Lecture 4: Time Series Econometrics

Vector Autoregressions (VARs), and Structural VARs

DANIEL BUNCIC

Institute of Mathematics & Statistics
University of St. Gallen
Switzerland

February 5, 2016 (tse4)

Table of Contents

1. Introduction and Background

We can think of Vector AutoregRessions (VARs) as multivariate generalisations of the univariate analogue AR models introduced earlier. It should be evident that with any univariate modelling approach that there are limitations as it is clearly difficult to think of a complex real world economic relation to be successfully described by one equation and one variable alone. Nonetheless, as we will see later on in Section 3.11, ARMA models can be viewed as *'final form equations'* coming from much bigger *'systems of equation'* models. Thus every variable in the structural system will have a ARMA structure when put in final equation form. Despite this fact, the final equation form is not very informative as it is a *'summative'* model, and therefore a just devise to summarise the properties of the data. It is not useful for policy analysis as there is no theoretical or structural model that can be attached to it to make recommendations of the form: *'when this control variable changes the response variable will change by . . . '*.

The main advantage of VAR models is that every variable in the system is interdependent and endogenously determined, as there are no *'incredible exogeneity assumptions'* (this is a term used famously by Chris Sims of Princeton now, formerly Minnesota), although these can of course be imposed. As we have seen with the univariate AR models, any set of variables will have an arbitrary large VAR representation under some conditions. It may seem natural to ask now why we do not generalise the treatment to VARMA type models? VARMA models do get a considerable textbook style treatment in Lütkepohl (2005), however, there are a number of identification issues when working with VARMA models, they are very often difficult to estimate numerically, and there exists the possibility of a non-fundamental representation so that it is not clear what VARMA model generated the data. Due to this, VARMA models are not used as widely as VAR models in empirical work.

Now it should be said here that there is a case to be made for VARMA models, especially when one tries to link the empirical model to some theoretical model such as a Dynamic stochastic general equilibrium (DSGE) model, as are commonly used in academic as well as central bank circles, to model the behaviour of individual agents as well as the whole economy in aggregate form. It is known that, in general, log-linearised DSGE models have a reduced form VARMA structure. Thus it may not always be possible to approximate the VARMA model *'well'* with a *'large'* lag order VAR model, given the size of available macroeconomic data. In fact, nearly all of the DSGE models that I know have a VAR(2) reduced from structure. Since it can be difficult to estimate and more so, interpret VARMA models and relate them back to the underlying and potentially hidden

theoretical model, we are not going to consider them in what follows in this Lecture.

VARs originally grew out of a dissatisfaction with traditional structural econometric models. The original development of the idea seems to have been due to Quenouille (1957) but it was not exploited much as a representation of multivariate series until Sims (1980) proposed that it be used for the analysis of macroeconomic data. Sims (1980) was concerned that the model building style of the 1960s tended to be rather restrictive in its treatment of dynamics and that it was desirable to leave the dynamics as free as possible so as to avoid any distortion of the evidence in the data. One could justify this stance by looking at theoretical developments at that time, particularly that due to the rational expectations school who had made a strong argument that it was difficult to claim that any lagged value should be excluded from any equation. Hence, it was obviously of interest to determine how far one could go with models that made very few assumptions. Clearly, this was very much in a *data analysis* tradition. Rather than *great ratios* or graphs, one was going to utilise observed data correlations to shed light upon questions of economic interest.

It should be pointed out here that there has been a bit of a re-emergence of VAR modelling in the last 15 years or so, where the popularity of VARs has increased again after a number of years of a slow down say from $1990 - 2000$ onwards and this has to do largely with the development of DSGE models and the close link to a (restricted) VAR(2) type of structure of the solved reduced form DSGE model. There has been a bit of a quarrel with respect to the usefulness of VAR models for policy analysis ever since Sims (1980) popularised the use of VARs and there have been many opponents to VAR models. Arnold Zellner[1] once coined the acronym VARs to stand for **V**ery **A**wful **R**egressions because of its near non-sense inclusion of what ever is available in terms of variables to include in the system, without any *a priori* consideration of the economic structure and validity of the variables. Another well known opponent is Andrew Harvey, formerly of LSE now of Cambridge University, who has been favouring a more structural approach to macroeconomic modelling. There is now a reconciling view that we can think of VARs as a form of a structural model and this view is outlined and discussed in more detail in Fernández-Villaverde *et al.* (2007).

It should be stressed here that there is an intertwined **relationship** between the **order of the VAR process**, that is the lag length, and the **variables** that are selected to be included in the VAR (the dimension of the VAR). One has to be very careful to clearly set out theoretical arguments to choose the variables that go into a VAR. There is no point to

[1] A very famous statistician and the father of modern Bayesian statistics as well as the SUR model.

just through into the model whatever and press buttons in a software package because it is available. There is always the danger of running a *'garbage in, garbage out'* VAR and then trying to use it for policy analysis by looking at the impulse responses and other quantities of interest without having a clear theoretical model or idea of what to make of the VAR (hence the Zellner label Very Awful Regression).

A short comment on notation. We have used α for the parameters of the AR model in the univariate case. I will continue to use α for the parameters of the AR model when I make comparisons between the conceptual similarity of VAR and AR models. Nonetheless, I will simply use lower case letters a_{ij}^{p} to denote the elements of the VAR parameter matrix of interest to be consistent with the notation commonly used in this literature. The subscripts (i, j) refer to row and column entries, while the superscripts denote the matrix on the p^{th} lag. I will always use k to denote the number of variables in the VAR and p for the lag order of the VAR. Also, I will use simple (non-bold) capital letters to denote matrices of the VAR model, and will use bold face letters for the corresponding companion form of the VAR(p) model.

2. VAR Model Structure and its Analysis

2.1. Overview within a bivariate model

To introduce the ideas behind VAR models, let us begin with a simple bivariate VAR(1) taking the form:

$$x_{1t} = a_{11}x_{1t-1} + a_{12}x_{2t-1} + u_{1t} \tag{1a}$$

$$x_{2t} = a_{21}x_{1t-1} + a_{22}x_{2t-1} + u_{2t} \tag{1b}$$

where $u_{jt}, \forall j = 1, 2$, are bivariate Normal with a mean of zero and covariance matrix Σ, where

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}. \tag{2}$$

We will generalise the model later.

One of the earliest uses of VARs was to test for causality and we will formalise the testing for causality a little later in Section 4. It seems obvious that, if a theory made a prediction that one variable caused another, one would expect the data to show those movements in such a way that the **causal** variable would precede the **caused** variable. Within the context of (1b), x_{1t} is said to **cause** x_{2t} if $a_{21} \neq 0$ and $a_{12} = 0$. Thus, VARs

provided a simple way of looking for **precedence** in relations since it just involved fitting the VAR in (1) and then testing if coefficients were zero. There seems no harm in such an exercise as that is essentially what we would do if we were looking at graphs of x_{1t} and x_{2t} and trying to determine which came first. Obviously, such methods are much more attractive then graphical methods once we have more than two variables involved in a relation. It would be surprising, however, if many questions could be settled in this way. Perhaps the biggest difficulty in using such an approach is that the method focuses on a timing relationship which may have only a tenuous connection with causality. Edward Leamer's famous "**Christmas cards cause Christmas**" example shows its weakness in this direction. It seems that this type of use of VARs has declined a lot in recent years, although most models do have something that is viewed as **exogenous** e.g. government expenditure, the foreign sector, productivity shocks, and it is natural to investigate whether the variables that measure such quantities need only their own history as explanatory.

Note that the equations in (1) by itself are simply just a convenient way to **summarise some of the correlations within the data**, whereas the sort of question asked by economists are of a different form, more along the lines of *"what will happen to x_1 if another variable changes?"*. Thus in regression analysis where

$$x_{1t} = \delta x_{2t} + error \tag{3}$$

one could ask what would happen to x_{1t} as x_{2t} changes. This is a very common mode of investigation. What the VAR structure points out is that this is a very poorly defined question. Because x_{2t} depends upon the past history of x_{1t}, x_{2t} as well as the current shocks (impulses), u_{1t} and u_{2t}, one has to be more specific about which component it is that *'causes'* x_{2t} to change. In VAR analysis it is always the contemporaneous shocks u_{1t} and u_{2t} that are varied to effect any change, and this precision is an advantage of VAR analysis.

A major part of Sims' argument was to point out that while the VAR might be a good format for statistical analysis it was **not** necessarily the best one for interpretation or for policy analysis, partly because there was no focus on theoretical economic models. It was more meaningful in a VAR to present the dynamic relations linking the outputs x_{1t} and x_{2t} with their **impulses**. Thus, the **impulse response function (IRF)** has become the primary tool when working with VAR models and it will be informative briefly to outline what is meant by that before we formalise things later.

Suppose we just had (1a) with $a_{12} = 0$. Then, it is a simple matter to express x_{1t} as a univariate AR(1) model taking the form

$$\begin{aligned} x_{1t} &= a_{11}x_{1t-1} + u_{1t} \\ (1 - a_{11}L)x_{1t} &= u_{1t} \end{aligned} \tag{4}$$

with $|a_{11}| < 1$ which we can then invert simply to its MA infinity representation of the AR in (1a), to yield

$$x_{1t} = u_{1t} + a_{11}u_{1t-1} + (a_{11})^2 u_{1t-2} + ... \tag{5}$$

We would then ask questions such as, "*what is the effect of a (unit) change in the innovation or impulse of shock u_{1t} upon x_{1t} in m periods time*" by looking at the the impulse response coefficients a_{11}, $(a_{11})^2$, $(a_{11})^3$ ... to see the effect of such a shock on the economic variable, m periods into the future. For example, the effect five periods from now of a shock in u_{1t} would be $(a_{11})^5$. These coefficients are termed the **impulse response coefficients** and they are commonly plotted against m to visualise the dynamic behaviour of x_{1t} to shocks that come into the model through u_t. In general k variable VAR(p) models, there will be various shocks in the model and what these exactly are needs to be carefully defined/identified in the system model. So this is the policy use of VAR models.

Fundamentally, VARs will always be used in these two ways:

1) as a summative, reduced form model, which can be used for forecasting and summarising the data.
2) as a policy model to *'tell stories'* about how shocks that come into the system affect the dynamic behaviour of the variables in the system.

3. VAR mechanics

Before we get into the policy use of VARs, let us initially outline the general structure and notation used in VARs and then introduce concepts related to stability/stationarity, second moment properties and cross-correlations in VARs.

3.1. General VAR notation

Let us formalise the k variable VAR to a more general structure, taking on the (reduced and normalised) VAR(p) form (what these two terms exactly mean will become clear later when we talk about structural VARs):

$$A(L)X_t = C + U_t \tag{6}$$

where

$$A(L) = I_k - A_1 L - A_2 L^2 - ... - A_p L^p \tag{7}$$

$$X_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{kt} \end{bmatrix}, C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}, U_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{kt} \end{bmatrix} \tag{8}$$

and $U_t \sim \mathsf{WN}(0, \Sigma)$ (White Noise in vector form) with mean 0 and variance/covariance matrix Σ given by

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22}^2 & \cdots & \sigma_{2k} \\ \vdots & \cdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk}^2 \end{bmatrix}. \tag{9}$$

The matrix of coefficients on the ℓ^{th} lag of X_t are given by

$$A_\ell = \begin{bmatrix} a_{11}^\ell & a_{12}^\ell & \cdots & a_{1k}^\ell \\ a_{21}^\ell & a_{22}^\ell & \cdots & a_{2k}^\ell \\ \vdots & \cdots & \ddots & \vdots \\ a_{k1}^\ell & a_{k2}^\ell & \cdots & a_{kk}^\ell \end{bmatrix}, \tag{10}$$

where the subscripts attached to elements in A index the row and column entries of the ℓ^{th} lag matrix linked to $X_{t-\ell}$.

3.2. Companion form of the VAR(p) model

As in the univariate AR(p) model, we can put the VAR(p) into **companion form**. That is, we can re-express the VAR(p) as:

$$\mathbf{X}_t = \mathbf{C} + \mathbf{A}\mathbf{X}_{\mathbf{t-1}} + \mathbf{U}_t \tag{11}$$

where

$$\underset{(pk\times 1)}{\mathbf{X}_t} = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix}, \underset{(pk\times 1)}{\mathbf{X}_{t-1}} = \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p} \end{bmatrix}, \underset{(pk\times 1)}{\mathbf{C}} = \begin{bmatrix} C \\ 0_k \\ \vdots \\ 0_k \end{bmatrix}, \underset{(pk\times 1)}{\mathbf{U}_t} = \begin{bmatrix} U_t \\ 0_k \\ \vdots \\ 0_k \end{bmatrix} \tag{12}$$

and

$$\underset{(pk\times pk)}{\mathbf{A}} = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_k & 0_k & \cdots & 0_k & 0_k \\ 0_k & I_k & \cdots & 0_k & 0_k \\ \vdots & \cdots & \ddots & \vdots & \vdots \\ 0_k & 0_k & \cdots & I_k & 0_k \end{bmatrix}, \underset{(k\times k)}{I_k} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{13}$$

$$\underset{(k\times k)}{0_k} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The vector white noise process $\mathbf{U}_t$ has mean $\mathbf{0}_{pk}$ and a variance/covariance matrix $\mathbf{\Sigma}_U$ which takes the form

$$\underset{(pk\times pk)}{\mathbf{\Sigma}_U} = \underbrace{\begin{bmatrix} \Sigma & 0_k & \cdots & 0_k \\ 0_k & 0_k & \cdots & 0_k \\ \vdots & \vdots & \ddots & \vdots \\ 0_k & 0_k & \cdots & 0_k \end{bmatrix}}_{(p \text{ times } 0_k)}. \tag{14}$$

3.3. Stationarity (Stability)

The rules for stationarity in VARs extend naturally from those in univariate AR model. Thus, the VAR in (6) is stable (or stationary) if

$$\det(I_k - A_1 z - A_2 z^2 - ... - A_p z^p) \neq 0 \text{ for } |z| \leq 1. \tag{15}$$

That is, the roots of the lag polynomial $A(L)$ (with z replacing L again) cannot be equal to zero for any z smaller than or equal to 1 in absolute value. That is, if the roots z are greater than 1 in absolute value or modulus ($|z| > 1$), then the process is stationary (or stable). Alternatively, we can again look at the $\mathbf{A}$ matrix of the companion form of the

VAR in (13) and check that $|\text{eig}(\mathbf{A})| < 1$

3.4. Vector Moving Average VMA(∞) representation

If the VAR is stable (or stationary), then (6) can be inverted into its vector MA(∞) (abbreviated as VMA(∞)) form:

$$\begin{aligned} A(L)X_t &= C + U_t \\ &= A(L)^{-1}C + A(L)^{-1}U_t \\ X_t &= \mu + \Psi(L)U_t \end{aligned} \tag{16}$$

where

$$\mu = A(1)^{-1}C \tag{17}$$

and

$$\Psi(L) = I_k + \Psi_1 L + \Psi_2 L^2 + \ldots \tag{18}$$

The Ψ coefficients can again be obtained from the same relation as in the AR(p) to MA(∞) transformation, that is, by solving

$$A(L)\Psi(L) = I_k \tag{19}$$

$$(I_k - A_1 L - A_2 L^2 - \ldots - A_p L^p)(\Psi_0 + \Psi_1 L + \Psi_2 L^2 + \ldots) = I_k \tag{20}$$

to yield

$$\begin{aligned} \Psi_0 &= I_k \\ \Psi_1 &= A_1\Psi_0 \\ \Psi_2 &= A_1\Psi_1 + A_2 \\ \Psi_3 &= A_1\Psi_2 + A_2\Psi_1 + A_3 \\ &\vdots \end{aligned}$$

or more generally, we have the recursion (difference equation)

$$\Psi_i = A_1\Psi_{i-1} + A_2\Psi_{i-2} + A_3\Psi_{i-3} + \ldots + A_p\Psi_{i-p} \tag{21}$$

for all $i = 1, 2, 3, \ldots$, $\Psi_0 = I_k$, and $\Psi_n = 0$ if $n < 0$.

A computationally convenient way to generate the sequence $\{\Psi_i\}_{i=0}^{\infty}$ is to make use

of the VAR(1) properties of the companion form of the VAR(p) model. For a VAR(1), we know that the Ψ weights are defined as $\Psi_i = A_1^i$ for all $i = 0, 1, 2, \ldots$, where A_1 is the parameter matrix on the first lag of X_t. For the VAR(p) in companion form, this translates into selecting the first row and column entries of the $\mathbf{A}$ matrix raised to the power i. That is,

$$\Psi_i = \mathbf{S}\mathbf{A}^i\mathbf{S}' \tag{22}$$

for all $i = 0, 1, 2, 3, \ldots$, where $\mathbf{S}$ is a $(k \times kp)$ selection matrix defined as

$$\mathbf{S} = \underbrace{[I_k \; 0_k \; \ldots \; 0_k]}_{p \text{ times}}. \tag{23}$$

3.5. Mean and Covariance

3.5.1. Mean

The mean of the VAR is most easily obtained from the VMA(∞) representation in (16) by simply taking the expectation operator through to obtain

$$\begin{aligned} E(X_t) &= \mu \\ \mu &= A(1)^{-1}C \end{aligned} \tag{24}$$

since $E[\Psi(L)U_t] = \Psi(L)E[U_t]$.

Using the **companion from** of the VAR(p) process defined in (11), the mean is found as:

$$\underset{(k\times 1)}{\mu} = \mathbf{S}\underbrace{(\mathbf{I}_{pk} - \mathbf{A})^{-1}\mathbf{C}}_{=\boldsymbol{\mu}} \tag{25}$$

where

$$\begin{aligned} \boldsymbol{\mu} &= (\mathbf{I}_{pk} - \mathbf{A})^{-1}\mathbf{C} \\ &= \left(\underset{(pk\times pk)}{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \vdots & 0 \\ \vdots & \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}} - \underset{(pk\times pk)}{\begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_k & 0_k & \cdots & 0_k & 0_k \\ 0_k & I_k & \cdots & 0_k & 0_k \\ \vdots & \cdots & \ddots & \vdots & \vdots \\ 0_k & 0_k & \cdots & I_k & 0_k \end{bmatrix}} \right)^{-1} \underset{(pk\times 1)}{\begin{bmatrix} C \\ 0_k \\ 0_k \\ \vdots \\ 0_k \end{bmatrix}} = \underset{(pk\times 1)}{\begin{bmatrix} \mu \\ \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix}} \end{aligned} \tag{26}$$

and **S** is the selection matrix defined in (23).

