
- etc.
- Which of iterated or direct forecasting is better?
- This seems to depend on the data set being used

Large VARs: A Promising Way of Dealing with Big Data

- Pioneering paper: Banbura, Giannone and Reichlin (2010, JAE) "Large Bayesian Vector Autoregressions"
- Banbura et al paper has 131 dependent variables (standard US macro variables)
- Many others, here is a sample (note range of types of applications in macro/finance and internationally):
- Carriero, Kapetanios and Marcellino (2009, IJF): exchange rates for many countries
- Carriero, Kapetanios and Marcellino (2012, JBF): US government bond yields of different maturities
- Giannone, Lenza, Momferatou and Onorante (2010): euro area inflation forecasting (components of inflation)
- Koop and Korobilis (2016, EER) eurozone sovereign debt crisis
- Bloor and Matheson (2010, EE): macro application for New Zealand
- Jarociński and Maćkowiak (2016, ReStat): Granger causality
- Banbura, Glannone and Lenza (2014, ECB): conditional forecasts/scenario analysis in euro area

Why large VARs?

- Availability of more data
- More data means more information, makes sense to include it
- Concerns about missing out important information (omitted variables bias, fundamentalness, etc.)
- The main alternatives are factor models
- Factors squeeze information in large number of variables to small number of factors
- But this squeezing is done without reference to explanatory power (i.e. squeeze first then put in regression model or VAR): "unsupervised"
- Large VAR methods are supervised and can easily see role of individual variables
- And they work: often beating factor methods in forecasting competitions

- BGR "medium" VAR has 20 dep vars and "large" VAR has 130
- Usually, when working with so many macroeconomic variables, factor methods are used
- We will discuss factor methods in later lecture
- However, BGR find that medium and large Bayesian VARs can forecast better than factor methods
- Perhaps Bayesian VARs should be used even when researcher has dozens or hundreds of variables?
- Dimensionality of α is key
- Large VAR with quarterly data might have $n = 100$ and $p = 4$ so α contains over 40000 coefficients.
- With monthly data it would have over 100000 coefficients.
- For a medium VAR, α might have about 1500 coefficients with quarterly data.
- Σ is parameter rich: $\frac{n(n+1)}{2}$ elements.

- Number of parameters may far exceed the number of observations.
- In theory, this is no problem for Bayesian methods.
- These combine likelihood function with prior.
- Even if parameters in likelihood function are not identified, combining with prior will (under weak conditions) lead to valid posterior density
- But how well do they work in practice?
- Role of prior information becomes more important as likelihood is less informative
- Methods I have discussed have been found to work well
- But very active area of research (both for econometric theory and empirical practice)

Conclusion

- 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
- Bayesian Estimation, Analysis and Regression (BEAR) Toolbox is easy to use computer program for Bayesian VARs
- Developed by researchers at the European Central Bank
- BEAR toolbox working paper good source for proofs and derivations relating to VARs
- Another good resource: Sune Karlsson: Forecasting with Bayesian Vector Autoregressions (Handbook of Economic Forecasting, volume 2, Elsevier)