Contents

Chapter 7

Multivariate or Vector Time Series Models

- Economic models typically focus on the interrelation of endogenous and exogenous variables and vector autoregressive models have become the linchpin in understanding these interrelationships
- A big obstacle for vector processes is notational baggage.
- One might reasonably expect that the estimation problems lie primarily with estimating vector MA processes (recall the nonlinearities even in the univariate setting) and that vector autoregressive processes (called $VARs$) do not pose as great a problem computationally
- While most modern time series books discuss the topics of vector processes they tend to be spread out all over the place, particularly with the arrival of the study of nonstationary (cointegration_ processes
- A good readable source is Hamilton.

7.1 Vector Autoregressive Moving Average Processes

- For this chapter we maintain the assumptions of stationarity and then in Chapter 11 move to nonstationary processes
- Such a dichotomy creates some difficulty in reading the current literature because there the presentation is more general encompassing both stationary and instigations vector processes.
- We may use the Wold representation theorem (as we did for the univariate case) to represent any vector stationary nondeterministic (indeterministic) process $\boldsymbol{y}_t$ $(N \times 1)$ as a **vector MA** process of infinite order:

$$\boldsymbol{y}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\epsilon}_{t-j} \tag{7.1}$$

where $\boldsymbol{\Psi}_j$ is an $(N \times N)$ matrix of parameters with $\boldsymbol{\Psi}_0 = \boldsymbol{I}$ and $\boldsymbol{\epsilon}_t$ a $(N \times 1)$ vector of white noise with mean zero (We have assumed any deterministic components like a constant or trend have been removed).

- Vector white noise may be thought of as:

  (i) $E[\,\boldsymbol{\epsilon}_t\,] = \mathbf{0}$ for all t

  (ii) $E\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s^T\right] = \begin{cases} \Omega & \text{for } t = s \\ \mathbf{0} & \text{for } t \neq s \end{cases}$

  where Ω is an $(N \times N)$ nonsingular covariance matrix.

- This general representation is applicable for all stationary vector processes.

- Note we have the same analogous conditions for **weak stationarity** with time invariant finite mean $E[\,\boldsymbol{y}_t\,] = \mathbf{0}$ (for our example) and autocovariances: $E[\,\boldsymbol{y}_t \boldsymbol{y}_{t+h}^T\,] = \boldsymbol{\Gamma}(h)$.

- Again we may specify a parsimonious representation of (7.1) by the **vector autoregressive moving average process:**

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \ldots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\epsilon}_t + \boldsymbol{\Theta}_1 \boldsymbol{\epsilon}_{t-1} + \ldots + \boldsymbol{\Theta}_q \boldsymbol{\epsilon}_{t-q} \tag{7.2}$$

  where the $\boldsymbol{\Phi}_i' s$ are $(N \times N)$ matrices of the AR parameters and $\boldsymbol{\Theta}_i' s$ are the MA parameters.

- We refer to this as a **vector** ARMA(p, q).

- Notice that the order of the AR and MA processes are the orders of the highest process for any of the individual equations.

- There can be zeros for particular equations.

- In practice, we seldom put restrictions on specific equations of the system but rather treat each equation symmetrically.

- However, we have to put restrictions on the $\boldsymbol{\Psi}_i' s$ in (7.1) to obtain the representation (7.2) –just as we did in the univariate case. (see Judge et al pp. 657 for a simple recursion formula).

7.2 Autocovariance and Autocorrelation Function

- We may obtain the autocovariance function for the $\boldsymbol{y}_t$ simply by evaluating the infinite moving average representation (7.1):

$$E\left[\boldsymbol{y}_t \boldsymbol{y}_{t+h}^T\right] = \boldsymbol{\Gamma}(h) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\Omega} \boldsymbol{\Psi}_{j+h}^T \tag{7.3}$$

- We may specialize (7.2) and consider the autocovariance functions for particular processes. For instance, the vector $MA(1)$ process:

$$\boldsymbol{y}_t \;=\; \boldsymbol{\epsilon}_t + \boldsymbol{\Theta}\boldsymbol{\epsilon}_{t-1} \tag{7.4}$$

is stationary regardless of the $\boldsymbol{\Theta}$ (recall stationarity is concerned with the AR part of the process). It follows from (7.4) that:

$$\boldsymbol{\Gamma}(0) \;=\; E\left[\,\boldsymbol{y}_t\boldsymbol{y}_t^T\,\right] \;=\; \boldsymbol{\Omega} + \boldsymbol{\Theta}\boldsymbol{\Omega}\boldsymbol{\Theta}^T \text{ for all } \; t \tag{7.5}$$

and

$$\boldsymbol{\Gamma}(h) \;=\; E\left[\boldsymbol{y}_t\boldsymbol{y}_{t-h}^T\right] = \begin{cases} \boldsymbol{\Theta}\boldsymbol{\Omega} & \text{if} \quad h = 1 \\ \boldsymbol{\Omega}\boldsymbol{\Theta}^T & \text{if} \quad h = -1 \\ \mathbf{0} & \text{if} \quad \mid h \mid > 1 \end{cases} \tag{7.6}$$

and that $\boldsymbol{\Gamma}(1) \;=\; \boldsymbol{\Gamma}^T(-1)$.

- Just as the univariate $MA(1)$ process has a memory of one, so does the vector MA.

- The intuition also extends to the vector $AR(1)$ process.

- Assume the conditions for stationarity are satisfied (dealt with in a moment), then the stationary vector $AR(1)$ model (called VAR(1)) is:

$$\boldsymbol{y}_t \;=\; \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t \tag{7.7}$$

- We follow the same trick as before to obtain the autocovariance function, namely multiplying by $\boldsymbol{y}_{t-h}$ and taking expectations of (7.7):

$$E\left[\,\boldsymbol{y}_t\boldsymbol{y}_{t-h}^T\,\right] \;=\; \boldsymbol{\Phi}E\left[\,\boldsymbol{y}_{t-1}\boldsymbol{y}_{t-h}^T\,\right] + E\left[\,\boldsymbol{\epsilon}_t\boldsymbol{y}_{t-h}^T\,\right] \tag{7.8}$$

The last term is a vector of zeros for $h \geq 1$ and so we get (compare to the univariate case):

$$\boldsymbol{\Gamma}(h) \;=\; \boldsymbol{\Phi}\boldsymbol{\Gamma}(h-1) \text{ for } h \geq 1 \tag{7.9}$$

This is a vector difference equation with the solution:

$$\boldsymbol{\Gamma}(h) \;=\; \boldsymbol{\Phi}^h\boldsymbol{\Gamma}(0) \text{ for } h \geq 0 \tag{7.10}$$

- To evaluate $\boldsymbol{\Gamma}(0)$ we need the last term in (7.8) for $h = 0$:

$$E\left[\,\boldsymbol{\epsilon}_t(\boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t)^T\,\right] \;=\; E\left[\,\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\,\right] \;=\; \boldsymbol{\Omega} \tag{7.11}$$

- The first term in (7.8) for $h = 0$ is:

$$\boldsymbol{\Phi}\boldsymbol{\Gamma}(-1) \;=\; \boldsymbol{\Phi}\boldsymbol{\Gamma}^T(1) \;=\; \boldsymbol{\Phi}\boldsymbol{\Gamma}(0)\boldsymbol{\Phi}^T \tag{7.12}$$

using (7.10).