3.5.2. Covariance matrix

The covariance matrix of X_t, $\Gamma_X(j)$ defined on the demeaned series $\tilde{X}_t = (X_t - \mu)$ is

$$\begin{aligned}\underbrace{\Gamma_X(j)}_{(k\times k)} &= E[\tilde{X}_t \tilde{X}'_{t-j}] \\ &= E\begin{bmatrix} \tilde{x}_{1t}\tilde{x}_{1t-j} & \tilde{x}_{1t}\tilde{x}_{2t-j} & \cdots & \tilde{x}_{1t}\tilde{x}_{kt-j} \\ \tilde{x}_{2t}\tilde{x}_{1t-j} & \tilde{x}_{2t}\tilde{x}_{2t-j} & \cdots & \tilde{x}_{2t}\tilde{x}_{kt-j} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{kt}\tilde{x}_{1t-j} & \tilde{x}_{kt}\tilde{x}_{2t-j} & \cdots & \tilde{x}_{kt}\tilde{x}_{kt-j} \end{bmatrix}\end{aligned} \tag{27}$$

which for the VAR in (6) become

$$\begin{aligned}E[\tilde{X}_t \tilde{X}'_{t-j}] = A_1 E[\tilde{X}_{t-1}\tilde{X}'_{t-j}] + A_2 E[\tilde{X}_{t-2}\tilde{X}'_{t-j}] + \ldots \\ + A_p E[\tilde{X}_{t-p}\tilde{X}'_{t-j}] + E\left[U_t \tilde{X}'_{t-j}\right]\end{aligned}$$

so that for $j > 0$ we have

$$\Gamma_X(j) = A_1\Gamma_X(j-1) + A_2\Gamma_X(j-2) + \ldots + A_p\Gamma_X(j-p) \tag{28}$$

and for $j = 0$ we get

$$\Gamma_X(0) = A_1\Gamma_X(-1) + A_2\Gamma_X(-2) + \ldots + A_p\Gamma_X(-p) + \Sigma. \tag{29}$$

Note that due to

$$\Gamma_X(-j) = \Gamma_X(j)' \tag{30}$$

we have further that

$$\Gamma_X(0) = A_1\Gamma_X(1)' + A_2\Gamma_X(2)' + \ldots + A_p\Gamma_X(p)' + \Sigma \tag{31}$$

where also the relevant quantities in (28) will be affected by this whenever $\Gamma_X(j-i) < 0, \forall i = 1, \ldots, p$.

The result in (30) that $\Gamma_X(-j) = \Gamma_X(j)'$ can be most easily seen from the following

bivariate VAR example. Let

$$X_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \tag{32}$$

thus

$$\underbrace{E\left[X_t X_{t-1}'\right]}_{\Gamma_X(1)} = E\left(\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}\begin{bmatrix} x_{1t-1} & x_{2t-1} \end{bmatrix}\right) = E\begin{bmatrix} x_{1t}x_{1t-1} & x_{1t}x_{2t-1} \\ x_{2t}x_{1t-1} & x_{2t}x_{2t-1} \end{bmatrix} \tag{33}$$

but

$$\underbrace{E\left[X_{t-1} X_t'\right]}_{\Gamma_X(-1)} = E\left(\begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix}\begin{bmatrix} x_{1t} & x_{2t} \end{bmatrix}\right) \quad = E\begin{bmatrix} x_{1-1t}x_{1t} & x_{1t-1}x_{2t} \\ x_{2-1t}x_{1t} & x_{2t-1}x_{2t} \end{bmatrix} \tag{34}$$

so we have $\Gamma_X(-1) = \Gamma_X(1)'$.

Now for a k dimensional VAR(1) we get from the relations in (28) and (31)

$$\begin{aligned} \Gamma_X(0) &= A_1\Gamma_X(-1) + \Sigma \\ &= A_1\Gamma_X(1)' \ + \Sigma \\ \Gamma_X(1) &= A_1\Gamma_X(0) \end{aligned}$$

Note here that for two conformable matrices A and B we have $(\mathsf{AB})' = \mathsf{B}'\mathsf{A}'$ and since $\Gamma_X(0)$ is a covariance matrix it is symmetric (and positive definite) so that $\Gamma_X(0) = \Gamma_X(0)'$. Thus

$$\begin{aligned} \Gamma_X(0) &= A_1\left[A_1\Gamma_X(0)\right]' + \Sigma \\ &= A_1\Gamma_X(0)A_1' + \Sigma. \end{aligned} \tag{35}$$

and more generally, we get the recursive relation

$$\Gamma_X(j) = A_1\Gamma_X(j-1).$$

Using the $\mathsf{vec}(\cdot)$ operator, which stacks a matrix into a vector, we can convert the relation involving the matrices in (35) into one involving vectors which yields

$$\begin{aligned} \mathsf{vec}[\Gamma_X(0)] &= \mathsf{vec}[A_1\Gamma_X(0)A_1' + \Sigma] \\ &= A_1 \otimes A_1\mathsf{vec}[\Gamma_X(0)] + \mathsf{vec}(\Sigma) \end{aligned}$$

$$\text{vec}[\Gamma_X(0)] - A_1 \otimes A_1 \text{vec}[\Gamma_X(0)] = \text{vec}(\Sigma)$$

$$\text{vec}[\Gamma_X(0)] = \underbrace{(I_{k^2} - A_1 \otimes A_1)^{-1}}_{\text{invertible if } |\text{eig}(A_1)|<1} \text{vec}(\Sigma) \tag{36}$$

where the relation in (36) depends on the invertibility of $(I_{k^2} - A_1 \otimes A_1)$. If the eigenvalues of A_1 are less than 1 in modulus, then the inverse of $(I_{k^2} - A_1 \otimes A_1)$ will exist.

Remark 1: *To get the result above, we have made use of the following* vec *rules for the three conformable matrices A, B, and C:*

1. $\text{vec}\,(\text{ABC}) = (\text{C}' \otimes A)\,\text{vec}(\text{B})$ *and*
2. $\text{vec}\,(\text{A} + \text{B}) = \text{vec}(\text{A}) + \text{vec}(\text{B})$.

as well as the Kroneker product $\otimes$, where $A \otimes B$ means that each element of A is multiplied by the matrix B, so that

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nm}B \end{bmatrix} \tag{37}$$

for the $n \times m$ matrix A and any sized matrix B.

We can generalise the results obtained for the VAR(1) process to that of a k dimensional VAR(p), which, again written in companion form (with zero mean) is:

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{U}_t. \tag{38}$$

We then get

$$\mathbf{\Gamma_X}(0) = \mathbf{A}\mathbf{\Gamma_X}(0)\mathbf{A}' + \mathbf{\Sigma}_U, \tag{39}$$

where

$$\mathbf{\Gamma_X}(0) = \begin{bmatrix} \Gamma_X(0) & \Gamma_X(1) & \cdots & \Gamma_X(p-1) \\ \underbrace{\Gamma_X(-1)}_{\Gamma_X(1)'} & \Gamma_X(0) & \cdots & \Gamma_X(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \underbrace{\Gamma_X[-(p-1)]}_{\Gamma_X(p-1)'} & \underbrace{\Gamma_X[-(p-2)]}_{\Gamma_X(p-2)'} & \cdots & \Gamma_X(0) \end{bmatrix}.$$

Performing the same operations on (39) as before (33) and (34), we obtain the relation:

$$\text{vec}[\mathbf{\Gamma_X}(0)] = (\mathbf{I}_{(pk)^2} - \mathbf{A} \otimes \mathbf{A})^{-1}\text{vec}(\mathbf{\Sigma}_U). \tag{40}$$

3.6. Autocorrelations

Autocorrelation matrices can be computed as

$$R_X(h) = d^{-1}\Gamma_X(h)d^{-1}, \tag{41}$$

where $d = \text{diag}\left[\Gamma_X(0)\right]^{1/2}$ and $\text{diag}(X)$ is the diagonal of the matrix X.

3.7. An Example

Let us now look at an example that goes through each of these computations with some numbers plugged in so that it is easier to conceptualise what these equations should look like.

Example 1: Suppose we have the following system

$$\begin{aligned} x_{1,t} &= c_1 + a_{11}^1 x_{1,t-1} + a_{11}^2 x_{1,t-2} + a_{12}^1 x_{2,t-1} + a_{12}^2 x_{2,t-2} + u_{1,t} \\ x_{2,t} &= c_2 + a_{21}^1 x_{1,t-1} + a_{21}^2 x_{1,t-2} + a_{22}^1 x_{2,t-1} + a_{22}^2 x_{2,t-2} + u_{2,t} \end{aligned}$$

where

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim \text{WN}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}\right)$$

In matrix polynomial form we can express the system as

$$A(L)X_t = C + U_t$$

where $A(L) = I_2 - A_1 L - A_2 L^2$.

With

$$\begin{aligned} A_1 &= \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.3 \end{bmatrix}, A_2 = \begin{bmatrix} -0.2 & 0.1 \\ -0.3 & 0.2 \end{bmatrix} \\ \Sigma &= \begin{bmatrix} 1.75 & 0.25 \\ 0.25 & 3.00 \end{bmatrix}, C = \begin{bmatrix} 0.2 \\ 0.3 \end{bmatrix} \end{aligned}$$

we get the roots of the lag polynomial as $\det(I_2 - A_1 L - A_2 L^2)$, that is, as:

$$\det\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}0.5 & 0.1\\ 0.4 & 0.3\end{bmatrix} z - \begin{bmatrix}-0.2 & 0.1\\ -0.3 & 0.2\end{bmatrix} z^2\right)$$

$$= \quad -z^4/100 + z^3 3/20 + z^2 11/100 - z4/5 + 1$$

which has roots $|z_i| > 1,\ i = 1, \ldots 4$ at -4.4981, $2.8686 \pm 2.0966i$, 1.7610.

Similarly, the eigenvalues of the companion form can be computed from

$$\mathbf{A} = \begin{bmatrix}A_1 & A_2\\ I_2 & 0_2\end{bmatrix}$$

as $\text{eig}\,(\mathbf{A}) = \det\,(\mathbf{A} - \lambda I_4)$

$$\text{eig}\,(\mathbf{A}) \quad = \quad \det\left(\begin{bmatrix}0.5 & 0.1 & -0.2 & 0.1\\ 0.4 & 0.3 & -0.3 & 0.2\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix} - \lambda \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\right)$$

$$= \quad \lambda^4 - 4/5\lambda^3 + 11/100\lambda^2 + 3/100\lambda - 1/100$$

with roots $|\lambda_i| < 1,\ i = 1, \ldots 4$ at 0.5679, -0.2223, $0.2272 \pm 0.1661i$. So the model is stable/stationary.

The unconditional mean we get from the relation $\mu = A(1)^{-1} C$ which is

$$\left(\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}0.5 & 0.1\\ 0.4 & 0.3\end{bmatrix} - \begin{bmatrix}-0.2 & 0.1\\ -0.3 & 0.2\end{bmatrix}\right)^{-1} \begin{bmatrix}0.2\\ 0.3\end{bmatrix} = \begin{bmatrix}0.4848\\ 0.6970\end{bmatrix}.$$

The variance/covariance matrices are computed from

$$\text{vec}(\mathbf{\Gamma}_{\mathbf{X}}(0)) \quad = \quad (\mathbf{I}_{(pk)^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec}(\mathbf{\Sigma}_W)$$

$$= \quad \left(\underbrace{\mathbf{I}_{(4)^2}}_{\mathbf{I}_{16}} - \begin{bmatrix}0.5 & 0.1 & -0.2 & 0.1\\ 0.4 & 0.3 & -0.3 & 0.2\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix} \otimes \begin{bmatrix}0.5 & 0.1 & -0.2 & 0.1\\ 0.4 & 0.3 & -0.3 & 0.2\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}\right)^{-1} \text{vec}(\mathbf{\Sigma}_W)$$

where

$$\mathbf{\Sigma}_W = \begin{bmatrix} \Sigma & 0_2 \\ 0_2 & 0_2 \end{bmatrix} = \begin{bmatrix} 1.75 & 0.25 & 0 & 0 \\ 0.25 & 3.00 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From Matlab we get (after reshaping to appropriate matrix dimensions)

$$\mathbf{\Gamma_X}(0) = \begin{bmatrix} \Gamma_X(0) & \Gamma_X(1) \\ \Gamma_X(1)' & \Gamma_X(0) \end{bmatrix}$$

$$= \left[\begin{array}{cc:cc} 2.4457 & 1.0531 & 1.1803 & 0.8820 \\ 1.0531 & 4.1304 & 1.1165 & 1.6568 \\ \hdashline 1.1803 & 1.1165 & 2.4457 & 1.0531 \\ 0.8820 & 1.6568 & 1.0531 & 4.1304 \end{array}\right]$$

so that the variance $\Gamma_X(0)$ and first autocovariance $\Gamma_X(1)$ of the bi-variate VAR(2) are:

$$\Gamma_X(0) = \begin{bmatrix} 2.4457 & 1.0531 \\ 1.0531 & 4.1304 \end{bmatrix} \text{ and } \Gamma_X(1) = \begin{bmatrix} 1.1803 & 0.8820 \\ 1.1165 & 1.6568 \end{bmatrix}$$

respectively, for $j = 0, 1$. The remaining autocovariances can be computed from the recursive formula:

$$\Gamma_X(j) = A_1\Gamma_X(j-1) + A_2\Gamma_X(j-2) \tag{42}$$

for $j \geq 2$ with $\Gamma_X(0)$ and $\Gamma_X(1)$ taken as initial conditions.

Thus, $\Gamma_X(j)$ for $j = 2, 3, ...$ is computed as

$$\Gamma_X(2) = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} 1.1803 & 0.8820 \\ 1.1165 & 1.6568 \end{bmatrix} + \begin{bmatrix} -0.2 & 0.1 \\ -0.3 & 0.2 \end{bmatrix} \begin{bmatrix} 2.4457 & 1.0531 \\ 1.0531 & 4.1304 \end{bmatrix}$$

$$\Gamma_X(2) = \begin{bmatrix} 0.3180 & 0.8091 \\ 0.2840 & 1.3600 \end{bmatrix}$$

$$\Gamma_X(3) = \begin{bmatrix} 0.5 & 0.1 \\ 0.4 & 0.3 \end{bmatrix} \begin{bmatrix} 0.3180 & 0.8091 \\ 0.2840 & 1.3600 \end{bmatrix} + \begin{bmatrix} -0.2 & 0.1 \\ -0.3 & 0.2 \end{bmatrix} \begin{bmatrix} 1.1803 & 0.8820 \\ 1.1165 & 1.6568 \end{bmatrix}$$

$$\Gamma_X(3) = \begin{bmatrix} 0.0630 & 0.5298 \\ 0.0816 & 0.7984 \end{bmatrix}$$

$$\vdots$$

To compute the autocorrelations, we need to construct d for the relation in (41). For

our example above we get

$$d = \begin{bmatrix} \sqrt{2.4457} & 0 \\ 0 & \sqrt{4.1304} \end{bmatrix} = \begin{bmatrix} 1.5639 & 0 \\ 0 & 2.0323 \end{bmatrix}$$

and thus

$$\begin{aligned} R_X(0) &= d^{-1}\Gamma_X(0)d^{-1} \\ &= \begin{bmatrix} \frac{1}{1.5639} & 0 \\ 0 & \frac{1}{2.0323} \end{bmatrix} \begin{bmatrix} 2.4457 & 1.0531 \\ 1.0531 & 4.1304 \end{bmatrix} \begin{bmatrix} \frac{1}{1.5639} & 0 \\ 0 & \frac{1}{2.0323} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0.3314 \\ 0.3314 & 1 \end{bmatrix} \end{aligned}$$

and for higher order autocorrelations

$$R_X(1) = \begin{bmatrix} 0.4826 & 0.2775 \\ 0.3513 & 0.4011 \end{bmatrix}$$

$$R_X(2) = \begin{bmatrix} 0.1300 & 0.2546 \\ 0.0894 & 0.3293 \end{bmatrix}$$

$$R_X(3) = \begin{bmatrix} 0.0258 & 0.1667 \\ 0.0257 & 0.1933 \end{bmatrix}$$

$$\vdots$$

1 The code below replicates these results with Matlab.

```
% VAR(2) stability and correlations
clear all;clc;
% set up parameters
p = 2; % lag order
k = 2; % number of variables

A1 = [0.5 0.1 ;...
      0.4 0.3];

A2 = [-0.2 0.1 ;...
      -0.3 0.2];

c  = [0.2;0.3];
```

```
14
15 sig = [1.75 0.25 ;...
16       0.25 3.00];
17
18 % set up companion form parameters
19 Ik = eye(k);
20 Ok = zeros(k,k);
21
22 A  = [A1 A2 ;...
23       Ik Ok];
24
25 C  = [c;Ok(:,1)];
26
27 Sig = [sig Ok;...
28       Ok Ok];
29
30 % lagploynomial (I - A1L - A2L) using symbolic math toolbox
31 syms z
32 AL  = Ik - A1*z - A2*z^2; % symbolic polynomial
33 detAL = det(AL);       % det of symbolic polynomial
34 %%
35 fprintf('Lag polynomial expression when evaluating det(AL) \n');
36 fprintf(' %s \n\n', char(detAL));
37
38 % convert back to numerical polynomial
39 lag_roots = roots(sym2poly(detAL));
40 fprintf('Roots of Lag-polynomial:\n');
41 disp(lag_roots);
42
43 % compute eig(A)
44 eigA = eig(A);
45 fprintf('\nRoots of charactistic-polynomial:\n');
46 disp(eigA);
47
48 %% unconditional mean
49 mu  = inv(Ik - A1 - A2)*c;
50
51 % or from companion from
52 S   = [Ik Ok];      % selection vector
53 Mu  = inv(eye(p*k)-A)*C;
54 muC  = S*Mu;
55
56 %% variance/covariance matrix
57 Gam0 = reshape(inv(eye(16)-kron(A,A))*Sig(:),4,4);
58 gam0 = Gam0(1:2,1:2);
59 gam1 = Gam0(1:2,3:4);
60
61 % computing the autocovariance and autocorrelation recursions for
62 J   = 10;       % highest number of ACVs and ACFs to compute
63 % ACVFs
64 Gams = zeros(k,k,J);
65 Gams(:,:,1) = gam0;
```

```
Gams(:,:,2) = gam1;
% ACFs
d   = sqrt(diag(diag(gam0)));
Rhos = zeros(k,k,J);
Rhos(:,:,1) = inv(d)*gam0*inv(d);
Rhos(:,:,2) = inv(d)*gam1*inv(d);

for j = (p+1):J
  Gams(:,:,j) = A1*Gams(:,:,j-1) + A2*Gams(:,:,j-2);
  Rhos(:,:,j) = inv(d)*Gams(:,:,j)*inv(d);
end;

fprintf('\n---------------------------------------\n');
fprintf('   ACVs    ACFs \n');
fprintf('---------------------------------------\n');
disp([Gams Rhos])
fprintf('---------------------------------------\n');
```

Matlab Code 1: `example_VAR2_stability_forecasting.m`

3.8. Forecasting with VAR models

Forecasts from VARs are easily constructed. Conceptually, we follow the same strategy as for AR(p) models. For a first order VAR in de-meaned form, that is:

$$(X_t - \mu) = A_1 (X_{t-1} - \mu) + U_t \tag{43}$$

where $E(X_t) = \mu = (I_k - A_1)^{-1}C$, we forecast $h-$periods ahead using the relation

$$X_{t+h|t} = \mu + A_1^h (X_t - \mu). \tag{44}$$

If we have a more general VAR(p), we can again use the companion form

$$(\mathbf{X}_t - \boldsymbol{\mu}) = \mathbf{A} (\mathbf{X_{t-1}} - \boldsymbol{\mu}) + \mathbf{U}_t$$

to construct the forecasts as:

$$\mathbf{X}_{t+h|t} = \boldsymbol{\mu} + \mathbf{A}^h (\mathbf{X}_t - \boldsymbol{\mu}) \tag{45}$$

for the $(kp \times 1)$ vector $\mathbf{X}_{t+h}$. To extract the forecast for the X_{t+h} vector of interest, we need to compute:

$$X_{t+h|t} = \mu + \mathbf{S}\mathbf{A}^h (\mathbf{X}_t - \boldsymbol{\mu}) \tag{46}$$

where $\mathbf{S}$ is again the $(k \times kp)$ dimensional selection matrix defined in (23) and $(\mathbf{X}_t - \boldsymbol{\mu})$ is the de-meaned (stacked) series of X_t, that is,

$$(\underset{(pk\times 1)}{\mathbf{X}_t - \boldsymbol{\mu}}) = \left(\begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix} - \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} \right), \tag{47}$$

where μ is again the unconditional mean of X_t as defined in (25).

The forecast errors, defined as $\mathcal{E}_{t+h} = X_{t+h} - X_{t+h|t}$ also follow an (vector) MA$(h-1)$ structure, as was the case for the the univariate AR(p) model forecasts. The forecast errors are computed as:

$$\mathcal{E}_{t+h} = \sum_{i=0}^{h-1} \Psi_i U_{t+h-i} \tag{48}$$

where Ψ_i are the coefficients of the VMA(∞) representation of the VAR(p) defined in (21) with $\Psi_0 = I_k$. The $h-$step ahead forecast error variance is calculated as

$$\Sigma_U(h) = \sum_{i=0}^{h-1} \Psi_i \Sigma \Psi_i'. \tag{49}$$

We can see again, that as $h \to \infty$, the $h-$step ahead forecast error variance $\Sigma_U(h)$ will go to the unconditional variance of X_t, ie., to $\Gamma_X(0)$ from as given in (31).

The Matlab code below shows how to implement these forecasts.

```
%% forecasting
H  = 20;       % upper forecast horizon cap.
Xt = [1 2 3 4]';   % (pkx1) vector of Observed X_t at time t to use a base to forecast
    from.
% constructing the Psi weights
Psi = zeros(k,k,H);
% create psi function
psi = @(A,h) (S*A^h*S');

% looping through different time horizons
for jj = 1:H
  Psi(:,:,jj) = psi(A,jj-1);
end;

for h = 1:H;
  Xhat_h(:,h)  = S*(Mu + A^h*Xt);      % h-step ahead forecasts
  SigU_h(:,:,h) = Psi(:,:,h)*sig*Psi(:,:,h)'; % h-step ahead Sig weiths that need to be
```

```
      summed.
end

% plot the forecasts
hold on;
plot(1:H,Xhat_h(1,:),'b');
plot(1:H,Xhat_h(2,:),'r');
hline(mu(1),':b');
hline(mu(2),':r');
hold off;

% forecast error variance converges to gam0
SigU_H = sum(SigU_h,3);
disp(' Forecast error variance matrix');
disp(SigU_H)
disp(' Gamma(0) covariance matrix');
disp(gam0)
```

Matlab Code 2: `example_VAR2_stability_forecasting.m`

3.9. Exogeneity and Model reduction

3.9.1. Exploiting Exogeneity

Although the initial argument of Sims (1980) and the reason for using VAR models in the first place was that one was unhappy with the *'incredible exclusion restrictions'* of the structural equation models (SEMs) of that time,[2] one can think of many cases where it does not seem such a good idea to leave the model unrestricted. One of these occurs when operating with data on a small open economy or perhaps modelling a small sector of an economy. In these cases it often makes a lot of sense to impose some zero restrictions upon the A_j. These are frequently referred to as *'block-exogeneity'* restrictions.

A simple example would be when modelling the Australian or Swiss (or of any other small open economy) and the US economy, where the US economy is often used as a proxy for the world economy. It would seem sensible to argue that movements in Australian/Swiss GDP do not influence US GDP so that lagged values of Australian/Swiss GDP would have zero coefficients attached to them in the equation that has US GDP on the left-hand side (LHS). Similar considerations would apply to domestic and overseas interest rates and equity prices. It may also be that one feels that the institutional details (or information) are such as to indicate that some lags should be zero. Such *'inertial restrictions'* have been popular in recent work in connection with the impact of monetary

[2] that is, one should not impose any restrictions upon lagged values of X_t, so that there are no zeros in the A_j matrix

policy in the sense that it is assumed that interest rates will have no effect on output for at least two quarters.

3.9.2. Model Reduction

It seems unlikely that one can fully capture a complex entity like an economy with a small set of variables in X_t. What happens if the data generating process (DGP) is a VAR in $X_t = [x_{1t}' \; x_{2t}']'$, where x_{1t} and x_{2t} are of of dimension $(k_1 \times 1)$ and $(k_2 \times 1)$ respectively, but the variables worked with, x_{1t}, are a sub-set of X_t?[3] The classic treatments of such a set up are discussed in detail in Zellner and Palm (1974) and Wallis (1977).

We can gain an appreciation of what such model reduction produces by using again a simple bi-variate VAR(1) model taking the form

$$x_{1t} = a_{11}^1 x_{1t-1} + a_{12}^1 x_{2t-1} + u_{1t}$$
$$x_{2t} = a_{21}^1 x_{1t-1} + a_{22}^1 x_{2t-1} + u_{2t}$$

or using lag operators

$$\begin{bmatrix} a_{11}(L) & \tilde{a}_{12}(L) \\ \tilde{a}_{21}(L) & a_{22}(L) \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \tag{50}$$

where $a_{jj}(L) = 1 - a_{jj}L$ is the standard lag polynomial and $\tilde{a}_{ij}(L) = -a_{ij}L, \forall i \neq j$, is a lag polynomial which is restricted to have the first element set to 0 rather than one.

Let us think of the simple scenario where $a_{21} = 0$ in (50) so that x_{2t} is exogenous to this system and ask what would happen if we reduced the two variable system to a single variable x_{1t}. Then the relation in (50) becomes:

$$\begin{bmatrix} (1 - a_{11}L) & -a_{12}L \\ 0 & (1 - a_{22}L) \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{51}$$

Expanding (51) yields:

$$(1 - a_{11}L)x_{1t} = a_{12}Lx_{2t} + u_{1t} \tag{52a}$$
$$(1 - a_{22}L)x_{2t} = u_{2t}. \tag{52b}$$

[3] In the simple bi-variate VAR(1) we have x_{1t} and x_{2t} as scalars but this clearly generalises to the case where x_{1t} and x_{2t} are vectors themselves.

Substituting $x_{2t} = (1 - a_{22}L)^{-1}u_{2t}$ into (52a) yields:

$$(1 - a_{11}L)x_{1t} = a_{12}\frac{Lu_{2t}}{(1 - a_{22}L)} + u_{1t} \tag{53}$$

$$(1 - a_{22}L)(1 - a_{11}L)x_{1t} = a_{12}u_{2t-1} + (1 - a_{22}L)u_{1t} \tag{54}$$

so that the process for x_{1t} is in fact an ARMA(2,1).

Note that this is a **general result**. **Reducing the number of variables** in a system that is a VAR in X_t **becomes a VARMA process in the sub-set** of X_t, x_{1t}. If one then tries to fit the process for x_{1t} with a VAR it may have to be of very high order. The **problem becomes worse the greater the reduction** in variables being performed, so it is sensible to keep a reasonable number of variables in the x_{1t} set and not to think one can get away with a small number like two or three.

Mostly people tend to resist making the number of variables in x_{1t} large as they say that this means too many right-hand side (RHS) variables in each equation, since, if there are k_1 variables in x_{1t}, there are $p \times k_1$ lagged values on the RHS of each equation in the VAR(p). This may seem justified as one wants to avoid to model too large a VAR system due to the large proliferation of number of parameters to be estimated. But the problem with this way of thinking is that we are keeping the lag order p fixed as we increases the number of variables included in the VAR. One may have less variables on the RHS as k_1 grows as one can see from the example above. There were only three coefficients in the VAR(1) model (a_{11}, a_{12}, a_{22}, because a_{21} is zero by assumption) in the simple two variable scenario, but there are also at least three in the equation with x_{1t}. In fact, there can be more in the latter case since one has to approximate an infinite order VAR with a large p VAR due to the fact that x_{1t} has a VARMA structure and people generally tend to prefer to approximate the VARMA model with a large VAR model as they are easier to estimate and interpret.

Reducing a system of k equations to a single variable equation is generally referred to as finding the *'final equations form'* of the system. These are the univariate equation forms of the k variables in a general VAR(p). It is instructive here to show how one can find these (at least conceptually). Let us use the general k variable VAR(p) denoted by $A(L)X_t = U_t$ to illustrate this. Then X_t can be inverted or solved to its VMA(∞) form to yield:

$$X_t = A(L)^{-1}U_t. \tag{55}$$

Now the inverse of a matrix can be written as

$$A(L)^{-1} = \frac{\mathsf{adj}[A(L)]}{\mathsf{det}[A(L)]}$$

where $\mathsf{adj}(A)$ is the adjoint matrix of A and $\mathsf{det}(A)$ is the determinant, so that $A(L)^{-1}$ can be written as

$$\begin{aligned} X_t &= \frac{\mathsf{adj}[A(L)]}{\mathsf{det}[A(L)]} U_t \quad (56) \\ \mathsf{det}[A(L)]X_t &= \mathsf{adj}[A(L)]U_t. \quad (57) \end{aligned}$$

In terms of our example earlier, we had

$$A(L) = \begin{bmatrix} 1 - a_{11}L & -a_{12}L \\ 0 & 1 - a_{22}L \end{bmatrix} \quad (58)$$

so that

$$\begin{aligned} \mathsf{adj}[A(L)] &= \begin{bmatrix} 1 - a_{22}L & a_{12}L \\ 0 & 1 - a_{11}L \end{bmatrix} \quad (59) \\ \mathsf{adj}[A(L)] &= (1 - a_{22}L)(1 - a_{11}L) \quad (60) \end{aligned}$$

and therefore

$$\underset{(1\times1)}{(1 - a_{22}L)(1 - a_{11}L)} \underset{(1\times2)}{\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}} = \underset{(2\times2)}{\begin{bmatrix} 1 - a_{22}L & a_{12}L \\ 0 & 1 - a_{11}L \end{bmatrix}} \underset{(2\times1)}{\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}} \quad (61)$$

$$\begin{aligned} (1 - a_{22}L)(1 - a_{11}L)x_{1t} &= (1 - a_{22}L)u_{1t} + a_{12}Lu_{2t} \\ (1 - a_{22}L)(1 - a_{11}L)x_{2t} &= (1 - a_{11}L)u_{2t} \end{aligned}$$

which then reduce to

$$\begin{aligned} (1 - a_{22}L)(1 - a_{11}L)x_{1t} &= (1 - a_{22}L)u_{1t} + a_{12}Lu_{2t} \\ (1 - a_{22}L)x_{2t} &= u_{2t}. \end{aligned}$$

This is the same outcome as we established directly above.

It's clear then that the AR part of the final equations for all elements of X_t are identical

and equal to $\det[A(L)]$, as is evident from the LHS relation in (61). So the differences between the nature of the univariate processes will reside in the MA part that comes from $\text{adj}[A(L)]U_t$.

3.10. Estimation of VARs

VARs are nothing else but a system of equations where on the left hand side we have the current values of X_t and on the right hand side we have lagged values of X_t, plus possibly some other exogenous variables (say Z_t) that are not determined within the system (ie., we do not have an equation written down for it).

Recall from your Econometrics II class, that the Seemingly Unrelated Regression (SUR) estimator of Zellner (1962) applies to the following system of equation model

$$y_{1i} = x_{i1}\beta_1 + u_{1i} \tag{62}$$

$$y_{2i} = x_{i2}\beta_2 + u_{2i} \tag{63}$$

which in stacked form is

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} = \begin{bmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} u_{1i} \\ u_{2i} \end{bmatrix}$$

$$Y_i = X_i\beta + U_i \tag{64}$$

where x_{ij} is a $(1 \times g_j)$ vector of *"regressors"* (including the intercept) for y_{ij}, $\forall j = 1, 2$, and $i = 1, \ldots, N$, β_j is a $(g_j \times 1)$ parameter vector of interest, and u_{ij} is a noise term with $\text{Var}(U_i) = \Sigma$. Notice here that x_{i1} and x_{i2} do not need to have the same regressors for each equation, so this specification is rather general. The dimension of X_i is thus $(2 \times G)$, where $G = (g_1 + g_2)$ and $\beta = [\beta_1' \ \beta_2']'$ with dimension $(G \times 1)$.

The SUR estimator of Zellner (1962) is a Generalised Least Squares (GLS) estimator which is the most efficient estimator when a general structure for the system in (64) is assumed, so that no restrictions on X_i or Ω are placed. As long as each element of X_i is uncorrelated with each element of U_i so that $E(X_i \otimes U_i) = 0$ holds, we can get efficient (GLS) estimates of β as

$$\hat{\beta}_{\text{GLS}} = \left(\sum_{i=1}^{N} X_i' \Sigma^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \Sigma^{-1} Y_i \right). \tag{65}$$

Two results from the SUR model are as follows. First, if Σ is diagonal, then there exists no efficiency gain from using (65) relative to OLS which would just obtain estimates of β as:

$$\hat{\beta}_{\mathsf{OLS}} = \left(\sum_{i=1}^{N} X_i' X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' Y_i \right). \tag{66}$$

This means that $\hat{\beta}_{\mathsf{GLS}} \equiv \hat{\beta}_{\mathsf{OLS}}$ if Σ is diagonal. Second, if $x_{i1} \equiv x_{i2}$, that is, the *'regressors'* are the same in each equation, then $\hat{\beta}_{\mathsf{GLS}} \equiv \hat{\beta}_{\mathsf{OLS}}$.

Now, since

$$X_i = \begin{bmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{bmatrix} \text{ and } Y_i = \begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix}$$

we have

$$\begin{aligned} X_i' X_i &= \begin{bmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{bmatrix}' \begin{bmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{bmatrix} \\ &= \begin{bmatrix} x_{i1}' & 0 \\ 0 & x_{i2}' \end{bmatrix} \begin{bmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{bmatrix} \\ &= \begin{bmatrix} x_{i1}' x_{i1} & 0 \\ 0 & x_{i2}' x_{i2} \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} X_i' Y_i &= \begin{bmatrix} x_{i1} & 0 \\ 0 & x_{i2} \end{bmatrix}' \begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \\ &= \begin{bmatrix} x_{i1}' y_{1i} \\ x_{i2}' y_{2i} \end{bmatrix} \end{aligned}$$

so that

$$\begin{aligned} \hat{\beta}_{\mathsf{OLS}} &= \left(\sum_{i=1}^{N} X_i' X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' Y_i \right) \\ &= \begin{bmatrix} \sum_{i=1}^{N} x_{i1}' x_{i1} & 0 \\ 0 & \sum_{i=1}^{N} x_{i2}' x_{i2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} x_{i1}' y_{1i} \\ \sum_{i=1}^{N} x_{i2}' y_{2i} \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \left[\sum_{i=1}^{N} x_{i1}' x_{i1}\right]^{-1} & 0 \\ 0 & \left[\sum_{i=1}^{N} x_{i2}' x_{i2}\right]^{-1} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} x_{i1}' y_{1i} \\ \sum_{i=1}^{N} x_{i2}' y_{2i} \end{bmatrix}$$

$$= \begin{bmatrix} \left[\sum_{i=1}^{N} x_{i1}' x_{i1}\right]^{-1} \sum_{i=1}^{N} x_{i1}' y_{1i} \\ \left[\sum_{i=1}^{N} x_{i2}' x_{i2}\right]^{-1} \sum_{i=1}^{N} x_{i2}' y_{2i} \end{bmatrix}. \tag{67}$$

The relation in (67) can be seen to imply that the OLS estimator

$$\hat{\beta}_j = \left[\sum_{i=1}^{N} x_{ij}' x_{ij}\right]^{-1} \sum_{i=1}^{N} x_{ij}' y_{1j}$$

for $j = 1, 2$, ie., the two regression equations of interest. Thus, we can get estimates of $\hat{\beta}_{\text{OLS}} = \left[\hat{\beta}_1' \; \hat{\beta}_2'\right]'$ by running two individual OLS regressions on the equations given in (62) and (63).