- Finally combining (7.11) and (7.12) gives

$$\boldsymbol{\Gamma}(0) \;=\; \boldsymbol{\Phi}\boldsymbol{\Gamma}(0)\boldsymbol{\Phi}^T + \boldsymbol{\Omega} \tag{7.13}$$

- Solving this can be done by either iterating on (7.13) from an initial guess (say from some estimate) or solving directly using:

$$vec\,[\,\boldsymbol{\Gamma}(0)\,] \;=\; \{\boldsymbol{I} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi}\}^{-1} vec\,[\,\boldsymbol{\Omega}\,]\,.$$

  where $\otimes$ is the Kronecker delta product and vec [ ] is the operator that stacks an $n \times m$ one on top of the other to form a column vector of length nm.

- We may obtain the autocovariance for the any vector ARMA process following the same basic rules applied for the univariate case.

- The notation becomes somewhat dense buried in vec and $\otimes$. See Fuller (pp. 70 -74) for a more comprehensive discussion.

- Typically the autocovariance function has little practical applications since identification is not straightforward from this object

7.3 Stationarity and Invertibility for Vector ARMA Models

- As before, the stationarity of the process depends upon the autoregressive part of the model.

- First consider the simple VAR(1) case (7.7). Repeated substitution for $\boldsymbol{y}_t$ up to the say, J lags, yields:

$$\boldsymbol{y}_t \;= \sum_{j=0}^{J} \boldsymbol{\Phi}^j \boldsymbol{\epsilon}_{t-j} + \boldsymbol{\Phi}^j \boldsymbol{y}_{t-J} \tag{7.14}$$

- For stationarity, we want the dependency to die out as the lag length increases which implies that the elements in $\boldsymbol{\Phi}^J$ tend to zero as J increases.

- We may diagonalize $\boldsymbol{\Phi}$ by finding a nonsingular matrix $\boldsymbol{Q}$ $(N \times N)$ such that

$$\boldsymbol{\Phi} \;=\; \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}$$

  where the $\boldsymbol{\Lambda} = \text{diag}\,\{\lambda_1, \lambda_2, \ldots, \lambda_N)$.

- We obtain the $\lambda_1, \lambda_2, \ldots, \lambda_N$ by solving the **determinantal vector equation**:

$$|\,\boldsymbol{\Phi} - \lambda \boldsymbol{I}\,| \;=\; 0$$

- Let $N = 2$ (a simple bivariate process) and the ϕ_{ij} refer to the ij^{th} element of $\mathbf{\Phi}$, and the determinantal vector equation is:

$$\begin{vmatrix} \phi_{11} - \lambda & \phi_{12} \\ \phi_{21} & \phi_{22} - \lambda \end{vmatrix} = 0$$

and the roots are obtained by solving the quadratic:

$$\lambda^2 - (\phi_{11} + \phi_{22})\lambda + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21}) \;=\; 0$$

which may be real or complex pairs.

- Now, we are concerned with the behaviour of $\boldsymbol{\phi}^J$ as J gets arbitrarily large:

$$\boldsymbol{\phi}^J \;=\; \boldsymbol{Q\Lambda Q}^{-1} \bullet \boldsymbol{Q\Lambda Q}^{-1} \bullet \ldots. \;=\; \boldsymbol{Q\Lambda}^J\boldsymbol{Q}^{-1}$$

where $\mathbf{\Lambda}^J = \text{diag}\{\lambda_1^J, \ldots, \lambda_p^J\}$ and therefore it follows:

$$\lim_{j\to\infty} \boldsymbol{\phi}^J \;=\; \lim_{j\to\infty} \boldsymbol{Q\Lambda}^J\boldsymbol{Q}^{-1} \;=\; \mathbf{0}$$

if the roots (eigenvalues) are **less than one in absolute values**.

- In the general vector ARMA model, the stationarity condition is

$$\left|\lambda^p \boldsymbol{I} - \lambda^{p-1}\boldsymbol{\phi}_1 - \ldots - \mathbf{\Phi}_p\right| \;=\; 0 \quad \text{for all} \quad |\lambda| \leq 1 \tag{7.15}$$

- There will be Np roots for (7.15) which all must be inside the unit circle.

- In applied work, we seldom see someone actually checking the stationarity or stability conditions (7.15).

- STATA command: *varstable, graph* calculates the roots of (7.15) and can graph them around the unit circle

- In a system context there are tests for unit roots/cointegration that we take up in Chapter 11.

- Instead, the standard procedure is to use the estimated model to forecast arbitrarily far out into the future and determine whether the forecasts are exploding or converging to the unconditional means.

- Of course the individual series may be nonstationary and forecasts themselves may not diverge since some of the vectors might be cointegrated (see Chapter 11)

7.3.1 Invertibility

- The conditions for invertibility follow the same kind of argument:

$$\left|\lambda^p \boldsymbol{I} - \lambda^{p-1}\boldsymbol{\Theta}_1 - \ldots - \boldsymbol{\Theta}_p\right| \;=\; 0 \quad \text{for all} \quad |\lambda| \leq 1 \tag{7.16}$$

- A short hand used in Harvey for denoting vector ARMA is:

$$\boldsymbol{\Phi}_p(L)\boldsymbol{y}_t \;=\; \boldsymbol{\Theta}_q(L)\boldsymbol{\epsilon}_t \tag{7.17}$$

and

$$\begin{aligned}\boldsymbol{\Phi}_p(L) &= \boldsymbol{I} - \boldsymbol{\Phi}_1 L - \ldots - \boldsymbol{\Phi}_p L^p \\ \boldsymbol{\Theta}_q(L) &= \boldsymbol{I} - \boldsymbol{\Theta}_1 L - \ldots - \boldsymbol{\Theta}_p L^q\end{aligned}$$

- Stationarity and invertibility in terms of the polynomial in the lag operator are from the determinantal polynomials $|\boldsymbol{\Phi}_p(L)|$ and $|\boldsymbol{\Theta}_q(L)|$, with the roots lying **outside** the unit circle for both sets of conditions.

7.3.2 Multivariate and Univariate Representations

- Just as in simultaneous equation models one can look at the 'reduced' form of a single equation.
- Consider the univariate representations from a vector stationary ARMA processes.
- First take the infinite MA representation of (7.17):

$$\boldsymbol{y}_t \;=\; \boldsymbol{\Phi}^{-1}(L)\boldsymbol{\Theta}(L)\boldsymbol{\epsilon}_t \tag{7.18}$$

and then expand the $\boldsymbol{\Phi}^{-1}(L)$ by a power series:

$$\boldsymbol{\Phi}^{-1}(L) \;=\; \frac{\boldsymbol{\Phi}^*(L)}{|\boldsymbol{\Phi}(L)|}$$

where $\boldsymbol{\Phi}^*(L)$ is the adjoint of the determinant of $\boldsymbol{\Phi}(L)$.

- Rearranging gives:

$$|\boldsymbol{\Phi}(L)|\,\boldsymbol{y}_t \;=\; \boldsymbol{\Phi}^*(L)\boldsymbol{\Theta}(L)\boldsymbol{\epsilon}_t \tag{7.19}$$

which is the **autoregressive final form**.

- This is a N equation system with an MA process of order $p+q$. Each of the individual MA processes may be expressed as another $MA(p+q)$ process ($i \;=\; 1,\ldots,N$):

$$\phi_i^*(L)y_{it} \;=\; \theta_i^*(L)\eta_{it} \tag{7.20}$$

where η_{it} is white noise, $\theta_i^*(L)$ is the MA component of order $p+q$ and $\phi_i^*(L) \;=\; |\boldsymbol{\Phi}(L)|$ is the AR part of order 2p.

- Thus even if there is a system of equations, we may nevertheless model the process as a univariate one. This also motivates Box-Jenkins focus on univariate processes.

- Keep in mind, that for the univariate representation it may be difficult to identify 'structure' given the convolution of parameters.

7.4 Estimation of Vector ARMA Processes

- There are very few packages available to estimate vector ARMA processes. (STATA can now do it using the state space command *sspace*). No doubt as the technology develops such an estimation problem will be handled routinely.

- Most econometric software can analyze the VAR processes (of particular interest is STATA). The reason that VAR's are so easy to estimate is that this is basically an OLS problem.

- Recall, from seemingly unrelated regressors problem, that if the $\boldsymbol{X}$ matrix (in this case the lagged values of all the $\boldsymbol{y}_t's$) are all the same then OLS estimation of each equation **individually** yields **efficient** estimates.

- Interestingly and thankfully the exact likelihood is not used and the estimation is conditional on dropping observations (order of VAR)

- Hence, if we do not place exclusion restrictions on individual equations, then we can estimate the equations by OLS efficiently.

- When constructing the appropriate variance-covariance matrix we must take into account the correlations of the errors across equations (*i.e.* the $\boldsymbol{\Omega}$).

- The programs that do MLE of the vector ARMA models use the conditional sum of squared errors together with some Gauss-Newton approach discussed for the univariate time series models. We will not go into a deep discussion of the problems associated with this kind of estimation.

- Instead, we shall concentrate on VAR(1). This is just a special case of the general VAR(p) and the same basic results hold.

- Estimate the model equation by equation by OLS and this will be the same as MLE.

- For the $VAR(1)$

$$\boldsymbol{y}_t \;=\; \boldsymbol{\Phi}\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t \tag{7.21}$$

  with N equations there are N^2 parameters together with $N(N+1)/2$ terms coming from the variance covariance matrix $E\left[\,\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t^T\,\right] \;=\; \boldsymbol{\Omega}$.

- Even a 1st order VAR, there are a sizeable number of parameters to estimate.

- The log likelihood function (given the first observation for each variable) under normal support is:

$$logL(\mathbf{\Phi}, \mathbf{\Omega}; \mathbf{y}|\mathbf{y}_1) = -\frac{N(T-1)}{2} log 2\pi - \frac{1}{2}(T-1) log\,|\mathbf{\Omega}| - \frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{y}_t - \mathbf{\Phi}\boldsymbol{y}_{t-1})^T \mathbf{\Omega}^{-1}(\boldsymbol{y}_t - \mathbf{\Phi}\boldsymbol{y}_{t-1}) \tag{7.22}$$

- We could incorporate the distribution of $\boldsymbol{y}_1$ and do full maximum likelihood.
- However, this probably only implies a modest efficiency gain at a cost of considerable nonlinear complexity. Instead, it is much simpler to work with the conditional likelihood.
- Notice that in (7.22), the MLE of $\mathbf{\Phi}$ are obtained by minimizing :

$$S = \frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{y}_t - \mathbf{\Phi}\boldsymbol{y}_{t-1})^T \mathbf{\Omega}^{-1}(\boldsymbol{y}_t - \mathbf{\Phi}\boldsymbol{y}_{t-1}) \tag{7.23}$$

- As discussed earlier, while this may appear to be dependent on $\mathbf{\Omega}$ (hence be a GLS problem), since the regressors are the same, minimizing the unweighted sum of squares (*i.e.* OLS) yields efficient estimates.
- The MLE of $\mathbf{\Omega}$ may be determined by using the OLS residuals and forming the matrix:

$$\tilde{\mathbf{\Omega}} = T^{-1}\sum_{t=2}^{T}\tilde{\boldsymbol{\epsilon}}_t\tilde{\boldsymbol{\epsilon}}_t^T \tag{7.24}$$

- The estimators $\tilde{\mathbf{\Phi}}$ are asymptotically normal with mean $\mathbf{\Phi}$ and variance-covariance:

$$\mathbf{Avar}(\tilde{\mathbf{\Phi}}) = T^{-1}(\mathbf{\Omega} \otimes \mathbf{\Gamma}(0)^{-1}) \tag{7.25}$$

  where $\mathbf{\Gamma}(0) = E[\,\boldsymbol{y}_t\boldsymbol{y}_t^T\,]$.
- The most direct way to estimate $\mathbf{\Gamma}(0)$ is take the OLS estimate once again:

$$\hat{\mathbf{\Gamma}}(0) = T^{-1}\sum \boldsymbol{y}_t\boldsymbol{y}_t^T \tag{7.26}$$

  This is the usual formula in the VAR routines.

7.5 Hypothesis Tests on VARs

- It is straightforward to test hypothesis using the Wald (W), Likelihood ratio (LR) or Lagrange multiplier (LM) principles.
- For example, we could **jointly** test whether the first lags for all variables in the system belong

$$
\begin{aligned}
H_0 &: \mathbf{\Phi} = \mathbf{0} \\
H_1 &: \mathbf{\Phi} \neq \mathbf{0} \\
W &= \tilde{\Phi}^T \left[\tilde{Avar}(\tilde{\Phi}) \right]^{-1} \tilde{\Phi} \sim \chi^2_{N^2} \text{ under } H_0
\end{aligned}
$$

$$
LR = T \log \left\{ \frac{\left| \hat{\mathbf{\Gamma}}(0) \right|}{\left| \tilde{\mathbf{\Omega}} \right|} \right\} \sim \chi^2_{N^2} \quad \text{under } H_0
$$

- Since tests in the VAR context are usually exclusion restrictions (omit lists of variables) LR tests are almost exclusively applied

- The LM test is more complex because of the interdependence (see Gregory and Veall, 1985 for a simple vector LM test of a rational expectation model).