We can translate the SUR model now to the case of a VAR(p), where, for the moment, suppose that $p = 2$. This yields

$$x_{1t} = c_1 + a_{11}^1 x_{1t-1} + a_{12}^1 x_{2t-1} + a_{11}^2 x_{1t-2} + a_{12}^2 x_{2t-2} + u_{1t} \tag{68a}$$

$$x_{2t} = c_2 + a_{21}^1 x_{1t-1} + a_{22}^1 x_{2t-1} + a_{21}^2 x_{1t-2} + a_{22}^2 x_{2t-2} + u_{2t} \tag{68b}$$

which can be written as

$$X_t = C + A_1 X_{t-1} + A_2 X_{t-2} + U_t \tag{69}$$

with

$$X_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}, \; C = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \; A_j = \begin{bmatrix} a_{11}^j & a_{12}^j \\ a_{21}^j & a_{22}^j \end{bmatrix}, \text{ and } U_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

and $\text{Var}(U_t) = \Sigma$.

We have seen earlier that the VAR(p) model in (69) can be written in companion form as:

$$\mathbf{X}_t = \mathbf{C} + \mathbf{A}\mathbf{X}_{t-1} + \mathbf{U}_t \tag{70}$$

where

$$\mathbf{X}_t = \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix}, \; \mathbf{C} = \begin{bmatrix} C \\ 0_k \end{bmatrix}, \; \mathbf{A} = \begin{bmatrix} A_1 & A_2 \\ I_k & 0_k \end{bmatrix}, \text{ and } \mathbf{U}_t = \begin{bmatrix} U_t \\ 0_k \end{bmatrix}.$$

The companion form is convenient to use when performing ARMA algebra on the VAR(p)

model and when forecasting.

The model in (69) can also be put in the following '*regression form*' (corresponding to (64)), that is,

$$X_t = \mathcal{A}' \mathcal{X}_{t-1} + U_t \tag{71}$$

where

$$\underset{(2\times 5)}{\mathcal{A}'} = [C \; A_1 \; A_2] \text{ and } \underset{(5\times 1)}{\mathcal{X}_{t-1}} = \left[1 \; X'_{t-1} \; X'_{t-2}\right]'$$

so that OLS estimates of $\mathcal{A}'$ can be obtained from the relation (see page 293 in Hamilton (1994))

$$\widehat{\mathcal{A}'}_{\text{OLS}} = \left(\sum_{t=1}^{T} \underset{(2\times 1)(1\times 5)}{\mathcal{X}_t \mathcal{X}'_{t-1}} \right) \left(\sum_{t=1}^{T} \underset{(5\times 1)(1\times 5)}{\mathcal{X}_{t-1} \mathcal{X}'_{t-1}} \right)^{-1}$$

where $\mathcal{A}'$ is of dimension $(2 \times (1 + 2k))$ with $k = 2$, the number of variables in the VAR(p). These results generalise to any k and any p.

To see the relation to ML estimation, we again need to make an assumption about the distribution of the U_t vector of residuals. If this distribution is multivariate normal, the likelihood function for the model in (71) is:

$$\begin{aligned} \mathcal{L}(\cdot) = & -(Tk/2)\log(2\pi) + (T/2)\log|\Sigma^{-1}| \\ & - \frac{1}{2}\sum_{t=1}^{T} \left(\mathcal{X}_t - \mathcal{A}'\mathcal{X}_{t-1}\right)' \Sigma^{-1} \left(\mathcal{X}_t - \mathcal{A}'\mathcal{X}_{t-1}\right) \end{aligned}$$

and we can see again that it is proportional to the negative of the OLS criterion function. Thus, ML estimates of the parameters of the VAR can simply be obtained by OLS which will have the same efficiency bounds as MLE with normal errors (see Hamilton, 1994, pages 293 onwards). From Proposition 11.1 on page 298 in Hamilton (1994) we know that these will be distributed asymptotically normal, just as our AR(p) regression estimator. Once estimates of the elements of A_j are available, we can compute $\hat{\Sigma}$ as $T^{-1}\sum_{t=1}^{T} \hat{U}_t \hat{U}'_t$, where $\hat{U}_t$ denotes the fitted residuals from the "*regressions*" in (71). This is the MLE estimate of Σ and we can get an unbiased OLS one by deflating by $T - (1 + pk)$ rather than by T.

In summary, if a plain vanilla VAR is of interest, we can use OLS equation by equation to get estimates of the model parameters. If restrictions on the lagged values on the right hand side are placed, then a GLS type estimator can be used to get "*more efficient*" estimates than OLS. The potential cost of this efficiency gain is nonetheless, that, if there

is any misspecification in any of the equations of the model, then all VAR parameters will be affected by this, and not only in those equations where the misspecification originated from. This is a (well known) drawback of using GLS in this context.

3.11. Lag and Model Selection

3.11.1. Selecting the Lag order

The lag order in the VAR(p) can, in principle, be selected with the same information criteria that were used for the univariate models discussed earlier. The multivariate version of the AIC, BIC and HQ criteria are computed as:

$$\begin{aligned} \text{AIC} &: \ln\left|\hat{\Sigma}\right| + 2k^2 p/T \\ \text{BIC} &: \ln\left|\hat{\Sigma}\right| + k^2 p \ln(T)/T \\ \text{HQ} &: \ln\left|\hat{\Sigma}\right| + k^2 p 2\ln[\ln(T)]/T \end{aligned} \tag{72}$$

where k is the number of variables in the VAR(p), p is the lag order and $\hat{\Sigma} = T^{-1}\sum_{t=1}^{T} U_t U_t'$ is the estimate of the variance covariance matrix of the residual vector U_t. The lag length p in the VAR should be chosen to minimise these criteria. Such methods are likely to be most reliable if the order of the VAR is low since making p very large can make it very hard to estimate any of the parameters with much accuracy.

Ivanov and Kilian (2005) recommend to use the AIC for monthly data and the HQ for quarterly data for commonly encountered sample sizes of more than 120 observations. If less than that are available, the BIC criterion is recommended. Their results are based on the "*accuracy*" of the IRFs so if the VAR model is used for policy analysis, than these recommendations appear to be useful. If one wants to use a VAR as a forecasting model only, then it is not clear whether these recommendations still apply.

3.11.2. Variable selection

A much less studied issue in VAR analysis is the **choice of variables** that are to appear in the VAR. So this is a **variable** rather then **lag** selection problem and as we saw earlier in the Model Reduction section, this choice is potentially a serious source of problems in VARs. Quite a bit of thought needs to be given to decide on which variables should appear in the system. Generally one would be ill-advised to reduce the number of variables in a VAR too quickly and one should try to use economic theory as much as possible in the selection of the variables that enter the VAR. VARs should never really be constructed without putting a good amount of thought into the (non-existing) model (and thus vari-

ables) that one is fitting. Although VARs can be used in a very powerful way due to the fact that we can think about many macro-economic models fundamentally having a VAR reduced form structure, one needs to think carefully about the economics and not be tempted to just *'press buttons'* in some software environment.

It should be emphasised here that the problem of variable selection and lag order selection is really one **joint selection problem**. That is, we should always think about selecting the number of variables and the lag order together, rather than trying to figure out the lag order after the number of variables have been fixed. This should be clear from the discussion in the Model Reduction Section (Section 3.9.2). Since we know that variable reduction in a VAR(p) means we will obtain a VARMA process, and a VARMA process can (or will) be approximated arbitrarily well by a high order VAR process with the order of this process being substantially larger than the original VAR(p), will need to think carefully about k as well as p. In general, we can think about this in terms of the following two implications:

1. A small number of variables (small k) will likely mean that a high order VAR(p) will be needed.
2. If we need a high order VAR(p) model to fit the data, then this is likely an indication that we need more variables (higher k) to more accurately describe the economic relation that is being modelled.

4. Causality in VARs

We have discussed some issues with causality already at the outline of this lecture. Nonetheless, since one of the earliest uses of VARs was to look at causality, we will go through some of the representations below to get a feel for the relations that are being looked at. One of the most popular ways to look at causality is Granger temporal causality. To simplify notation a bit, let us use x and y to denote a bi-variate system. Causality is then defined within a predictive power argument as follows.

Definition 1 (Granger Causality): *Let MSE be the Mean Squared Forecast Error. Then, if x causes y, x has predictive power over y such that*

$$MSE\left(y_{t+h}|\Omega_t\right) < MSE\left(y_{t+h}|\Omega_t - X^t\right) \tag{73}$$

*where $X^t = \{x_i\}_1^t$ and Ω_t is the (full) information set available at time t to forecast y. x is then said to **Granger Cause** y.*

Causality can be determined through instantaneous influences (including current time t values) or influences that are adjusted to over different time periods, ie., with lags. When we think of instantaneous causality, we would include the current value of x in the information set. In reduced form VARs, we will generally look at the off-diagonal entries of the covariance matrix of residuals to determine instantaneous Causality. Let us now look at various transformations of the relations between the variables of interest in the simple bivariate case. We will again use lag operator notation, so that $a(L) = 1 - aL - aL^2 - \dots$ as before and also $\tilde{a}(L)$ to denote a restricted form of the lag operator, ie., $\tilde{a}(L) = -aL - aL^2 - \dots$.

Consider the following reduced form and normalised VAR(1)

$$\begin{aligned} y_t &= a_{11}y_{t-1} + a_{12}x_{t-1} + u_t \\ x_t &= a_{21}y_{t-1} + a_{22}x_{t-1} + v_t \end{aligned}$$

which, in lag operator form, is

$$\underbrace{\begin{bmatrix} a_{11}(L) & \tilde{a}_{12}(L) \\ \tilde{a}_{21}(L) & a_{22}(L) \end{bmatrix}}_{A(L)} \underbrace{\begin{bmatrix} y_t \\ x_t \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} u_t \\ v_t \end{bmatrix}}_{W_t} \tag{74}$$

where

$$W_t \begin{bmatrix} u_t \\ v_t \end{bmatrix}$$

with covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}.$$

Evidently, here $a_{11}(L) = 1 - a_{11}L$ and $\tilde{a}_{12}(L) = -a_{12}L$ for the simple case considered here, but it will be fruitful to leave the set up as general as possible so that we can see how this could generalise to any lag order model

If we make the assumption that Σ is diagonal for the moment, so that $\sigma_{21} = \sigma_{12} = 0$ and hence u_t and v_t are uncorrelated, then non-instantaneous Causality fundamentally looks at the off-diagonal terms of the $A(L)$ matrix in (74) by testing if the $\tilde{a}_{21}(L)$ and $\tilde{a}_{12}(L)$ coefficients of the $A(L)$ lag polynomial matrix are statistically different from zero. The 4 possible outcomes we can arrive at in terms of Causality are:

i) if $\tilde{a}_{12}(L) = 0$ (all else the same), then x_t does not Granger Cause y_t.

ii) if $\tilde{a}_{21}(L) = 0$ (all else the same), then y_t does not Granger Cause x_t.

iii) if $\tilde{a}_{21}(L) = \tilde{a}_{12}(L) = 0$ (all else the same), then no Granger Causality exist.

iv) if $\tilde{a}_{21}(L) \wedge \tilde{a}_{12}(L) \neq 0$ (all else the same), then bi-directional Granger Causality exist.

If one wants to make an allowance for the possibility of instantaneous Causality, one would need to look at the correlation between u_t and v_t and hence essentially test if σ_{21} (or σ_{12}) is different from 0 in Σ. Note here that this will only determine whether there does or does not exist instantaneous Causality, but one will not be able to determine which way it is running. For this to be determined, more economic structure will need to be put on the model and this is thus very similar then to the scenario of how best to decompose Σ to get the IRFs. Let us thus abstract for now from instantaneous Causality and focus on different ways of looking at Causality through the lag polynomial $A(L)$.

4.1. Alternative representations of Causality

4.1.1. VMA representation

We can invert the VAR form in (74) to the following VMA(∞) form

$$\underbrace{\begin{bmatrix} y_t \\ x_t \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} b_{11}(L) & \tilde{b}_{12}(L) \\ \tilde{b}_{21}(L) & b_{22}(L) \end{bmatrix}}_{B(L)} \underbrace{\begin{bmatrix} u_t \\ v_t \end{bmatrix}}_{W_t} \tag{75}$$

where we have $B(L) = A(L)^{-1}$ and inverting $A(L)$ yields:

$$A(L)^{-1} = \det[A(L)]^{-1} \begin{bmatrix} a_{22}(L) & -\tilde{a}_{12}(L) \\ -\tilde{a}_{21}(L) & a_{11}(L) \end{bmatrix} \tag{76}$$

and hence

$$\underbrace{\begin{bmatrix} y_t \\ x_t \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} \frac{a_{22}(L)}{\det(A)} & -\frac{\tilde{a}_{12}(L)}{\det(A)} \\ -\frac{\tilde{a}_{21}(L)}{\det(A)} & \frac{a_{11}(L)}{\det(A)} \end{bmatrix}}_{B(L)} \underbrace{\begin{bmatrix} u_t \\ v_t \end{bmatrix}}_{W_t} \tag{77}$$

With Σ diagonal as assumed above, Causality in the VMA(∞) representation is also determined by the off-diagonal terms in the $B(L)$ lag polynomial matrix, so that x_t causes y_t through lags of v_t if $\tilde{b}_{12}(L) \neq 0$ (all else constant) and y_t causes x_t through lags of u_t if $\tilde{b}_{21}(L) \neq 0$ (all else constant). If $\tilde{b}_{12}(L) = \tilde{b}_{21}(L) = 0$, then no flow of information from

past u_t or v_t to y_t and x_t takes place so that both variables are determined independently from one another. We can easily see from the VAR and VMA representation that the causal flow is determined by $\tilde{b}_{12}(L) = -\frac{\tilde{a}_{12}(L)}{\det(A)}$ and $\tilde{b}_{21}(L) = -\frac{\tilde{a}_{21}(L)}{\det(A)}$, so we have merely expressed the relations in an alternative form.

4.1.2. Univariate ARMA representations

Yet another way to look at Causality is to formulate univariate ARMA representations for x_t and y_t. To illustrate this, let $Z_t = B(L)W_t$ be the VMA representation from (75). Now define $\Psi(L)$ as:

$$\Psi(L) = \begin{bmatrix} \psi_{11}(L) & 0 \\ 0 & \psi_{22}(L) \end{bmatrix}.$$

Then

$$\begin{aligned} Z_t &= \overbrace{\Psi(L)\underbrace{\Psi(L)^{-1}B(L)}_{H(L)}}^{I_2} W_t \\ &= \Psi(L)\underbrace{H(L)W_t}_{G_t} \\ &= \Psi(L)\, G_t. \end{aligned} \tag{78}$$

The term $H(L)$ in (78) is computed as

$$\begin{aligned} H(L) &= \Psi(L)^{-1} B(L) \\ \begin{bmatrix} \eta_{11}(L) & \tilde{\eta}_{12}(L) \\ \tilde{\eta}_{21}(L) & \eta_{22}(L) \end{bmatrix} &= \begin{bmatrix} \psi_{11}(L)^{-1} & 0 \\ 0 & \psi_{22}(L)^{-1} \end{bmatrix} \begin{bmatrix} b_{11}(L) & \tilde{b}_{12}(L) \\ \tilde{b}_{21}(L) & b_{22}(L) \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} \eta_{11}(L) &= \psi_{11}(L)^{-1}b_{11}(L), \quad \tilde{\eta}_{12}(L) = \psi_{11}(L)^{-1}\tilde{b}_{12}(L) \\ \tilde{\eta}_{21}(L) &= \psi_{22}(L)^{-1}\tilde{b}_{21}(L), \quad \eta_{22}(L) = \psi_{22}(L)^{-1}b_{22}(L) \end{aligned}$$

so that we obtain

$$\underbrace{\begin{bmatrix} y_t \\ x_t \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} \psi_{11}(L) & 0 \\ 0 & \psi_{22}(L) \end{bmatrix}}_{\Psi(L)} \underbrace{\begin{bmatrix} g_{1,t} \\ g_{2,t} \end{bmatrix}}_{G_t}. \tag{79}$$

Note here that $G = [g_{1,t} \;\; g_{2,t}]'$ is a transformed residual vector with ARMA components

in it.

Because of the block diagonal relation in (79), y_t and x_t have two individual univariate MA representations of the form

$$y_t = \psi_{11}(L)g_{1,t} \tag{80}$$

$$x_t = \psi_{22}(L)g_{2,1} \tag{81}$$

where y_t and x_t are each an MA(∞) model. We can then proceed to approximate the univariate MA(∞) models by finite order ARMA(p, q) models as done before. Notice here that we ignore the bi-variate nature of Z_t, and treat them as individual series. We could then test for correlation between $g_{1,t}$ and $g_{2,t}$ at all leads and lags do determine Causality.