- The vector $MA(1)$ process is considered in Harvey pp.142 and the vector ARMA process in the advanced Judge pp. 679. There are also certain exclusion restrictions that may make estimation simpler.

- For example, restricting the own variable to appear in the system and no MA terms leads to a relatively simpler model. Of course, without any prior information we can not be sure that this is the right thing to do.

- We shall leave the entire discussion of **identification** of vector ARMA models aside. This is a difficult subject which is not of special direct interest. Instead, we shall move into model building with systems.

7.6 Systems Model Building

- The estimation procedures for vector models (as well as the univariate ones earlier) assumed that the vector ARMA(p, q) is known.

- While we could rely on parameter significance testing and test from the most general model down such a procedure will undoubtedly favor large models (together with the residual analysis for white noise errors equation by equation).

- Sometimes the order of the process is decided by the nature of the data; say quarterly data indicates $p = 4$ for a VAR model. The MA term is more problematic in this case.

- Often in macro economics, the theory itself may suggest the appropriate MA structure (with the AR part coming from the underlying dynamics specified in the theory).

- At this particular level of generality, we assume that there are no guidelines suggesting the appropriate order and that the researcher prefers to follow some of the model selection criteria set out earlier for the univariate case.

- For VAR models (no MA components) researchers have employed several criteria (see Judge et al pp. 686-689).

- In STATA a simple command *varsoc* provides the log likelihood, some likelihood ratio tests and various information criteria (they include the constant which is usually ignored as in below)

- These include:

1. **Akaike' criterion AIC:** $AIC = -2ln(\text{maximum likelihood}) + 2(\text{number of parameters})$ which for the VAR model with normal support is:

$$AIC(p) \;=\; ln \mid \tilde{\mathbf{\Omega}}_p \mid + 2N^2 p/T \tag{7.27}$$

   In which case we choose p to minimize AIC(p).

2. **Schwarz criterion** (SBIC) to select the order of the VAR(p)

$$SC(n) = ln \mid \tilde{\mathbf{\Omega}}_p \mid + \frac{N^2 p \; lnT}{T} \tag{7.28}$$

3. **Final Prediction Error** (FPE) for the p in VAR(p) models

$$FPE(p) = \left(\frac{(T+p)}{(T-p)}\right)^N \mid \tilde{\mathbf{\Omega}}_p \mid \tag{7.29}$$

4. **Hannan Quinn Information Criterion** (HQIC):

$$HQIC = \ln \mid \tilde{\mathbf{\Omega}}_p \mid + \frac{2\ln(lnT)N^2 p}{T}$$

- There are several others reviewed in Judge et al.

- The FPE may be used equation by equation to get a vector AR process, where the order of the process is equation specific.

- In general, researchers have not tried to obtain the most parsimonious specification within each equation, but rather get one for the entire system. Of course one could always employ the techniques for the univariate system and the re-estimate then system given the results for the individual equation pretesting.

- In univariate applications, both AIC and FPE have a tendency to over-parameterize and consequently the SC is favoured. Interestingly, for $VAR's$ Gonzalo and Pitarakis (1999) have shown that the probability of overfitting is a decreasing function of the dimension of the system. Large systems will have less chance of overfitting. Their Monte Carlo work establishes that AIC as the best preforming criterion for systems.

- We might also check to see what the error structure is like for each of the equations estimated using the type of LM test we first saw in the univariate context: $varlmar, mlag(5)$

7.7 Prediction

- The good news about vector ARMA prediction is that the rules are the same. The exact forecasting equation is a generalization of the formulae for univariate time series and will not be developed here. See Judge et al pp. 659 -660 as well as the STATA manual

- the command: $fcast$ provides dynamic forecasts (see the $var.do$ program)

7.8 Exogeneity and Causality

- The distinction between exogeneity and causation has been a source of much contention in the economics literature.

- There is an economic sense in which one variable is said to 'cause' another, say consumers consume on the basis of their wealth.

- Whereas exogeneity is a straightforward statistical concept (Engle, Hendry and Richard, 1983, **Econometrica**) that has to do with factoring joint probabilities into products of marginals.

- This latter form of causation does not necessarily have anything to do with economic causation and hence we have reserved a particular name for the kind of causation (statistical) tests that we conduct –namely **Granger causation**.

- A set of variables $\boldsymbol{z}_t$ is said to be Granger caused by a set of variables $\boldsymbol{x}_t$ if the information in the past and present $\boldsymbol{x}_t$ helps improve the **forecast** of $\boldsymbol{z}_t$.

- Formally, let I_t be the information set at time t (I_t includes as many other variables as you want to put into the information set).

- Let the 1-step forecast of $\boldsymbol{z}_t$ given I_t be (linear) $E[\boldsymbol{z}_{t+k} \mid I_t]$. We say that $\boldsymbol{x}_t$ **Granger causes** $\boldsymbol{z}_t$ if for some t

$$E[\boldsymbol{z}_{t+1} \mid I_t] \neq E\left[\boldsymbol{z}_{t+1} \mid \bar{I}_t\right] \tag{7.30}$$

where $\bar{I}_t$ is the set I_t with the set $\{\boldsymbol{x}_s \mid s \leq t\}$ removed.

- If the equality holds, we say that $\boldsymbol{z}_t$ is not Granger caused by $\boldsymbol{x}_t$. We say that $\boldsymbol{x}_t$ **causes** $\boldsymbol{z}_t$ **instantaneously** if:

$$E\left[\boldsymbol{z}_{t+1} \mid I_t \; U \; \{\boldsymbol{x}_{t+1}\}\right] \neq E\left[\boldsymbol{z}_{t+1} \mid I_t\right] \tag{7.31}$$

- If the equality holds, we say that $\boldsymbol{z}_t$ is not *Granger caused instantaneously* by $\boldsymbol{x}_t$.

- Note we can do the tests with the roles reversed. If we find that $\boldsymbol{x}_t$ Granger causes $\boldsymbol{z}_t$ and that $\boldsymbol{z}_t$ Granger causes $\boldsymbol{x}_t$ we say that there is **feedback**.

- The tests of Granger causation are usually significance testing. For instance suppose we take the VAR(p)

$$\begin{bmatrix} z_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Phi_{11,1} & \Phi_{12,1} \\ \Phi_{21,1} & \Phi_{22,1} \end{bmatrix} \begin{bmatrix} z_{t-1} \\ x_{t-1} \end{bmatrix} + \cdots + \begin{bmatrix} \Phi_{11,p} & \Phi_{12,p} \\ \Phi_{21,p} & \Phi_{22,p} \end{bmatrix} \begin{bmatrix} z_{t-p} \\ x_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \nu_t \end{bmatrix} \tag{7.32}$$

- We see that $\boldsymbol{z}_t$ does not Granger cause $\boldsymbol{x}_t$ iff $\boldsymbol{\Phi}_{21,i} = \mathbf{0}\,, \forall\, i = 1 \ldots, p$. Similarly z_t does not Granger-cause x_t if $\Phi_{12,i} = 0, \forall\; i = 1, \cdots p$.

- Note that for the testing to be complete

$$I_t \;=\; \{\boldsymbol{z}_s \mid s \leq t\}\, U\, \{\boldsymbol{x}_s \mid s \leq t\} \tag{7.33}$$

  which assumes that $\boldsymbol{z}_t$ and $\boldsymbol{x}_t$ contain all the relevant information in the determination of $\boldsymbol{z}_t$.

- For example if there is some **third** variable that Granger causes both, we may find that $\boldsymbol{x}_t$ may cause $\boldsymbol{z}_t$ even though in reality its the third variable influence. Thus there is an omitted variable in (7.32) and our results are tainted by specification error.

- STATA command: $v\arg ranger$ gives granger causation for each variable for each equation (including all variables (excluding own lags)

  - Wald tests are used (which may not be the best but of course these can be done from a single system estimation)

- It is easy to see how this framework might be extended to detect instantaneous causation (identifying assumption is that there is no instantaneous causation).

- Non Granger causation does not necessarily imply exogeneity (since we are only consider regression directions and not the probabilities themselves). See Engle, Hendry and Richard, 1983 and also Geweke, Meese and Dent, 1983 *Journal of Econometrics*.

7.9 Impulse Response Functions (Innovation Accounting)

- The basic idea of calculating impulse response functions is to determine the timepath of variables in response to an innovation or shock to one (or possibly some linear combination) of the variables in the model. In this way, we attempt to follow the interdependence over time.

- I am going to start the impulse response discussion following the notation I have used but soon we will adopt the standard notation for structural VAR. And to cap off the confusion, I will link to the STATA code which flips our notation right around from the conventional

- So we have the usual $VAR(p)$ in N variables

$$\begin{aligned} y_t &= \alpha + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + ... + \Phi_p y_{t-p} + \epsilon_t \\ \epsilon_t &\sim (0, \Omega) \end{aligned} \tag{7.34}$$

There are (at least) two difficulties in conducting such exercises (7.34).

1. The fact that we are interested in the innovations requires us to rewrite the VAR in terms of an infinite moving average representation (7.1). This permits a meaningful innovation exercise that avoids conditioning on some lag values of the $\boldsymbol{y}'s$.
$$y_t = \sum \Psi_i \epsilon_{t-i}$$

2. If we are interested in the response from a single shock it becomes difficult to identify the effects, if the errors are correlated (i.e. if $\boldsymbol{\Omega}$ is not block diagonal). With this in mind, it has been suggested by Sims and Litterman and others, that it is worthwhile to orthogonalize the variance-covariance structure. Cholesky decompositions for example

3. Others have argued that the restrictions implicit on the VAR from the standard triangular orthogonalizations are not reasonable and recently many other structures have been placed on the vector errors for the impulse response functions

7.9.1 Cholesky Orthogonalization

- There are many ways in which we may orthogonalize the variance covariance matrix $\boldsymbol{\Omega}$.

- If we normalize $\boldsymbol{\Psi}_0 = \boldsymbol{I}$, the identity matrix, and take a matrix $\boldsymbol{G}$ nonsingular such that:
$$\boldsymbol{G}^{-1}\boldsymbol{\Omega}(\boldsymbol{G}^T)^{-1} = \boldsymbol{I}. \tag{7.35}$$

- The transformed innovations are $\boldsymbol{\nu}_t = \boldsymbol{G}^{-1}\boldsymbol{\epsilon}_t$ and thus we are replacing $\boldsymbol{\epsilon}_t$ by $\boldsymbol{G}\boldsymbol{\nu}_t$ in (7.1):

$$\boldsymbol{y}_t = \sum_{j=0}^{k} \boldsymbol{\Psi}_j \boldsymbol{G}\boldsymbol{\nu}_{t-j} \tag{7.36}$$

  where we have assumed a finite truncation length k.

- This leaves the error term with the desirable property:

$$E\left[\boldsymbol{\nu}_t \boldsymbol{\nu}_t^T\right] = \boldsymbol{I}$$

  .

- These **orthogonalized innovations** have the convenient property that the are uncorrelated both **across time** and **across equations**.

- The errors in this equation may be shocked *individually* and the response to the shock say at time t may be followed using the above.

- This graphing of the timepath of $\boldsymbol{y}_t$ to a shock is called an **impulse response function**.

- Representation (7.36) can be used to decompose the variance for any forecast horizon from a unit shock (just as we did in the $ARMA$) –called the **forecast-error variance decomposition** ($FEVD$)

- The total FEVD can normalized to 1 (or 100%) for each variable at each horizon and the fraction or share of variance attributable from a one unit innovation in one of the orthogonalized variances be calculated.

- There is also no need to have zero mean variables and STATA permits there to be $\boldsymbol{X}$ variables in the model (These are called near VAR's).

7.9.2 STATA Commands

STATA has gone fairly nuts on the $IRF\imath s$ and there many many options

1. • The command syntax is:

$[TS]irf$– Create and analyze IRFs, dynamic-multiplier functions, and FEVDs
Syntax
irf subcommand ... [, ...]
subcommand description

---

create create IRF file containing IRFs, dynamic-multiplier functions, and FEVDs
set set the active IRF file
graph graph results from active file

cgraph combine graphs of IRFs, dynamic-multiplier functions, and FEVDs
ograph graph overlaid IRFs, dynamic-multiplier functions, and FEVDs
table create tables of IRFs, dynamic-multiplier functions, and FEVDs from active file
ctable combine tables of IRFs, dynamic-multiplier functions, and FEVDs
describe describe contents of active file
add add IRF results from one file to active file
drop drop IRF results from active file
rename rename IRF results within a file
The various subcommands for irf

1. irf shock the ordinary VAR residuals (almost never done since they are correlated in general and it makes little sense to shock just one error)
2. $oirf$ -orthogonalized impulse response functions
3. $dm-$dynamic multipliers when there are exogenous X variables (for example suppose we have a simple $VAR(1)$

$$\begin{aligned} y_t &= \Phi y_{t-1} + \Gamma X_{t-1} + \epsilon_t \text{ we could get the infinite } MA \\ y_t &= \sum_{i=0}^{\infty} \Psi_i \epsilon_{t-i} + \sum_{i=0}^{\infty} \Lambda_i X_{t-i} \text{ the } \Lambda_i \text{ are the dynamic multipliers} \end{aligned}$$

4. $cirf$- cumulative impulse response functions
5. $coirf$- cumulative orthogonalized impulse response functions
6. cdm- cumulative dynamic multipliers
7. $fevd$- Cholesky $fevd$
8. $sirf$- structural impulse response functions (to be discussed)
9. $sfevd-$structural forecast-error variance decompositions (to be discussed)

- In the Stata program *var.do* we conduct an impulse response exercise which can be done in a number of ways with various orderings and methods of calculating the standard errors.