To see how these relations fit together, assume that we are interested in the case $\tilde{b}_{12}(L) = 0$ which thus implies that $a_{12}(L) = 0$. As before, the off-diagonal terms of G_t determine if Causality exists or not. In (80) we have

$$\begin{aligned} y_t &= \psi_{11}(L)g_{1,t} \\ &= \psi_{11}(L)\left[\eta_{11}(L)u_t + \tilde{\eta}_{12}(L)v_t\right]. \end{aligned}$$

If $\tilde{\eta}_{12}(L) = 0$, we get

$$\begin{aligned} y_t &= \psi_{11}(L)[\eta_{11}](L)u_t \\ &= \cancel{\psi_{11}(L)}[\cancel{\psi_{11}(L)^{-1}}b_{11}(L)](L)u_t \\ &= b_{11}(L)u_t \end{aligned}$$

so have the same MA(∞) representation for y_t as in (75) with $\tilde{b}_{12}(L) = 0$ is obtained which is mirrored in $\tilde{\eta}_{12}(L) = 0$.

4.2. Testing for Granger Causality

There are various forms of testing for Causality. We will list three fairly common ones here.

1. The VAR form takes
$$A(L)Z_t = W_t$$

   as the starting point and tests if $a_{12}(L)$ and $a_{21}(L)$ are equal to zero using standard

$F-$test procedures, ie.,

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(T-k)}, \quad \begin{matrix} m = \text{No. of restrictions} \\ k = \text{No. of parameters} \end{matrix}.$$

This is known as the *direct Granger Procedure*.

2. The Haugh-Pierce test procedure takes the individual ARMA froms of (80) and (81) and tests the cross-correlations between g_{1t} and $g_{2,t}$ for different leads and lags by constructing the test statistic S as:

$$S = T \sum_{k=k_1}^{k_2} \rho^2_{g_1 g_2}(k),\ k = k_1,...,k_2. \tag{82}$$

Under the null hypothesis that with g_{1t} and g_{2t} are uncorrelated, S is distributed as a Chi-squared random variable with $k_2 - k_1 + 1$ degrees of freedom.

3. The Hsiao procedure looks initially at individual regressions of y_t on its own lags and finds the optimal AR lag order p_y^0. It then includes x_t in the regression of y_t on its own p_y^0 lags and determines the optimal lag order for x_t, that is, p_x^0. One then proceeds by fixing p_x^0 and try to re-optimise the lag order for y_t. This procedure is repeated for x_t. Lastly, the whole system is then estimated by SUR and the individual $A(L)$ components are tested for significance in the SUR model.

Using the Haugh-Pierce test and Hsiao procedure to test for causality with more than two variables becomes messy. Many empirical papers with larger VARs focus on Granger form of Causality tests so fundamentally only ever estimate a reduced form VAR and then look at the off-diagonals of $A(L)$. Causality tests in general can be quite sensitive to the lag length specifications used and also to the number of variables included in the tests. So results are frequently found to be not very robust without any fundamentally important and decisive conclusions.

5. Structural VARs and VARs in policy analysis

As was outlined earlier, although VARs may be useful for summarising the dynamics of a set of variables, they are largely used to study how shocks that come into the economic system modelled by the VAR equations are absorbed and how they propagate through the economic system. So in VAR models, we are much less interested in the ACFs of the system *per se*, as was the case in AR models. Rather, we are more interested

in *'telling a story'* about what impact the different *'shocks'* have on the economic system (see the paper by Levtchenkova *et al.*, 1998 for an interesting read). This is commonly done through impulse response functions (IRFs) and decompositions of the forecast error variance (FVDs). Both of these are often displayed graphically and the focus is then on how large the reaction or impact of various policy variables of interest is at different time horizons, due to the *'artificially induced shock'* that came into the economic system.

5.1. Impulse response functions

Let us go back and consider the simple univariate AR(1): $x_t = \alpha x_{t-1} + u_t$. Let $x_0 = 0$ be the initial condition and let $u_1 = 3$ be the exogenous shock (or impulse) that comes into the economic model at time period $t = 1$ and which is 0 for all $t > 1$ (ie., $u_t = 0, \forall t = 2, 3, 4 \ldots$). Then we have the following evolution of the AR(1) over time:

$$\begin{aligned} x_1 &= \underbrace{\alpha x_0}_{0} + \underbrace{u_1}_{3} &&= 3 \\ x_2 &= \alpha x_1 + 0 &&= \alpha 3 \\ &\vdots &&\vdots \\ x_{t+1} &= \alpha x_t + 0 &&= \alpha^t 3 \end{aligned} \tag{83}$$

Thus, the impulse (or shock) response of x_t (due to $u_1 = 3$) for $t = 1, \ldots, t+1$ is determined by the initial size of the shock (ie., 3) and is transmitted through the system by the dynamics of the AR(1) process, that is, the α coefficient.

More formally, we can define the IRF as:

$$\frac{\partial x_{t+j}}{\partial u_t}, \forall j = 0, 1, 2, \ldots$$

that is, the change in the variable of interest x_{t+j} $(\forall j = 0, 1, 2, \ldots)$ periods into the future, due to an impulse of a given size in u_t (at time t). Let us now consider the same set up as before in (83), but here we extend it to a bi-variate VAR(1) model:

$$\underbrace{\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix}}_{A_1} \underbrace{\begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}}_{U_t}. \tag{84}$$

Giving again the same impulse $U_1 = \begin{bmatrix} 3 & 0 \end{bmatrix}'$ as before,[4] and assuming once more that the initial vector of X_t is $X_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}'$, our system evolves as

$$
\begin{aligned}
X_1 = A_1 X_0 + U_1 &= \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \\
X_2 = A_1 X_1 + U_2 &= \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.1 \\ 0.6 \end{bmatrix} \\
X_3 = A_1 X_2 + U_3 &= \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 1.1 \\ 0.6 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.54 \\ 0.54 \end{bmatrix} \\
\vdots \qquad & \qquad \vdots \qquad \qquad \vdots \\
X_{t+1} = A_1 X_t + U_{t+1} = & \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix}^t \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix}^t}_{A_1^t} \underbrace{\begin{bmatrix} 3 \\ 0 \end{bmatrix}}_{U_1}.
\end{aligned}
\tag{85}
$$

As is clear from the univariate and bi-variate examples above, to get the impulse responses, we iterated forward from a given initial condition X_0 (or x_0 for the univariate example) and then looked at how the shock that comes in at time period $t = 1$ propagates through the system. The coefficient vector A_1 controls the pass-through of the shock from period to period and U_1 determines the size of the impulse. Here we were able to explicitly keep the size and structure of U_1 fixed at $\begin{bmatrix} 3 & 0 \end{bmatrix}'$ because of the synthetic set-up. Evidently, this will not be possible in general and later on we will need to say something about the covariance matrix of U_t, as it controls the co-movement between the shocks u_{1t} and u_{2t} in the system. Depending on how we proceed, we will be able to choose the size of the shocks to be of a **one unit** or of a **one standard deviation** size. What this means will become clear in a couple of pages down.

The exposition of the IRF as outlined in the above few equations in (85) is a convenient way to look at and conceptually understand the propagation of shocks in a system, particulary if one has not seen such a treatment before. Nonetheless, it is not useful for computational purposes and it may not be clear what the link is to (V)AR and (V)MA models in general. Let us now then formalise the treatment of IRFs within the Moving Average representation of the IRFs that was used earlier in the AR(1) to MA(∞) representation in (4) and (5). The equivalent treatment in the vector from is to invert the VAR(p) to its VMA(∞) representations for the system and then look at how a shock at time t

[4] Now the impulse is the vector U_1 in the system VAR(1) model.

affects X_{t+j} for $j = 1, 2, 3, \ldots$. That is, we would form the VMA(∞) representation as:

$$X_{t+j} = \underbrace{A(1)^{-1}C}_{\mu} + \underbrace{A(L)^{-1}}_{\Psi(L)} U_{t+j}$$
$$= \mu + \Psi(L)U_{t+j}$$

and then look at the responses as captured by the partial derivatives

$$\frac{\partial x_{1t+j}}{\partial u_{1t}}, \frac{\partial x_{1t+j}}{\partial u_{2t}}, \frac{\partial x_{2t+j}}{\partial u_{1t}}, \frac{\partial x_{2t+j}}{\partial u_{2t}}, \forall j = 1, 2, 3, \ldots \tag{86}$$

The partial derivatives in (86) will be equal to the coefficient entries of the polynomial matrix $\Psi(L)$ at the different lag values. For example, let $j = 1$, then we get

$$X_{t+1} = \mu + \Psi_0 U_{t+1} + \Psi_1 U_t + \Psi_2 U_{t-1} + \ldots \tag{87}$$

$$\begin{bmatrix} x_{1t+1} \\ x_{2t+1} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \psi_{11}^0 & \psi_{12}^0 \\ \psi_{21}^0 & \psi_{22}^0 \end{bmatrix} \begin{bmatrix} u_{1t+1} \\ u_{2t+1} \end{bmatrix} + \begin{bmatrix} \psi_{11}^1 & \psi_{12}^1 \\ \psi_{21}^1 & \psi_{22}^1 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} + \ldots \tag{88}$$

and so

$$\frac{\partial x_{1t+1}}{\partial u_{1t}} = \psi_{11}^1, \frac{\partial x_{1t+1}}{\partial u_{2t}} = \psi_{12}^1,$$
$$\frac{\partial x_{2t+1}}{\partial u_{1t}} = \psi_{21}^1, \frac{\partial x_{2t+1}}{\partial u_{2t}} = \psi_{22}^1,$$

For $j = 2$, we get

$$X_{t+2} = \mu + \Psi_0 U_{t+2} + \Psi_1 U_{t+1} + \Psi_2 U_t + \Psi_3 U_{t-1} + \Psi_4 U_{t-2} + \ldots \tag{89}$$

$$\begin{bmatrix} x_{1t+2} \\ x_{2t+2} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \psi_{11}^0 & \psi_{12}^0 \\ \psi_{21}^0 & \psi_{22}^0 \end{bmatrix} \begin{bmatrix} u_{1t+2} \\ u_{2t+2} \end{bmatrix} + \begin{bmatrix} \psi_{11}^1 & \psi_{12}^1 \\ \psi_{21}^1 & \psi_{22}^1 \end{bmatrix} \begin{bmatrix} u_{1t+1} \\ u_{2t+1} \end{bmatrix} + \begin{bmatrix} \psi_{11}^2 & \psi_{12}^2 \\ \psi_{21}^2 & \psi_{22}^2 \end{bmatrix} \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} + \ldots \tag{90}$$

and so the IRFs become

$$\frac{\partial x_{1t+2}}{\partial u_{1t}} = \psi_{11}^2, \frac{\partial x_{1t+2}}{\partial u_{2t}} = \psi_{12}^2,$$
$$\frac{\partial x_{2t+2}}{\partial u_{1t}} = \psi_{21}^2, \frac{\partial x_{2t+2}}{\partial u_{2t}} = \psi_{22}^2,$$

We can see how this generalises to any j value.

One issue that is important when dealing with impulse response functions is that in the previous example in (84), we were able to impose the shock to u_{1t} to be of a specific size and we implicitly held u_{2t} fixed at zero to isolate the effect of u_{1t} on the system. To be able to do this, we need these shocks to be **orthogonal or uncorrelated**. This is the *'ceteris paribus'* analogue to cross-sectional analysis. In our simple linear setting, this means that the variance/covariance matrix of U_t (ie., $\Sigma = \text{Var}(U_t)$) needs to be diagonal. Now this will barely ever be the case in general when dealing with empirical data and one will thus need to find a way to enforce this by imposing a structure on the contemporaneous relationship between the elements of U_t. This will evidently impact on the assumption about how the **contemporaneous** variables of X_t are related to one another. Two frequently used ways to impose an orthogonal structure on U_t are the following:

a) *Cholesky decomposition* (or structural decomposition) of Σ (this is often considered to be atheoretical and is purely a matrix decomposition result.)

b) *Structural restrictions* (based on loose *'economic theory'*, which is mostly structural and ad hoc as it has no micro-foundations. When this is done the VAR model is often referred to as a Structural VAR (SVAR) model.)

Let me emphasise here again that the choice of how to make Σ diagonal is important as it **is an assumption that is imposed on the VAR model by the investigator** and **does not come from the data**. The first option that is listed under *a*), the Cholesky decomposition is a simple and convenient way of taking the *'square-root'* of a matrix and has certain consequences with it in terms of how the contemporaneous variables will be related to one another. The Cholesky decomposition is a purely mathematical tool to decompose a matrix into the product of an upper and lower triangular matrix. Due to its triangular structure, it imposes a **recursive ordering** on the contemporaneous variables in the VAR. We will see this explicitly in (99) below. Therefore, the use of the Cholesky decomposition implicitly imposes a structure on the contemporaneous relations of the variables X_t, so that the first variable x_{1t} is contemporaneously exogenous, the second variable x_{2t} in X_t is contemporaneously dependent on x_{1t}, and so forth. One needs to be aware of this implication of the Cholesky decomposition when working with VAR models for policy analysis.

The use of Structural restrictions listed under *b*) is a more coherent attempt to explicitly state and highlight the contemporaneous relations in X_t by formulating them explicitly and discussing the assumptions, rather than have them hidden away in the triangular implications of the Cholesky decomposition. Nonetheless, the assumptions

under the Structural restrictions are merely explicitly stated and there is no theoretic micro-foundations in terms of corroborating the restrictions from first principles of consumer choices, producer behaviour, frictions, equilibriums etc, as in fully micro-founded DSGE models. Although there exist larger Structural VAR models that impose the restrictions based on knowledge from DSGE model restrictions, the first uses of Structural restrictions were not formed with micro-foundations in mind.

5.2. VARs with Cholesky decomposition

Let us now look at the mechanics of computing IRFs. We can first have a look at how to get the IRFs with the Cholesky approach. In the normalised (ie., $A_0 = I_k$) and reduced form VAR we have

$$A(L)X_t = C + U_t \tag{91}$$

with $U_t \sim \mathsf{WN}(0, \Sigma)$, where $A(L) = I - A_1 L - ... - A_p L^p$ and Σ is not diagonal, ie., with non-zero entries on the off-diagonals. The Cholesky decomposition of a positive definite matrix Σ splits $\Sigma = PP'$ where P is a lower triangular matrix.

In the context of a simple bivariate scenario we have then:

$$\begin{aligned}
\Sigma &= PP' \\
\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix} &= \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} \\ 0 & p_{22} \end{bmatrix} \\
\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix} &= \begin{bmatrix} p_{11}^2 & p_{11}p_{21} \\ p_{11}p_{21} & p_{21}^2 + p_{22}^2 \end{bmatrix}
\end{aligned}$$

in general and matching the matrix entries on the left and on the right then yields the relations

$$\begin{aligned}
\sigma_{11}^2 &= p_{11}^2 \\
\sigma_{21} &= p_{11}p_{21} \\
\sigma_{22}^2 &= p_{21}^2 + p_{22}^2
\end{aligned}$$

which can be easily solved recursively for

$$\begin{aligned}
p_{11} &= \sigma_{11} \\
p_{21} &= \sigma_{21}/\sigma_{11}
\end{aligned}$$

$$p_{22} = \sqrt{\sigma_{22}^2 - \sigma_{21}^2/\sigma_{11}^2}.$$

Given this decomposition, the VAR in (91) can now be re-written as

$$\begin{aligned} X_t &= A(L)^{-1}C + A(L)^{-1}U_t \\ &= \mu + \Psi(L)U_t \qquad (92) \\ &= \mu + \underbrace{\Psi(L)P}_{\Phi(L)}\underbrace{P^{-1}U_t}_{W_t} \\ &= \mu + \Phi(L)W_t \qquad (93) \end{aligned}$$

where

$$\begin{aligned} \text{Var}\,(W_t) &= \text{Var}\left(P^{-1}U_t\right) \\ &= P^{-1}\underbrace{\text{Var}\,(U_t)}_{\Sigma = PP'}P^{-1\prime} \\ &= I_k \qquad (94) \end{aligned}$$

and I_k is the identity matrix, which is clearly diagonal. The term $\Phi(L) = \Phi_0 + \Phi_1 L + \Phi_2 L^2 + \ldots$ is another lag polynomial whose coefficients we can again work out. If we have a first order VAR, we get for

$$\begin{aligned} \Phi(L) &= \underbrace{\Phi_0}_{P} + \underbrace{\Phi_1 L}_{A_1 P} + \underbrace{\Phi_2 L^2}_{A_2 P} + \ldots \\ &= P + A_1 P + A_1^2 P + \cdots \end{aligned}$$

so that the set of j^{th} impulse responses can be computed as $A_1^j P$ for $j = 1, 2, \ldots$. Note here that the mean (or the constant term) have no influence on the computation of the impulse responses.

Let us look at a simple numerical example to visualise the computations.

Example 2 (IRFs of Bivariate VAR(1)): Suppose we have the following system

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

with

$$\Sigma = \begin{bmatrix} 0.25 & 0.3 \\ 0.3 & 0.9 \end{bmatrix}$$

and

$$P = \text{chol}(\Sigma) = \begin{bmatrix} 0.5 & 0 \\ 0.6 & 0.7348 \end{bmatrix}.$$

Then the first few impulse response matrices Φ are

$$\begin{aligned}
\Phi_0 &= P = \begin{bmatrix} 0.5 & 0 \\ 0.6 & 0.7348 \end{bmatrix} \\
\Phi_1 &= A_1 P = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0.6 & 0.7348 \end{bmatrix} = \begin{bmatrix} 0.26 & 0.0735 \\ 0.4 & 0.3674 \end{bmatrix} \\
\Phi_2 &= A_1^2 P = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix}^2 \begin{bmatrix} 0.5 & 0 \\ 0.6 & 0.7348 \end{bmatrix} = \begin{bmatrix} 0.1440 & 0.0661 \\ 0.2520 & 0.1984 \end{bmatrix} \\
&\vdots
\end{aligned}$$

and $\Phi_j = A_1^j P$ in general.

The IRF recursions above are commonly plotted to get a visual representation of the responses. A Plot of this relation is shown in Figure 1 below.

Note here that when using the Cholesky decomposition in this way, we implicitly assume that the innovations that are used to shock the system are of a **one-standard deviation** size. This is so because the transformed innovations vector W_t in (93) has unit variance and standard deviation, so that a unit innovation has a one-standard deviation size. This is evident from the first impulse at Φ_0 as it captures this magnitude.