7.9.3 Confidence Intervals for Impulse Response Functions

- Impulse response functions are complicated highly nonlinear functions of the estimated VAR parameters.
- STATA has a wide variety of ways to calculate confidence intervals

1. Use the Delta method (a linearization of the nonlinear function around the estimated values using the estimated variance covariance matrix) called asymptotic

2. Parametric Bootstrapping (Monte Carlo simulation drawing error structure from a normal distribution with variance covariance $\hat{\Omega}$ invoked by subcommand *bsp*

3. Bootstrap residuals using subcommand *bs*

Steps for Bootstrap or Monte Carlo (parametric bootstrap)

1. Draw vector of residuals or simulation from normal (T observations will be generated)

2. Use initial p observations from data, the sampled (simulated) residuals and the estimated coefficients to get T observations for each variable

3. Fit $VAR(p)$ and calculate the IRF and forecast error decompositions

4. Calculate average response and standard deviation over the draws for each step (however many steps you want to calculate)

5. Graph the average response ± 2 standard deviations for each step.

7.9.4 More on Orthogonalization

- We note that the matrix $\boldsymbol{G}$ can be any solution to $\boldsymbol{G}\boldsymbol{G}^T = \boldsymbol{\Omega}$ so that there are many possible choices for factorizing the positive definite $\boldsymbol{\Omega}$.

- For instances:

1. *Cholesky factorizations*, where $\boldsymbol{G}$ is chosen to be *lower triangular*. Given the $\boldsymbol{\Omega}$, there is only one lower triangular matrix $\boldsymbol{G}$ such that $\boldsymbol{G}\boldsymbol{G}^T = \boldsymbol{\Omega}$. However, a drawback of the procedure is that if we rearrange $\boldsymbol{\Omega}$ by simply rewriting the equations in a different order, we shall **change** the composition in $\boldsymbol{G}$ and hence change the impulse response function.

Example Let

$$\Omega = \begin{bmatrix} 1.0 & 4.0 \\ 4.0 & 25.0 \end{bmatrix} \text{ then the Cholesky factor } G = \begin{bmatrix} 1.0 & 0.0 \\ 4.0 & 3.0 \end{bmatrix}$$

rearranging the first and second equation gives:

- $$\Omega = \begin{bmatrix} 25.0 & 4.0 \\ 4.0 & 1.0 \end{bmatrix} \text{ then the Cholesky factor } G = \begin{bmatrix} 5.0 & 0.0 \\ .80 & .60 \end{bmatrix}$$

- Clearly the $G's$ are not invariant to simple rearrangement of the equations.

- The first type is referred to as a 1-2 order and the second is 2-1 order.

- It has been argued that the equations for this kind of decomposition should be ordered in such a way as the 'most important or related' variables appear first in the VAR, followed by less relevant ones. Of course, such nonspecifics are unhelpful to the practitioner and it is sometimes advisable that several different combinations and permutations should be considered.

- This interpretation suffers from a structural argument and is something that we are presumably trying to get away from in the VAR framework.

2. *Eigenvalue decompositions*

3. *Structural decompositions* suggested by Bernanke and Sims. These authors have noted the asymmetry in a mechanical VAR model and the imposition of semi-structural interpretation on the impulse response exercise. Instead they advocate a procedure that puts more economic structure on the problem from the outset.

- Let us consider a VAR(p) with N variates:

$$\boldsymbol{y}_t = \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{y}_{t-i} + \boldsymbol{\epsilon}_t \qquad E\left[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T\right] = \boldsymbol{\Omega} \tag{7.37}$$

- We can estimate the $\boldsymbol{\Phi}'s$ by OLS regardless of the assumptions on $\boldsymbol{\Omega}$.

- Now imagine the errors in (7.37) may be written in some form (whose specification is due to some economic causation argument), for instance for $N = 3$:

$$\begin{aligned} \epsilon_{1t} &= \nu_{1t} \\ \epsilon_{2t} &= \gamma \epsilon_{1t} + \nu_{2t} \\ \epsilon_{3t} &= \delta \epsilon_{1t} + \nu_{3t} \end{aligned} \tag{7.38}$$

  where the ν_t are uncorrelated.

- This system puts 1 restriction on the $\boldsymbol{\Omega}$, since there are six elements in the variance-covariance matrix but only five in (7.38): γ, δ and the variances of the ν_t.

- The idea is to place structure on $\boldsymbol{\Omega}$ explicitly as in (7.38) and then possibly test the restrictions.

- In (7.38) we see that the model has innovations in ϵ_{2t} and ϵ_{3t} related only through the ϵ_{1t}.

- The model is estimated by minimizing the following nonlinear problem:

$$\ln \mid \boldsymbol{\Omega} \mid + trace\ \boldsymbol{\Omega}^{-1} \boldsymbol{S}$$

  where $\boldsymbol{S}$ is the **sample** covariance matrix of residuals. The nonlinear minimization is accomplished by Newton's method.

7.10 Structural VARs

- These notes are an amalgamation of an article by Ben Bernanke (1986) and the notes in time series manual of STATA on Structural VARs.
- This section is divided into 3 parts.

1. Set-up the structural model and discuss the differences between the structural VAR analysis of Bernanke and a typical VAR analysis.
2. Estimation of the structural VAR as suggested by Bernanke and implemented by STATA respectively. The third section presents an example, from Bernanke, on how STATA estimates a structural VAR.
3. Conclude with the likelihood theory implicit in the STATA estimation.

7.10.1 Structural Model (Short-Run Restrictions)

With a slight change of notation (STATA notation in their manual completely flips standard notation but should be easy to follow) , let $\boldsymbol{Y}_t$ be an $n \times 1$ vector of macro variables, for relatively small n. We then specify the following structural model (VAR) for $\boldsymbol{Y}_t$:

$$\boldsymbol{Y}_t = \sum_{i=0}^{\ell} \boldsymbol{B}_i \boldsymbol{Y}_{t-i} + \boldsymbol{A}\boldsymbol{u}_t. \tag{7.39}$$

Notes:

1. $\boldsymbol{u}_t$ is the $n \times 1$ vector of structural errors. The number of structural errors in this setup is equal to the number of macro variables.
2. $\boldsymbol{A}$ is an $n \times n$ matrix of parameters. If $\boldsymbol{A} \neq \boldsymbol{I}$ then a structural error may affect more than one macro variable.
3. $E(\boldsymbol{u}_t \boldsymbol{u}_t^T) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is an $n \times n$ diagonal matrix. Thus the structural errors are uncorrelated.
4. Contemporaneous $\boldsymbol{Y}_t$ appears on the right-hand side in (7.39).
5. The model (7.39) has potentially many parameters to estimate and clearly would not be identified without some additional assumptions.
6. Bernanke suggests imposing no restrictions on the $\boldsymbol{B}$'s for $\boldsymbol{B}_i$, $i > 0$ except for the maximum lag length ℓ.
7. The identification restrictions are placed on the matrices $\boldsymbol{B}_0$ and $\boldsymbol{A}$. The model (7.39) is written in reduced-form.