To get a representation in terms of a **one unit shock**, we can use a scaling matrix D to rewrite the Cholesky decomposition so that we have a one unit shock rather than a one standard deviation shock, or alternatively, have the standard deviations in such a way that they are of a unit magnitude. This type of decomposition is called **"orthogonalised decomposition"** in Hamilton (1994, see page 322). The benefit of this is that, when the variables are log transformed,[5] then this naturally means that a one unit shock is a 1% shock, so that the impulse responses that follow out of this can be interpreted as elasticities, ie., a positive 1% interest rate shock or a tightening surprise then leads to a K% decrease in GDP in say j periods from now.

[5] natural log of course, with base e.

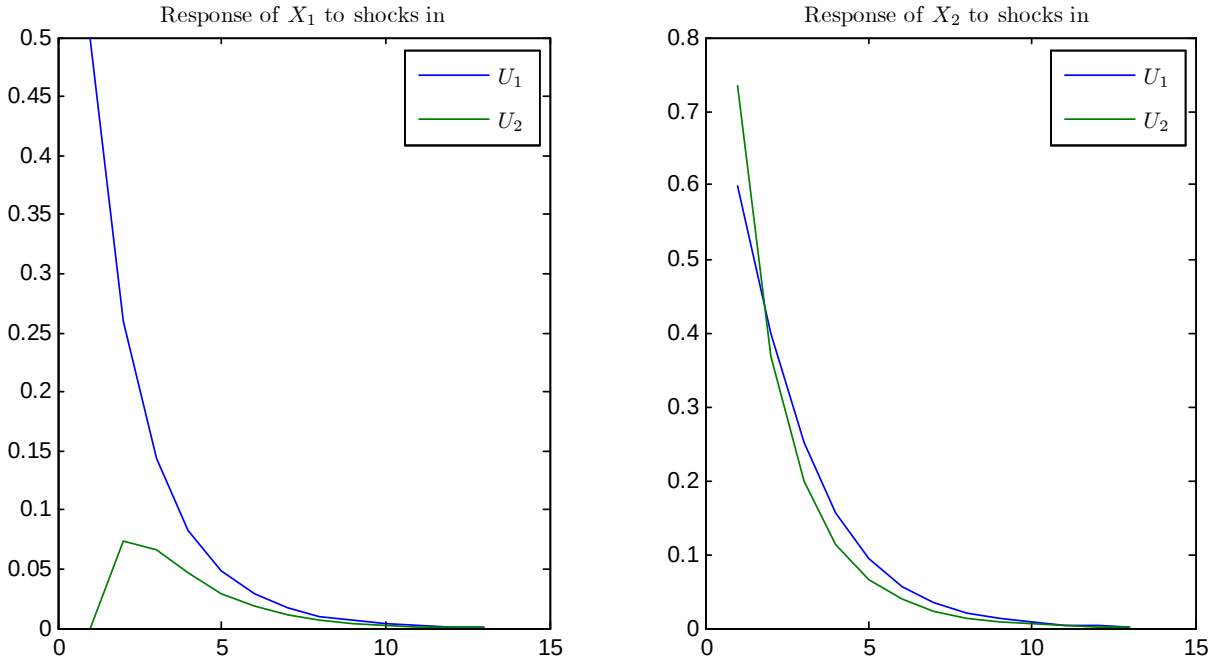

Figure 1: Plot of Cholesky based impulse response, ie. to one-standard deviation shock in U_t.

This is implemented as follows. Let $P = \text{chol}(\Sigma)$ as before. Now we seek a decomposition of the form

$$\begin{aligned}\Sigma &= (\tilde{P}D)(\tilde{P}D)' \\ &= PP'\end{aligned}$$

with the obvious relation $P = \tilde{P}D$. Now define $D = \text{diag}(P)$ so that for the bivariate scenario above we get

$$D = \begin{bmatrix} p_{11} & 0 \\ 0 & p_{22} \end{bmatrix}$$

and we can create a lower triangular matrix $\tilde{P}$ to have unit diagonal entries (and is thus normalised) by computing

$$\underbrace{PD^{-1}}_{=\tilde{P}} = \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} p_{11} & 0 \\ 0 & p_{22} \end{bmatrix}^{-1}$$

$$\tilde{P} = \begin{bmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{bmatrix} \begin{bmatrix} 1/p_{11} & 0 \\ 0 & 1/p_{22} \end{bmatrix}$$

$$\tilde{P} = \begin{bmatrix} 1 & 0 \\ \frac{p_{21}}{p_{22}} & 1 \end{bmatrix}.$$

Going back to the earlier VMA(∞) relation in (92) we have again

$$\begin{aligned} X_t &= \mu + \Psi(L)U_t \\ &= \mu + \underbrace{\Psi(L)\tilde{P}}_{\tilde{\Phi}(L)} \underbrace{\tilde{P}^{-1}U_t}_{\tilde{W}_t} \\ &= \mu + \tilde{\Phi}(L)\tilde{W}_t \end{aligned} \tag{95}$$

where

$$\begin{aligned} \text{Var}\left(\tilde{W}_t\right) &= \text{Var}\left(\tilde{P}^{-1}U_t\right) \\ &= \tilde{P}^{-1} \underbrace{\text{Var}\left(U_t\right)}_{\Sigma=(\tilde{P}D)(\tilde{P}D)'} \tilde{P}^{-1\prime} \\ &= \tilde{P}^{-1}(\tilde{P}D)(\tilde{P}D)'\tilde{P}^{-1\prime} \qquad (96) \\ &= D^2 \qquad (97) \end{aligned}$$

with $\tilde{W}_t \sim \text{WN}(0, D^2)$. Using the same logic as for the Cholesky case above, we can then see that the recursions become

$$\begin{aligned} \tilde{\Phi}_0 &= \tilde{P} = \begin{bmatrix} 1 & 0 \\ 1.2 & 1 \end{bmatrix} \\ \tilde{\Phi}_1 &= A_1\tilde{P} = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1.2 & 1 \end{bmatrix} = \begin{bmatrix} 0.52 & 0.1 \\ 0.8 & 0.5 \end{bmatrix} \\ \tilde{\Phi}_2 &= A_1^2\tilde{P} = \begin{bmatrix} 0.4 & 0.1 \\ 0.2 & 0.5 \end{bmatrix}^2 \begin{bmatrix} 1 & 0 \\ 1.2 & 1 \end{bmatrix} = \begin{bmatrix} 0.288 & 0.09 \\ 0.504 & 0.27 \end{bmatrix} \\ &\vdots. \end{aligned}$$

A plot of the IRFs based on one unit shock in U_t are shown below in Figure 2.

Note the distinction here. The *"own response"* (or elasticity with itself) in the first period is always 1. It should be kept in mind that the shock that is imposed is always of the one unit size, it is just that under the Cholesky decomposition it is so that the shock itself also has a standard deviation of unity, so that it has the interpretation of a one

standard deviation shock. We can see from the above that one can always move between these two definitions of the IRFs by simply rescaling the $\Phi(L)$ terms in (93) by D to get $\tilde{\Phi}(L)$ in (95) and vice versa.

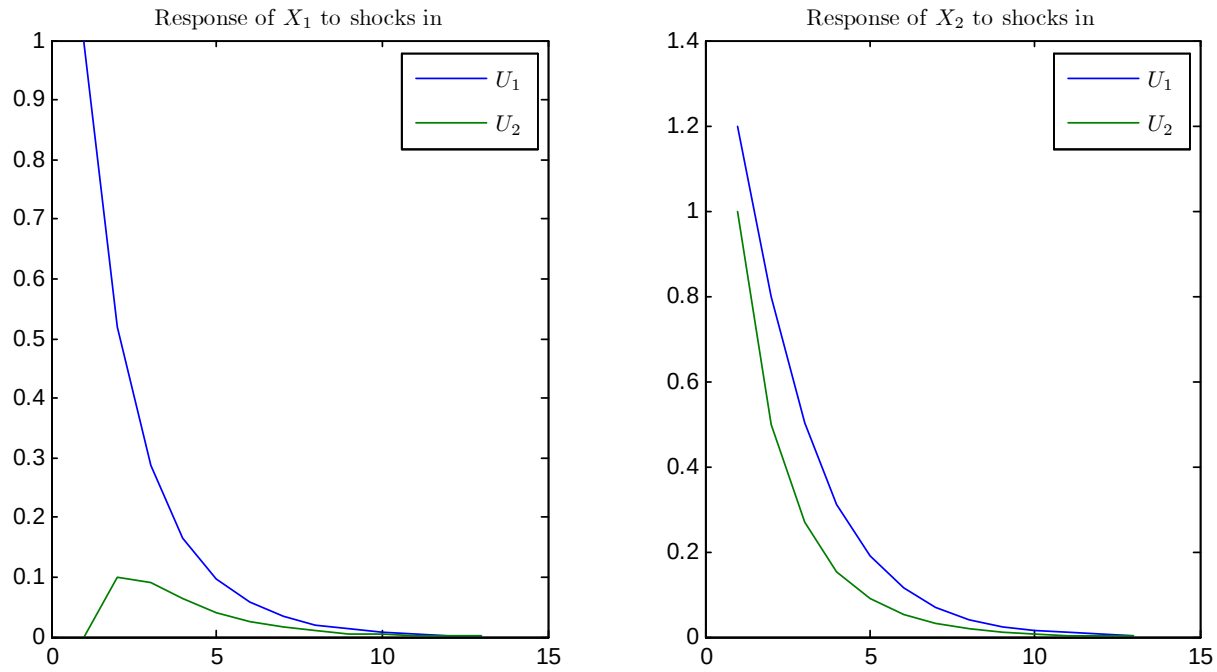

Figure 2: Plot of impulse responses based on a one unit shock in U_t.

5.2.1. Structural interpretation

As pointed out earlier, the Cholesky decomposed VAR system becomes a recursive VAR and hence has an implication about the causal ordering (sometimes referred to as *'Wold Causal ordering'* as he was the first to point it out) and thus the economic structure of the system. As an example, consider the simplest bi-variate VAR(1) with a zero mean:

$$\begin{aligned} X_t &= A_1 X_{t-1} + U_t \\ P^{-1}X_t &= \underbrace{P^{-1}A_1}_{=B_1} X_{t-1} + \underbrace{P^{-1}U_t}_{=W_t} \\ P^{-1}X_t &= B_1 X_{t-1} + W_t \end{aligned} \tag{98}$$

with $W_t \sim \mathsf{WN}(0, I_k)$. Note that the inverse of P, $P^{-1} = P^*$ is also lower triangular due to P being lower triangular. Writing out the system we get:

$$P^* X_t = B_1 X_{t-1} + W_t$$

$$\begin{bmatrix} p^*_{11} & 0 \\ p^*_{21} & p^*_{22} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{1t} \end{bmatrix}$$

and multiplying out

$$p^*_{11} x_{1t} = b_{11} x_{1t-1} + b_{12} x_{2t-1} + w_{1t} \tag{99a}$$

$$p^*_{22} x_{2t} = -p^*_{21} x_{1t} + b_{21} x_{1t-1} + b_{22} x_{2t-1} + w_{2t}. \tag{99b}$$

So we can see from (99) that there is no contemporaneous effect on x_{1t} from the other variable x_{2t} and also that x_{2t} is only affected by x_{1t} in the same time period. This can be generalised to higher order systems but the implications should be clear.

The use of the Cholesky decomposition was Sims' idea. It imposes strong assumptions about the economic order and hence the causal relations of the system. Sims originally suggested to use different orderings in X_t to see whether the economic story in terms of the IRFs remains the same. This does not seem very scientific and indeed there is not solution to the problem if the ordering does alter the shape and magnitude of the responses. Nonetheless, other Structural relations are possible. Some are focused on specifying the contemporaneous relations based on some weak form of economic theory. These are referred to as Structural VAR (SVAR) models. Yet, another approach focuses on specifying long-run restrictions, which was introduced by Blanchard and Quah (1989). A fairly new approach is based on sign restrictions of the IRFs. Fundamentally, the question and solution approaches are all related to identification questions of the system that is analysed, so it will make sense to look at VAR models form a more general formulation.

5.3. General formulation of VAR models and Identification

Let us step back from the reduced form VAR that we introduced earlier for a moment and think about writing the VAR out as a general system of simultaneous equations with dynamics in them. To simplify things, we will cut out the constant term and look at a simple first order VAR model. Also, since it is instructive to think about some macroeconomic system, rather than just some x_{jt} variables, let us think about a four dimensional macroeconomic VAR relating interest rates (r), money (m), prices (p) and output (y), taking the form

$$B_0 X_t = B_1 X_{t-1} + \epsilon_t \tag{100}$$

where $X_t = [r_t \; m_t \; p_t \; y_t]'$, $\epsilon_t = [\varepsilon_{rt} \; \varepsilon_{mt} \; \varepsilon_{pt} \; \varepsilon_{yt}]'$, B_0 and B_1 are the coefficient matrices of the contemporaneous and lagged terms of X_t, respectively, and $\epsilon_t \sim \mathsf{WN}(0_4, \Omega)$, where Ω

is a diagonal matrix implying that the shocks ϵ_t are **uncorrelated**. The ϵ_t are frequently referred to as **structural shocks**, because the system in (100) is known as a structural model (or structural VAR). Ignoring the $B_1 X_{t-1}$ term for the moment, written out, the relation in (100) looks like this:

$$b_{11}^0 r_t + b_{12}^0 m_t + b_{13}^0 p_t + b_{14}^0 y_t = \varepsilon_{rt} \tag{101}$$

$$b_{21}^0 r_t + b_{22}^0 m_t + b_{23}^0 p_t + b_{24}^0 y_t = \varepsilon_{mt} \tag{102}$$

$$b_{31}^0 r_t + b_{32}^0 m_t + b_{33}^0 p_t + b_{34}^0 y_t = \varepsilon_{pt} \tag{103}$$

$$b_{41}^0 r_t + b_{42}^0 m_t + b_{43}^0 p_t + b_{44}^0 y_t = \varepsilon_{yt} \tag{104}$$

where b_{ij}^0 are the coefficients of B_0.

How would we estimate a system such as the one in (100) with traditional simultaneous equation methods? It should be clear that the B_1 coefficients attached to X_{t-1} can always be estimated as the variables X_{t-1} are predetermined. It is the coefficients in the contemporaneous matrix B_0 that need to be *identified*. Note that we cannot just run OLS on the individual equations in (100) as they are contemporaneously correlated or endogenous, so this is the same as the standard endogenous variable regression problem in econometrics. Now to identify the model, one can think about six possible ways to restrict B_0. Fundamentally, we might consider imposing some constraints upon B_0 either from the first moments of the system, the second moments or possibly higher order moments. These are as follows:

First Order (relating to the mean)

- **Normalisation:** we can choose to set k of the elements in B_0 to unity. In the system above we could set $b_{ii}^0 = 1, i = 1, \ldots, 4$. This reduces the number of unknown parameters in B_0 to $k^2 - k$.
- **Identities:** reduces the number of unknown parameters by a structure specific amount. In an identity the elements in the row of B_0 corresponding to the identity are known and do not need to be estimated. There are rarely any in VAR systems.
- **Exclusion restrictions:** some of the b_{ij}^0 are set to zero. This reduces the number of unknown parameters again by a structure specific amount.
- **Linear restrictions:** coefficients add to a pre-specified value. This might either be among the B_0 coefficients or between these and the coefficients in B_1^*.

Second (and higher) Order (relating to the variance, skewness etc.)

- **Restrictions on Var(ϵ_t):** in particular, if we make this a diagonal matrix as we have

done above, we can use the fact that $E(\varepsilon_{jt}\varepsilon_{it}) = 0$ which means we have moment restrictions that can be used for the estimation of B_0.

Non-linearities: restrictions on the shape of the impulse responses as well as the symmetry of the density are place.

In practice, one will work with a combination of the above listed **restrictions**.

As we saw before, the Cholesky decomposition that was used in the original work with VARs by Sims (1980) made the assumption that B_0 was lower triangular. This puts a recursive structure on the system and thus, as pointed out by Cooley and Leroy (1985) implies some view of the way in which the economic system works ie., it implies a *"causal ordering"*. To put this in a broader perspective, the normalisation and exclusion restrictions that are implicitly used in the recursively ordered system, and hence in the Cholesky orthogonalisation of the model, are:

$$\begin{aligned} b_{ii}^0 &= 1 \\ b_{12}^0 = 0, b_{13}^0 &= 0, b_{14}^0 = 0 \\ b_{23}^0 = 0, b_{24}^0 &= 0 \\ b_{34}^0 &= 0 \end{aligned}$$

for all $i = 1, \ldots, k$.

This reduces the structural system in (100) to (again excluding the dynamics of the VAR(1) for simplicity)

$$B_0 X_t = \epsilon_t$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ b_{21}^0 & 1 & 0 & 0 \\ b_{31}^0 & b_{32}^0 & 1 & 0 \\ b_{41}^0 & b_{42}^0 & b_{43}^0 & 1 \end{bmatrix}}_{B_0} \underbrace{\begin{bmatrix} r_t \\ m_t \\ p_t \\ y_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}}_{\epsilon_t}$$

and expanding out yields:

$$\begin{aligned} r_t &= \varepsilon_{1t} \\ b_{21}^0 r_t + m_t &= \varepsilon_{2t} \\ b_{31}^0 r_t + b_{32}^0 m_t + p_t &= \varepsilon_{3t} \\ b_{41}^0 r_t + b_{42}^0 m_t + b_{43}^0 p_t + y_t &= \varepsilon_{4t} \end{aligned}$$

Sims (1980) was well aware of the difficulty that the ordering is rather arbitrary, and initially seems to have suggested that one experiment with different causal orderings to see how different the outcomes would be in terms of the impulse responses, and one sees this action followed fairly commonly in the literature. It is not very convincing, particularly if the system is very large, when there are an enormous number of possible orderings and authors are unlikely to look at all possible ones. Moreover, since the magnitude of any changes depends directly upon the covariance of the reduced form residuals U_t, one might as well have just tested if this was zero and made an appropriate response.