$$\boldsymbol{Y}_t = \sum_{i=1}^{\ell} \boldsymbol{C}_i \boldsymbol{Y}_{t-i} + \boldsymbol{\kappa}_t \tag{7.40}$$

$$\boldsymbol{C}_i = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{B}_i$$

$$\boldsymbol{\kappa}_t = \boldsymbol{B}_0 \boldsymbol{\kappa}_t + \boldsymbol{A}\boldsymbol{u}_t \qquad \left[\boldsymbol{\kappa}_t = (\boldsymbol{I} - \boldsymbol{B}_0)^{-1} \boldsymbol{A}\boldsymbol{u}_t\right] \tag{7.41}$$

- The first step in a structural analysis of a VAR is to estimate the reduced-form VAR (7.40) and obtain the residuals $\boldsymbol{\kappa}_t$.
- The second step is to attempt to identify and estimate the purely contemporaneous model of these residuals (7.41).
- A typical VAR analysis would decompose the $\boldsymbol{\kappa}_t$'s into n orthogonal time series which are then used in economically meaningful ways. Bernanke argues that we actually want the structural errors (the $\boldsymbol{u}$'s). In order to obtain the $\boldsymbol{u}$'s we must postulate a *specific* structural model.
- For instance in the usual VAR setting, the decomposition into orthogonal series (conventionally a Cholesky decomposition) is equivalent to assuming $\boldsymbol{A} = \boldsymbol{I}$. For a specified ordering of the variables, $\hat{\boldsymbol{B}}$ is a lower-triangular matrix. This procedure does obtain an orthogonalization but assumes the structural model for $\boldsymbol{\kappa}$ is strictly **recursive**.
- The process of choosing different orderings is consistent if an investigator believes the system is recursive with an unknown ordering. Should the system not be recursive the "shock" series are meaningless.

7.10.2 Estimation

Estimation of Contemporaneous Model

Note: In his article Bernanke deals with a just-identified model, in the sense to be described below.

We already have the estimated $\boldsymbol{\kappa}$'s from the estimation of the reduced-form VAR. Then making use of a Method-of-Moments technique and equating population moments to sample moments of the $\boldsymbol{y}$'s, we proceed as follows. Set up the equation

$$\boldsymbol{u}_t \boldsymbol{u}_t^T = \boldsymbol{A}^{-1}(\boldsymbol{I} - \boldsymbol{B}_0)\boldsymbol{\kappa}_t \boldsymbol{\kappa}_t^T (\boldsymbol{I} - \boldsymbol{B}_0)^T (\boldsymbol{A}^{-1})^T \tag{7.42}$$

Let $\boldsymbol{M} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\kappa}_t \boldsymbol{\kappa}_t^T$ be the sample covariance matrix of the $\boldsymbol{\kappa}$'s. This matrix will not be a diagonal matrix if $\boldsymbol{A} \neq \boldsymbol{I}$.

Averaging (7.42) over the sample and equating population and sample moments we obtain

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{A}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{B}}_0)\boldsymbol{M}(\boldsymbol{I} - \hat{\boldsymbol{B}}_0)^T (\hat{\boldsymbol{A}}^{-1})^T \tag{7.43}$$

Now (7.43) defines the estimates $\hat{\mathbf{\Gamma}} = (\hat{\boldsymbol{A}}, \hat{\boldsymbol{B}}_0, \hat{\mathbf{\Sigma}})$ if the system is just-identified.
Identification: The nonzero elements of $\hat{\Gamma}$ will be identified when

1. The number of parameters does not exceed the number of distinct covariances in $\boldsymbol{M}$. This restricts the number of identifiable parameters to $\frac{n(n+1)}{2}$. We can thus identify the diagonal elements of $\hat{\mathbf{\Sigma}}$ and however many parameters are left in $\boldsymbol{A}$ and $\boldsymbol{B}_0$, which can equal $\frac{n(n-1)}{2}$.

2. Almost always the diagonal of B_0 is a diagonal matrix of 1's

3. The system of nonlinear equations (7.43) has at least one solution.

STATA Estimation of Structural VARS

STATA uses a Maximum Likelihood estimation procedure. In the notes we have assumed (1) is a just-identified system, and (2) Normality of the $\boldsymbol{u}$'s.[1]

Let $\boldsymbol{\phi}\boldsymbol{\kappa}_t = \boldsymbol{u}_t$, where $\boldsymbol{\phi} = \boldsymbol{A}^{-1}(\boldsymbol{I} - \boldsymbol{B}_0)$ and we have that

$$\begin{aligned} E(\boldsymbol{u}_t \boldsymbol{u}_t^T) &= \mathbf{\Sigma} \\ E(\boldsymbol{\kappa}_t \boldsymbol{\kappa}_t^T) &= \boldsymbol{M} \end{aligned}$$

Recall that $\mathbf{\Sigma}$ is a diagonal matrix and that $\boldsymbol{M}$ is non-diagonal.

The objective of the MLE procedure is to minimize over the free parameters of $\boldsymbol{\phi}$ and $\mathbf{\Sigma}$, the following likelihood function (excluding constants that do not depend on any of the parameters in $\boldsymbol{\phi}$ or $\mathbf{\Sigma}$):

$$\min_{\boldsymbol{\phi}, \mathbf{\Sigma}} \quad -2log|\boldsymbol{\phi}| + log|\mathbf{\Sigma}| + trace(\mathbf{\Sigma}^{-1}\boldsymbol{\phi}\boldsymbol{M}\boldsymbol{\phi}^T)$$

or when we concentrate the likelihood function with respect to $\mathbf{\Sigma}$

$$\min_{\boldsymbol{\phi}} \quad -2log|\boldsymbol{\phi}| + \sum_{i=1}^{T} log(\boldsymbol{\phi}\boldsymbol{M}\boldsymbol{\phi}^T)_{ii}$$

For the STATA procedure, we must a set of restrictions on the SVAR. The exact form of the restrictions will influence the estimates as well as the interpretations.

7.10.3 Example Linked to Stata

The following example deals with the residuals of the first stage VAR estimation (i.e. it is in the form of the first part of equation (7.41)). We have that $\boldsymbol{\kappa}^T = [\, g\ b\ m\ p\ y\ c\,]$,

[1] The STATA estimation procedure is not limited to just-identified systems. It can easily accommodate over-identified systems. These notes amalgamate the Bernanke article with the STATA procedure, and Bernanke assumes the just-identified system.

and the system is given by:

$$\begin{aligned} g &= u_1 \\ b &= \beta_1 g + \beta_2 m + \beta_3 p + \beta_4 y + u_2 \\ c &= \beta_5 b + \beta_6 p + \beta_7 y + u_3 \\ m &= \beta_8 b + \beta_9 p + \beta_{10} y + u_4 \\ p &= \beta_{11} y + u_5 \\ y &= \beta_{12} g + \beta_{13}(c-p) + \beta_{14}(m-p) + \alpha_1 u_5 + u_6 \end{aligned}$$

The only off-diagonal element of $\boldsymbol{A}$ is α_1. We also have that $n = 6$ which implies that we can estimate 15 parameters in the concentrated likelihood $(\beta_1, \ldots, \beta_{14})$ and, α_1.