Given the recursive structure, it is clear that we can apply OLS to all equations since the regressors in each equation are uncorrelated with its error owing to the assumption that the shocks ϵ_t are uncorrelated with one another. Thus the *triangular assumption* for B_0, and the assumption of uncorrelated errors, allows one to estimate the remaining unknown elements in B_0 sequentially by exploiting the recursive nature of the system.

We can now generalise the treatment of structural VARs (SVARs) by looking at how to recover the structural model parameters given by the SVAR in (100) from the reduced from VAR. That is, how many restrictions do we need to put on the system in terms of B_0 and $\text{Var}(\epsilon_t) = \Omega$ so as to be able to recover them uniquely from Σ and A_1 in the **reduced form VAR**

$$X_t = A_1 X_{t-1} + U_t \tag{105}$$

where $\text{Var}(U_t) = \Sigma$ as before and A_1 is our standard reduced form VAR coefficient on the first lag. Now we can easily invert B_0 in (100) to go from the **structural VAR** to the **reduced form VAR** as:

$$B_0 X_t = B_1 X_{t-1} + \epsilon_t \tag{106}$$

$$\begin{aligned} X_t &= \underbrace{B_0^{-1} B_1}_{A_1} X_{t-1} + \underbrace{B_0^{-1} \epsilon_t}_{U_t} \\ &= A_1 X_{t-1} + U_t \end{aligned} \tag{107}$$

where

$$\begin{aligned} \text{Var}(U_t) &= \Sigma \\ &= \text{Var}(B_0^{-1} \epsilon_t) \\ &= B_0^{-1} \text{Var}(\epsilon_t) B_0^{-1\prime} \end{aligned}$$

$$\Sigma = B_0^{-1} \Omega B_0^{-1\prime} \tag{108}$$

and Ω is a diagonal matrix as defined before.

How can we recover B_0 and $\Omega = \text{Var}(\epsilon_t)$ in the SVAR in (106) from the reduced form VAR parameters in (107)? One way is to look at the relation in (108), that is, from $\Sigma = B_0^{-1} \Omega B_0^{-1\prime}$. Note that any general (non-symmetric) $k \times k$ matrix has k^2 terms in it. Now Σ is a symmetric matrix (due to it being a covariance matrix, such that $\text{Cov}(u_{it}, u_{jt}) = \text{Cov}(u_{jt}, u_{it})$ for all $j \neq i$). This means that terms above and below the diagonal (the variances of u_{it}) will be the same. So we will have $(k^2 - k)$ terms that are not on the diagonal. Half of them above and half of them below the main diagonal. Thus, the symmetric matrix Σ will then have k terms on the diagonal and $(k^2 - k)/2$ terms off the diagonal, so a total of $k + (k^2 - k)/2 = k(k+1)/2$ terms of variance and covariance terms.

To solve for B_0 and Ω from the relation in (108), all that needs to be done is to match the $k(k+1)/2$ (known or estimated) variances/covariances from the reduced form model with the right hand side terms $B_0^{-1} \Omega B_0^{-1\prime}$. Note that Ω will have only k unknown terms in it, due to it being a diagonal matrix (uncorrelated structural shocks ϵ_t assumption). If we leave B_0 unrestricted, there will be k^2 terms in it, yielding a total of $k(k+1)$ unknowns to recover from Σ. The difference in the number of terms is $k(k+1) - k(k+1)/2 = k(k+1)/2$, meaning that $k(k+1)/2$ will need to be placed on B_0 to identify the model.[6]

What kind of (suitable) restrictions can be placed on B_0 was discussed in above. The starting point is generally to place k normalisation restrictions on B_0, so that it has ones on the main diagonal. That leaves another $k(k+1)/2 - k = k(k-1)/2$ exclusion (or other) restrictions that need to be imposed to uniquely recover the remaining elements of B_0. The Cholesky decomposition of Sims (1980) which implies a recursive (triangular) causal relation puts zero (exclusion) restrictions on the remaining $k(k-1)/2$ un-identifiable parameters in B_0. This is just one way of going about recovering the B_0. There are non-triangular systems that fulfill the necessary condition (also known as order condition) to be able to recover the structural parameters. One fairly popular one is discussed below.

[6] This is a necessary condition but not sufficient. The caveat is ignored in what follows to keep the exposition simple.

5.3.1. An empirical model: The IS-LM SVAR

If one is going to introduce structural information into a VAR, there is a strong case for doing it much more explicitly than Sims (1980) did. Subsequently, Sims (1986), Bernanke (1986) and Blanchard and Watson (1987) pointed out that one might prefer a different pattern to B_0, with the aim of providing a more accurate description of how the economy was thought to operate. Provided the number of unknowns in B_0 remains the same, any such alternative patterns will be observationally equivalent, as they give rise to the same reduced from VAR that is fitted and any choice between them must be made on grounds other than the information that is in the data.

As an example of a *'non-triangular'* pattern set $k = 3$, $X_t = [y_t \; i_t \; m_t]'$, where y_t is output, m_t is real money and i_t is an interest rate, and postulate an economy governed by IS/LM relations with a money supply rule:

$$y_t = b_{12}^0 i_t + \varepsilon_{IS,t} \qquad \text{(IS curve)} \tag{109a}$$

$$i_t = b_{21}^0 y_t + b_{23}^0 m_t + \varepsilon_{LM,t} \qquad \text{(inverse LM curve)} \tag{109b}$$

$$m_t = \varepsilon_{MS,t} \qquad \text{(money supply rule).} \tag{109c}$$

This simple model has the contemporaneous structure

$$B_0 X_t = \epsilon_t$$

$$\underbrace{\begin{bmatrix} 1 & -b_{12}^0 & 0 \\ -b_{21}^0 & 1 & -b_{23}^0 \\ 0 & 0 & 1 \end{bmatrix}}_{B_0} \underbrace{\begin{bmatrix} y_t \\ i_t \\ m_t \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} \varepsilon_{IS,t} \\ \varepsilon_{LM,t} \\ \varepsilon_{MS,t} \end{bmatrix}}_{\epsilon_t}$$

so

$$B_0 = \begin{bmatrix} 1 & -b_{12}^0 & 0 \\ -b_{21}^0 & 1 & -b_{23}^0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{110}$$

Now B_0 in (110) has the same number of unknown parameters as a *'triangular representation'* would have here, that is, 3, but seems to have a better theoretical base than the assumption that there is a recursive structure.

The emergence of '*structural VAR's*' represents a much more satisfactory mode of analysis than the recursive assumptions imposed in early VAR work, and has had the effect of moving the VAR back towards structural ideas. In doing so the differences between the

Cowles Commission philosophy and the VAR methodology have become more sharply etched. The *'modus operandi'* of the Cowles Commission was to leave $\Omega = \text{Var}(\epsilon_t)$ unrestricted, but to allow B_0 and B_1 to have **no more than k^2 elements**, thereby enabling those parameters to be recovered from the k^2 elements in the reduced form A_1. In contrast the VAR approach puts restrictions on Ω and B_0, but leaves A_1 unrestricted.[7]

Most of the reasoning that goes into specifying B_0 is very similar to that for traditional structural models — one examines the influences on aggregate demand, unemployment, prices, wages and money demand. In the VAR methodology no constraints are placed upon B_1, since **all predetermined variables enter every equation**. Consequently, B_0 has to be identified by restricting the variance/covariance matrix of ϵ_t (ie., Ω), rather than by the conventional method of exclusion restrictions on B_1. But, when it comes down to the wire, we are back to traditional modelling, albeit with the use of covariance restrictions for identification, a possibility long realised but little used. Of course one might validly ask why this is any more credible than restricting B_1? One answer is that theoretical models have little to say about the structure of errors ϵ_t and much more to say about conditional means, and from a rational expectations perspective it is hard to exclude variables from the mean of any equation.

5.4. VARs with long-run restrictions

5.4.1. A simple Bivariate Model

Another popular approach to structural VARs was introduced by Blanchard and Quah (1989). They have a bivariate system[8]

$$x_{1t} = b_{12}^0 x_{2t} + b_{11} x_{1t-1} + b_{12} x_{2t-1} + \varepsilon_{1t} \tag{111a}$$

$$x_{2t} = b_{21}^0 x_{1t} + b_{21} x_{1t-1} + b_{22} x_{2t-1} + \varepsilon_{2t} \tag{111b}$$

or

$$B_0 X_t = B_1 X_{t-1} + \epsilon_t \tag{112}$$

[7] The interpretation given here to the *'Cowles Commission philosophy'* applies only to exactly identified systems, i.e., the number of parameters in the reduced form equals the number in the structure, so that all estimation is being done with the reduced form and we are just re-interpreting these estimates in a structural way. This appears to be a sensible way to effect a comparison between the methodologies, although it should be borne in mind that the Cowles Commission went much further and concentrated upon *over-identifying* restrictions, albeit recommending that these be tested. Moreover, the Cowles Commission did not use the VAR as they invoked exogeneity assumptions.

[8] Note here that b_{ij}^1 is usually our coefficient (elements of B_1 matrix) on the first lag. Since we only have a first order model, I will just write b_{ij} instead.

$$\underbrace{\begin{bmatrix} 1 & -b_{12}^0 \\ -b_{21}^0 & 1 \end{bmatrix}}_{B_0}\underbrace{\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}}_{X_t} = \underbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}}_{B_1}\underbrace{\begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix}}_{X_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}}_{\epsilon_t}$$

with $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ being supply and demand shocks, respectively. Blanchard and Quah (1989) argue that a demand shock ($\varepsilon_{2,t}$) should have a zero long-run effect on output, ie., only be transitory, while a supply shock will not be zero and persist forever, ie., have a permanent effect on output. Note from the relations in (111) that due to the presence of x_{1t} and x_{2t} one the left and right hand side of the relations (**simultaneity**), we cannot estimate this model by *'running OLS'*.

Let x_{1t} be the output variable. Now the system in (112) can be written as

$$B(L)X_t = \epsilon_t \tag{113}$$

where $B(L) = B_0 - B_1 L$ with L again being the lag operator. We then have:

$$\begin{aligned} B(L) &= \begin{bmatrix} 1 & -b_{12}^0 \\ -b_{21}^0 & 1 \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} L \\ &= \begin{bmatrix} 1 - b_{11}L & -b_{12}^0 - b_{12}L \\ -b_{21}^0 - b_{21}L & 1 - b_{22}L \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} X_t &= B(L)^{-1}\epsilon_t \\ &= \Psi(L)\epsilon_t \\ &= (\Psi_0 + \Psi_1 L + \ldots)\epsilon_t. \end{aligned}$$

The elements Ψ_j are the j period ahead impulse responses of X_t to a transitory unit rise in the shocks ϵ_t, ie., $\Psi_j = \frac{\partial X_{t+j}}{\partial \epsilon_t}$. The sum of these Ψ_j are the **long-run responses**, that is, the cumulated effect of a change in the shocks. Now $\Psi(1) = \sum_{j=1}^{\infty} \Psi_j$ and, in the Blanchard and Quah (1989) case, $\Psi(1)$ has the form

$$\Psi(1) = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} = \begin{bmatrix} \psi_{11} & 0 \\ \psi_{21} & \psi_{22} \end{bmatrix} \tag{114}$$

because we assume that ε_{2t} (demand shock) has a zero long-run impact on x_{1t} (output).

Now we know that $\Psi(L) = B(L)^{-1}$ means that $\Psi(L)B(L) = I_2$ and $\Psi(1)B(1) = I_2$. Hence

$$\begin{bmatrix} \psi_{11} & 0 \\ \psi_{21} & \psi_{22} \end{bmatrix} \begin{bmatrix} 1 - b_{11}L & -b_{12}^0 - b_{12}L \\ -b_{21}^0 - b_{21}L & 1 - b_{22}L \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

from which we get the relation $\psi_{11}(-b_{12}^0 - b_{12}) = 0$. So there are two possible scenarios here to make this true. Either $\psi_{11} = 0$ or $(-b_{12}^0 - b_{12}) = 0$. If $\psi_{11} = 0$ then $\Psi(1)$ would be a singular matrix and there would not be a VAR representation. This cannot be the case as we started of with a VAR. So we have to have $(-b_{12}^0 - b_{12}) = 0$ which gives the restriction on the parameters of the first equation as $b_{12} = -b_{12}^0$. This can be written as

$$x_{1t} = b_{12}^0 \Delta x_{2t} + a_{11} x_{1t-1} + \varepsilon_{1t}. \tag{115}$$

Consider the estimation of this equation (115). The relation in (115) can be estimated by using x_{1t-1}, x_{2t-1} as **instruments** for x_{1t-1} and Δx_{2t}. Note that we cannot just regress x_{1t} on Δx_{2t} because of the simultaneity problem outlined before. Having estimated these coefficients we can get the residuals from (115) $\hat{\varepsilon}_{1t}$ and then use $\hat{\varepsilon}_{1t}$, x_{1t-1} and x_{2t-1} as **instruments** for x_{1t-1}, x_{2t-1} and x_{1t} in the equation

$$x_{2t} = b_{21}^0 x_{1t} + b_{21} x_{1t-1} + b_{22} x_{2t-1} + \varepsilon_{2t} \tag{116}$$

or (111b) above. Note that, unlike in the recursive case, we cannot just use x_{1t} as an instrument because $E(x_{1t}\varepsilon_{2t}) = b_{21}^0 \text{Var}(\varepsilon_{2t}) \neq 0$ even if $E(\varepsilon_{1t}\varepsilon_{2t}) = 0$, that is, even with the *'structural shocks are uncorrelated'* assumption.

5.4.2. A more general model

The Blanchard and Quah (1989) model for a 2 variable system given above generalises to higher order systems. It is based on the result that $\Psi(1)B(1) = I_k$. Effectively long-run constraints replace the triangular assumption in B_0 in a recursive system with *"long-run recursivity"* in which $\Psi(1)$ is **triangular**. To see the generalisation, take a three dimensional SVAR(1) system, i.e., $k = 3$ with

$$\begin{aligned} x_{1t} &= b_{12}^0 x_{2t} + b_{13}^0 x_{3t} + b_{11} x_{1t-1} + b_{12} x_{2t-1} + b_{13} x_{3t-1} + \varepsilon_{1t} \\ x_{2t} &= b_{21}^0 x_{1t} + b_{23}^0 x_{3t} + b_{21} x_{1t-1} + b_{22} x_{2t-1} + b_{23} x_{3t-1} + \varepsilon_{2t} \\ x_{3t} &= b_{31}^0 x_{1t} + b_{32}^0 x_{2t} + b_{31} x_{1t-1} + b_{32} x_{2t-1} + b_{33} x_{3t-1} + \varepsilon_{3t} \end{aligned} \tag{117}$$

and thus

$$\Psi(1) = \begin{bmatrix} \psi_{11} & 0 & 0 \\ \psi_{21} & \psi_{22} & 0 \\ \psi_{31} & \psi_{32} & \psi_{33} \end{bmatrix}, \; B(1) = \begin{bmatrix} 1 - b_{11} & -b_{12}^0 - b_{12} & -b_{13}^0 - b_{13} \\ -b_{21}^0 - b_{21} & 1 - b_{22} & -b_{23}^0 - b_{23} \\ -b_{31}^0 - b_{31} & -b_{32}^0 - b_{32} & 1 - b_{33} \end{bmatrix}. \tag{118}$$

We see that $\Psi(1)B(1) = I_k$ yields the relations:

$$\underbrace{\begin{bmatrix} \psi_{11} & 0 & 0 \end{bmatrix} \begin{bmatrix} -b_{13}^0 - b_{13} \\ -b_{23}^0 - b_{23} \\ 1 - b_{33} \end{bmatrix}}_{= \psi_{11}(-b_{13}^0 - b_{13})} = 0$$

$$\underbrace{\begin{bmatrix} \psi_{11} & 0 & 0 \end{bmatrix} \begin{bmatrix} -b_{12}^0 - b_{12} \\ 1 - b_{22} \\ -b_{32}^0 - b_{32} \end{bmatrix}}_{= \psi_{11}(-b_{12}^0 - b_{12})} = 0$$

implies that

$$-b_{12}^0 - b_{12} = 0 \tag{119a}$$

$$-b_{13}^0 - b_{13} = 0 \tag{119b}$$

$$-b_{23}^0 - b_{23} = 0 \tag{119c}$$

due to

$$\underbrace{\begin{bmatrix} \psi_{21} & \psi_{22} & 0 \end{bmatrix} \begin{bmatrix} -b_{13}^0 - b_{13} \\ -b_{23}^0 - b_{23} \\ 1 - b_{33} \end{bmatrix}}_{= \psi_{21}(-b_{13}^0 - b_{13}) + \psi_{22}(-b_{23}^0 - b_{23})} = 0$$

$$\psi_{21} \underbrace{(-b_{13}^0 - b_{13})}_{= 0 \text{ from (119b)}} + \psi_{22}(-b_{23}^0 - b_{23}) = 0$$

$$-b_{23}^0 - b_{23} = 0$$

because ψ_{22} cannot be zero as there would again be no VAR representation of the process.[9]

[9] Recall that the inverse of a diagonal matrix has inverses of the diagonal terms on the diagonal, so that if

With these long-run restrictions in place, the system becomes:

$$\begin{aligned} x_{1t} &= b_{11}x_{1t-1} + b^0_{12}\Delta x_{2t} + b^0_{13}\Delta x_{3t} + \varepsilon_{1t} \\ x_{2t} &= b_{21}x_{1t-1} + b^0_{21}x_{1t} + b_{12}x_{2t-1} + b^0_{23}\Delta x_{3t} + \varepsilon_{2t} \\ x_{3t} &= b_{31}x_{1t-1} + b^0_{31}x_{1t} + b_{32}x_{2t-1} + b^0_{32}x_{2t} + b_{33}x_{3t-1} + \varepsilon_{3t}. \end{aligned} \tag{120}$$

One can now estimate the individual equations of this system in the following way.