- To assist you in reading the STATA manual (notation is completely opposite from standard for some reason) we have the following equivalence

$$\begin{aligned} A\kappa_t &= Bu_t \text{ where the A's and B's follow STATA} \\ A(I_N - A_1 L - A_2 L^2 - \ldots A_p) y_t &= A\epsilon_t = Be_t \\ &(\epsilon_t \text{ correlated VAR errors, } e_t \text{uncorrelated unit errors} \end{aligned}$$

Re-write this system in matrix form (following STATA notation) with ones on the main diagonal.

$$A = \begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ -\beta_1 & 1 & 0 & -\beta_2 & -\beta_3 & -\beta_4 \\ & -\beta_5 & 1 & 0 & -\beta_6 & -\beta_7 \\ 0 & -\beta_8 & 0 & 1 & -\beta_9 & -\beta_{10} \\ & & & & 1 & -\beta_{11} \\ -\beta_{12} & 0 & -\beta_{13} & -\beta_{14} & \beta_{13}+\beta_{14} & 1 \end{array}$$

and

$$B = \begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \alpha_1 & 1 \end{array}$$

so we input the matrix of constraints for B (in the usual way using *beq*) as diagonal matrix for the first 5 entries and last entry as

$$(0, 0, 0, 0, ., 1)$$

Notice the . (missing) is interpreted as a free parameter (not contrained) anw will be estimated. The A matrix for free and restricted(0 or 1) are written in the same way as we did for B. To impose the restriction that only real consumption and real money affecting real output we need to use the additional contraint

$$contraint\ 1\ [y]c+m+p=0$$

and add it as an option in the *svar* command line as $aconstraints(1)$

7.10.4 Deriving the STATA Estimation Procedure

Suppose we have the following specification

$$\boldsymbol{\phi}\boldsymbol{\kappa}_t = \boldsymbol{u}_t \qquad \boldsymbol{u}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

and we want to estimate the parameters of the matrices $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$. We proceed by building the likelihood function for $\boldsymbol{\kappa}_t$ which is derived from the likelihood for $\boldsymbol{u}_t$. The likelihood for $\boldsymbol{u}_t$ is given by

$$L_t = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left(-\frac{1}{2}\boldsymbol{u}_t^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u}_t\right)$$

The factor of $-\frac{n}{2}$ arises since we have a multivariate system (i.e. $\boldsymbol{u}_t$ is a $n \times 1$ vector and $\boldsymbol{\Sigma}$ is an $n \times n$ matrix). The contribution to the log likelihood is given by

$$\ell_t = -\frac{n}{2} log(2\pi) - \frac{1}{2} log|\boldsymbol{\Sigma}| - \frac{1}{2}\boldsymbol{u}_t^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u}_t. \tag{7.44}$$

In order to derive the log likelihood in terms of $\boldsymbol{\kappa}_t$ we need to calculate the Jacobian of the transformation. The Jacobian is:

$$\left|\frac{\partial \boldsymbol{u}_t}{\partial \boldsymbol{\kappa}_t}\right| = |\boldsymbol{\phi}|\,.$$

Using the Jacobian, the contribution to the log likelihood of $\boldsymbol{\kappa}_t$ is

$$\ell_t = -\frac{n}{2} log(2\pi) - \frac{1}{2} log|\boldsymbol{\Sigma}| + log|\boldsymbol{\phi}| - \frac{1}{2}\boldsymbol{\kappa}_t^T \boldsymbol{\phi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}\boldsymbol{\kappa}_t.$$

Now $\boldsymbol{\kappa}_t^T \boldsymbol{\phi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}\boldsymbol{\kappa}_t$ is a scalar, and the trace of a scalar equals itself (i.e. tr(a) = a). Hence

$$\begin{aligned} \boldsymbol{\kappa}_t^T \boldsymbol{\phi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}\boldsymbol{\kappa}_t &= trace\ (\boldsymbol{\kappa}_t^T \boldsymbol{\phi}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}\boldsymbol{\kappa}_t) \\ &= trace\ (\boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}\boldsymbol{\kappa}_t \boldsymbol{\kappa}_t^T \boldsymbol{\phi}^T) \\ &= trace\ (\boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}\boldsymbol{M}\boldsymbol{\phi}^T), \end{aligned}$$

where $\boldsymbol{M}$ is again the sample covariance of the $\boldsymbol{\kappa}_t$'s.

Excluding the constants and multiplying by -2, the contribution to the log likelihood for $\boldsymbol{\kappa}_t$ which is now *minimized* over the relevant parameters is:

$$\ell_t = -2log|\boldsymbol{\phi}| + \frac{1}{2} log|\boldsymbol{\Sigma}| + trace(\boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}\boldsymbol{M}\boldsymbol{\phi}^T).$$

This can be concentrated with respect to $\boldsymbol{\Sigma}$, giving the equation that is minimized by STATA :

$$\ell_t = -2log|\boldsymbol{\phi}| + \sum_{i=1}^{T} log(\boldsymbol{\phi}\boldsymbol{M}\boldsymbol{\phi}^T)_{ii}.$$

7.11 Structural Model (Long-Run Restrictions)

We can take the structural model from (equation 7.39)

$$Y_t = \sum_{i=0}^{l} B_i Y_{t-i} + Au_t$$

with a diagonal identification restrictions of 1, we can express as the infinite moving average process with $iid(0,1)$ u_t as

$$Y_t = C(L)u_t \tag{7.45}$$

Restictions on the $C(L)$ will imply long run restrictions. For example we may have 2 variables labelled 1 and 2 (say real income growth and money growth) where we want the long run response of variable 1 (income) to structural shocks from variable 2 (money) to be 0. In this case $C[1,2] = 0$ can be seens as imposing this restriction. In the example svar_lr.do we impose that neither variable has a long-run impact.

$$C[1,2] = \sum C_{12}(L) = C[2,1] = \sum C_{21}(L) = 0$$

Indentification

Recall if we have N variables there will be N^2 long-run parameters in C. SInce we have no restictions placed on the nature of the $SVAR$ here we have to assume that all variables appear in all equations (excluding the own variable assume zero) which means that the covariance matrix of the ordinary VAR must be restricted completely (no extra contemporanepous correlations which is ok given that all variables are in there). So this means for identification we must have

$$\frac{N(N+1)}{2} \quad \text{restrictions}$$

So in the case of $N = 2$ are example we can only test the exclusion of the two together (giving is one restirction). That is the order condtion for the number of restictions is

$$\begin{aligned} N^2 &= \frac{N(N+1)}{2} \quad \text{exact identification} \\ N^2 - \frac{N(N+1)}{2} &> 0 \quad \text{overidentification} \end{aligned}$$