1) Estimate the first equation using x_{1t-1}, x_{2t-1} and x_{3t-1} as **instruments** for x_{1t-1}, Δx_{2t} and Δx_{3t}.
2) Construct the (fitted) residuals from this equation $\hat{\varepsilon}_{1t}$ and use this plus x_{1t-1}, x_{2t-1} and x_{3t-1} as **instruments** in the second equation.
3) Finally use x_{1t-1}, x_{2t-1}, x_{3t-1}, $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$ as **instruments** in the third equation.

5.5. The Uncorrelated Shocks Assumption

A difficulty with SVARs that is often glossed over is the assumption that the structural shocks, the $\epsilon_t = [\varepsilon_{1t}\ \varepsilon_{2t} \dots \varepsilon_{kt}]'$, are **uncorrelated** with each other. One might argue that this is not unreasonable to assume if the system X_t is of large enough dimension but, in other cases, it may well be that a factor that is common to the explanation of more than one member of X_t has been omitted. In this case the factor would be included in the error terms and, being common, will mean that the assumption of uncorrelated errors is invalid. Consequently, imposing an incorrect restriction will generally bias estimators of the parameters of the SVAR.

To think about the likelihood of that happening one might return to the economic models higher up on the curve. A popular miniature model has been the New Keynesian policy model which takes the form

$$\begin{aligned} y_t - y^*_t &= a_1(i_{t-1} - \pi_{t-1}) + a_2(y_{t-1} - y^*_{t-1}) + \varepsilon_{IS,t} \\ \pi_t - \overline{\pi} &= b_1(\pi_{t-1} - \overline{\pi}) + b_2(y_{t-1} - y^*_{t-1}) + \varepsilon_{AS,t} \\ i_t - \pi_t &= c_1(y_t - y^*_t) + c_2(\pi_t - \overline{\pi}) + c_3(y_{t-1} - y^*_{t-1}) \\ &\quad + c_4(\pi_{t-1} - \overline{\pi}) + c_5(i_{t-1} - \pi_{t-1}) + \varepsilon_{MP,t} \end{aligned} \tag{121}$$

where the equations are the IS curve, the Phillips curve (AS curve) and the interest rate or monetary policy (MP) rule respectively. The variable $y_t - y^*_t$ is an output gap, y^*_t is

ψ_{22} is 0, we would have $1/0$ as its inverse term.

potential output, π_t is the inflation rate, $\overline{\pi}$ is the target inflation rate, i_t is the nominal interest rate and $\varepsilon_{IS,t}, \varepsilon_{AS,t}$ and $\varepsilon_{MP,t}$ are demand side, supply side and monetary policy shocks. In the interest of simplicity, I have abstracted from any expectation terms and have made the responses in the first two equations lagged ones rather then contemporaneous. Now the SVAR literature would imply that the first two equations can be estimated (separately) by OLS and, with the assumption that $\varepsilon_{IS,t}$ and $\varepsilon_{AS,t}$ are **uncorrelated** with $\varepsilon_{MP,t}$, one can use the estimated residuals $\hat{\varepsilon}_{IS,t}$ and $\hat{\varepsilon}_{AS,t}$ from those two equations as instruments for $(y_t - y_t^*)$ and $(\pi_t - \overline{\pi})$ in the third equation to estimate c_1 and c_2.

Now there is a potential catch here. The SVAR literature has invariably worked with equations in y_t and **NOT the output gap**, and so it is the following system that is actually estimated, rather than (121):

$$\begin{aligned} y_t &= a_1(i_{t-1} - \pi_{t-1}) + a_2 y_{t-1} + \{\varepsilon_{IS,t} + y_t^* - a_2 y_{t-1}^*\} \\ \pi_t - \overline{\pi} &= b_1(\pi_{t-1} - \overline{\pi}) + b_2 y_{t-1} + \{\varepsilon_{AS,t} - b_2 y_{t-1}^*\} \\ i_t - \pi_t &= c_1 y_t + c_2(\pi_t - \overline{\pi}) + c_3 y_{t-1} + c_4(\pi_{t-1} - \overline{\pi}) \\ &\quad + c_5(i_{t-1} - \pi_{t-1}) + \{\varepsilon_{MP,t} - c_1 y_t^* - c_3 y_{t-1}^*\} \end{aligned} \tag{122}$$

In (122) the errors (in curly brackets) are clearly correlated due to the common y_{t-1}^* term. Therefore the residuals from estimating the first two equations in (122) **cannot** be used as instruments in the last equation. If they are, then one will get inconsistent estimates of the parameters, leading to biases in the estimated impulse responses for the monetary shock. The study by Giordani (2004) showed that this is a possible explanation of the *price puzzle* in many SVAR studies, wherein a contractionary monetary policy shock (an unanticipated increase in interest rates) leads to a rise, and not fall, in the price level (inflation). There is nothing wrong about the fit of the (reduced form) VAR in this instance, but rather it is how one moves closer to the '*theoretical*' economic model, the SVAR in this instance, that creates the problems (bias).

5.6. Computing uncertainties about the IRFs

The impulse responses based on the Cholesky decomposition that were outline above were given for known matrix polynomial $A(L)$ and known covariance matrix Σ. When these are estimated from a data set, then there will naturally be uncertainties in the $A(L)$ as well as in the covariance matrix Σ and hence also in the IRFs. So it is necessary to find some sort of standard errors for theses impulse responses. We will initially go through an example of how to get asymptotic standard errors, but it should be noted here that

there are a number of studies that point to asymptotic standard errors performing rather poorly in finite samples (see for example Kilian, 1998, 1999). This is due to the same bias that one gets in AR models in general and in IRF analysis this bias gets amplified due to non-linear transformations of these coefficients.

5.6.1. Asymptotic standard errors of IRFs

To get some understanding how to compute asymptotic standard errors of the IRFs, let us look at how to do this again in the simple AR(1) model that is set up the same way as before. Let

$$x_t = \alpha_1 x_{t-1} + u_t \tag{123}$$

with $u_t \sim \mathsf{WN}(0, \sigma^2)$ and $|\alpha_1| < 1$. The MA(∞) representation is

$$x_t = \alpha_1 u_t + \alpha_1^2 u_{t-2} + \alpha_1^3 u_{t-3} + \alpha_1^4 u_{t-4} + \dots \tag{124}$$

where each one of the $\alpha_1^j, \forall j = 0, 1, 2, \dots$ represents the impulse responses at the different time periods to a shock in u_{t-j}. We know from OLS theory that $\hat{\alpha}_1 \sim \mathsf{N}\left(\alpha_1, \sigma^2_{\alpha_1}\right)$, with $\sigma^2_{\alpha_1} = T^{-1}(1 - \alpha_1^2)$, which can be estimated by $T^{-1}(1 - \hat{\alpha}_1^2)$. Now it is clear that the $j-$step ahead impulse response is estimated to be $\hat{\alpha}_1^j$, which is a non-linear transformation of $\hat{\alpha}_1$.

Due to the non-linearity caused by the powering up of $\hat{\alpha}_1$, the standard approach to get asymptotic standard errors is to argue that $\hat{\alpha}_1^h$ can be replaced by its **linear Taylor series** expansion around the true value α_1^j, ie.,

$$\hat{\alpha}_1^j \simeq \alpha_1^j + G(\hat{\alpha}_1 - \alpha_1) \tag{125}$$

where $G = \frac{\partial \alpha_1^j}{\partial \alpha_1}$ is evaluated at $\alpha_1 = \hat{\alpha}_1$. Since we know that $\hat{\alpha}_1 \sim \mathsf{N}\left(\alpha_1, \sigma^2_{\alpha_1}\right)$, we would have that, approximately,

$$\hat{\alpha}_1^j \sim \mathsf{N}\left(\alpha_1^j, G^2 \sigma^2_{\alpha_1}\right) = \mathsf{N}\left(\alpha_1^j, \left[j\alpha_1^{j-1}\right]^2 \sigma^2_{\alpha_1}\right), \tag{126}$$

where $\sigma^2_{\alpha_1}$ is the variance of $\hat{\alpha}_1$ as before. In large samples the higher terms neglected in the first order approximation in (125) are dominated by the linear part, although there is no certainty that this is so in typical sample sizes. The result that is used above in (126) is known as the Delta Method as stated in Lecture 1 and defined here again for convenience.

Definition 2 (Delta Method): *Let $\hat{\theta}$ be an asymptotically normally distributed estimator of θ, such that*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma_\theta^2). \tag{127}$$

Let $g(\theta)$ be a continuously and differentiable function at θ and let $g'(\theta) = \frac{\partial g(\theta)}{\partial \theta}$. Then the Delta method yields

$$\sqrt{T}[g(\hat{\theta}) - g(\theta)] \xrightarrow{d} N(0, g'(\theta)^2 \sigma_\theta^2).$$

General formulae using this approach to get the standard errors of the IRFs are given in Lütkepohl (2005) (see Proposition 3.6. on pages $110 - 112$). One thing to observe is that we cannot find the **joint distribution** of all estimated impulse response coefficients, since it must be **singular** because from (125) each estimated impulse response is a linear combination of a *single* random variable $(\hat{\alpha}_1 - \alpha_1)$. In general, they will be functions of as many parameters as are estimated in the VAR and therefore one can only find joint distributions of impulse response functions up to that number by this method. This is a theoretical restriction. The more important practical issue is that the asymptotic standard errors can perform very poorly, so let us look at a few re-sampling techniques that are commonly employed to get '*good/reasonable*' estimates of standard errors and confidence intervals.

5.6.2. Bootstrapped methods for IRFs

As outlined before, it may be inappropriate to use asymptotic standard errors to get confidence intervals for IRFs. There are a number of different ways that one can simulate the uncertainties regarding impulse responses. Two common approaches are parametric and non-parametric bootstraps. The methods are similar in the recursions of how to get the simulated values of the VAR parameters and thus the IRFs. But they differ in regards to the assumption about the residual vector U_t. Let us briefly first look at how to get the bootstrapped sequences of $A(L)$ and Σ in general. The steps are as follows:

1) Estimate the VAR(p) in (71) and obtain the fitted values of $\hat{A}(L)$ and $\hat{\Sigma}$
2) Draw random vector U_t^*.
3) Generate a new pseudo VAR(p) sequence X_t form the $\hat{A}(L)$ and $\hat{\Sigma}$ and the U_t^* vector from 2.
4) Get the set of i^{th} bootstrap estimates $\hat{A}^{(i)}(L)$ and $\hat{\Sigma}^{(i)}$ and compute the corresponding IRFs.

5) Repeat this many times, say a total number of $S = 10\,000$ times.

So the only question that remains is where the U_t^* come from? The parametric bootstrap assumes that U_t^* is a multivariate normal random variable with mean 0 and variance $\hat{\Sigma}$. So one can simply just generate draws from such a random variable in the set up above and then build up the recursions. If one believes that there are particular features in the fitted $\hat{U}_t$ that one wants to preserve in the bootstrap, then one can draw from the $\hat{U}_t$ directly. How this is done depends on the properties of $\hat{U}_t$. If they are *i.i.d.*, then random (single) draws with replacement can be made from the pool of $\hat{U}_t$. If the $\hat{U}_t$ are not *i.i.d.*, then a block-bootstrap will need to be used which selects a block of the $\hat{U}_t$ jointly from the pool and then puts them back. How to decide on the length of the block that is drawn at once depends on the data and there are methods for choosing the block length optimally. One of the earlier approaches is discussed in Politis and Romano (1994) and there exist refinements to that.

Once the bootstrapped IRFs are generated, there are again a few different ways to compute the confidence intervals. Let $\widehat{\text{irf}}$ and $\{\widehat{\text{irf}}^{(i)}\}_{i=1}^{S}$ be respectively the estimates of the IRFs from the VAR and the sequence of B bootstrapped IRFs from the simulation above. The confidence intervals are then constructed according to the following:

1) **Standard percentile method:** Use the sequence $\{\widehat{\text{irf}}^{(i)}\}_{i=1}^{S}$ to find the say 5^{th} and 95^{th} percentiles.

2) **Hall's studentised interval method:** Compute the sequence of

$$t - \text{ratios}^{(i)} = \left(\widehat{\text{irf}}^{(i)} - \widehat{\text{irf}}\right) / \sqrt{\text{Var}(\widehat{\text{irf}})} \tag{128}$$

where $\text{Var}(\widehat{\text{irf}})$ is computed as

$$\text{Var}(\widehat{\text{irf}}) = S^{-1} \sum_{i=1}^{S} \left(\widehat{\text{irf}}^{(i)} - \overline{\text{irf}}\right)^2 \tag{129}$$

with $\overline{\text{irf}} = S^{-1} \sum_{i=1}^{S} \widehat{\text{irf}}^{(i)}$ and then combine the values corresponding to the say 5^{th} and 95^{th} percentiles of $t - \text{ratios}^{(i)}$ and $\sqrt{\text{Var}(\widehat{\text{irf}})}$ to form the confidence interval as

$$\widehat{\text{irf}} - t_L^* \sqrt{\text{Var}(\widehat{\text{irf}})} \tag{130}$$

$$\widehat{\text{irf}} + t_U^* \sqrt{\text{Var}(\widehat{\text{irf}})} \tag{131}$$

where t_U^* and t_L^* are the $t-$value corresponding to the upper/lower 5^{th} and 95^{th} percentiles of the $t-$ratios$^{(i)}$ from the bootstrap sequence.

5.7. Forecast error variance decompositions

Forecast error variance decompositions give us the proportion of the variation in one variable explained by the different shocks originating in the other equations in the system. This is sometimes referred to as **innovations accounting**. Using the Cholesky decomposition (or the more general structural restrictions involving B_0 discussed before), we can write (48) as

$$\begin{aligned}\mathcal{E}_{t+h} &= \sum_{i=0}^{h-1}\Psi_i U_{t+h-i} \\ &= \sum_{i=0}^{h-1}\underbrace{\Psi_i P}_{\Phi_i}\underbrace{P^{-1}U_{t+h-i}}_{W_{t+h-i}} \\ &= \sum_{i=0}^{h-1}\Phi_i W_{t+h-i} \end{aligned} \tag{132}$$

similar to before when we derived the IRFs, with $\text{Var}(W_t) = I_k$.

The proportion of the $h-$step ahead forecast error variance (FEV) of variable j (x_j) explained by innovations in the ℓ^{th} equation (w_ℓ) is computed as

$$\omega_{j\ell,h} = \sum_{i=0}^{h-1}(e_j'\Phi_i e_\ell)^2 \Big/ \sum_{i=0}^{h-1}\Phi_i\Phi_i' \tag{133}$$

where e_j is a selection vector corresponding to the j^{th} column of the $k \times k$ identity matrix I_k. For example, we have that $\sum_{i=0}^{h-1}(e_j'\Phi_i e_\ell)^2 = \phi_{j\ell,0}^2 + \phi_{j\ell,1}^2 + \ldots + \phi_{j\ell,h-1}^2$ with j and ℓ being the row and column indices of Φ and $\phi_{j\ell,h}$ the corresponding elements of Φ_h. The FEV can thus be seen as a rescaled version of the accumulation of the squared IRFs of the variable of interest, where the scaling is equal to the h period mean squared forecast error

$$\begin{aligned}\Sigma_u(h) &= \sum_{i=0}^{h-1}\Psi_i\Sigma\Psi_i' \\ &= \sum_{i=0}^{h-1}\Phi_i\Phi_i'. \end{aligned}$$

References

Bernanke, Ben S. (1986): "Alternative explanations of the money-income correlation," *Carnegie-Rochester Conference Series on Public Policy*, **25**(1), 49–99.

Blanchard, Olivier Jean and Danny Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *The American Economic Review*, **79**(4), 655–673.

Blanchard, Olivier J. and Mark W. Watson (1987): "Are Business Cycles All Alike?" *NBER Working Paper No. 1382*, National Bureau of Economic Research.

Cooley, Thomas F. and Stephen F. Leroy (1985): "Atheoretical macroeconometrics: A critique," *Journal of Monetary Economics*, **16**(3), 283–308.

Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent and Mark W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, **97**(3), 1021–1026.

Giordani, Paolo (2004): "Evaluating New-Keynesan Models of a Small Open Economy," *Oxford Bulletin of Economics and Statistics*, **66**(Supplement), 713–733.

Hamilton, James D. (1994): *Time Series Analysis*, Princeton University Press.

Ivanov, Ventzislav and Lutz Kilian (2005): "A practitioner's guide to lag order selection for VAR impulse response analysis," *Studies in Nonlinear Dynamics and Econometrics*, **9**(1), Article 2.

Kilian, Lutz (1998): "Small-Sample Confidence Intervals For Impulse Response Functions," *The Review of Economics and Statistics*, **80**(2), 218–230.

———— (1999): "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses," *The Review of Economics and Statistics*, **81**(4), 652–660.

Levtchenkova, Sofia, Adrian R. Pagan and John C. Robertson (1998): "Shocking Stories," *Journal of Economic Surveys*, **12**(5), 507–532.

Lütkepohl, Helmut (2005): *New introduction to multiple time series analysis*, Springer.

Politis, Dimitris N. and Joseph P. Romano (1994): "The Stationary Bootstrap," *Journal of the American Statistical Association*, **89**(428), 1303–1313.

Quenouille, Maurice H. (1957): *The Analysis of Multiple Time Series*, Statistical Monographs and Course No. 1, Griffin, London.

Sims, Christopher A. (1980): "Macroeconomics and Reality," *Econometrica*, **48**(1), 1–48.

———— (1986): "Are Forecasting Models Usable for Policy Analysis?" *Federal Reserve Bank of Minneapolis Quarterly Review*, **10**(1), 2–16.

Wallis, Kenneth F (1977): "Multiple Time Series Analysis and the Final Form of Econometric Models," *Econometrica*, **45**(6), 1481–97.

Zellner, Arnold (1962): "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests forAggregation Bias," *Journal of the American Statistical Association*, **57**(298), 348–368.

Zellner, Arnold and Franz Palm (1974): "Time series analysis and simultaneous equation econometric models," *Journal of Econometrics*, **2**(1), 17–54.