Chapter 4: VAR Models

This chapter describes a set of techniques which stand apart from those considered in the next three chapters, in the sense that economic theory is only minimally used in the inferential process. VAR models, pioneered by Chris Sims about 25 years ago, have acquired a permanent place in the toolkit of applied macroeconomists both to summarize the information contained in the data and to conduct certain types of policy experiments. VAR are well suited for the first purpose: the Wold theorem insures that any vector of time series has a VAR representation under mild regularity conditions and this makes them the natural starting point for empirical analyses. We discuss the Wold theorem, and the issues connected with non-uniqueness, non-fundamentalness and non-orthogonality of the innovation vector in the first section. The Wold theorem is generic but imposes important restrictions; for example, the lag length of the model should go to infinity for the approximation to be "good". Section 2 deals with specification issues, describes methods to verify some of the restrictions imposed by the Wold theorem and to test other related implications (e.g. white noise residuals, linearity, stability, etc.). Section 3 presents alternative formulations of a VAR(q). These are useful when computing moments or spectral densities, and in deriving estimators for the parameters and for the covariance matrix of the shocks. Section 4 presents statistics commonly used to summarize the informational content of VARs and methods to compute their standard errors. Here we also discuss generalized impulse response functions, which are useful in dealing with time varying coefficients VAR models analyzed in chapter 10. Section 5 deals with identification, i.e with the process of transforming the information content of reduced form dynamics into structural ones. Up to this point economic theory has played no role. However, to give a structural interpretation to the estimated relationships, economic theory needs to be used. Contrary to what we will be doing in the next three chapters, only a minimalist set of restrictions, loosely related to the classes of models presented in chapter 2, are employed to obtain structural relationships. We describe identification methods which rely on conventional short run, on long run and on a sign restrictions. In the latter two cases (weak) restrictions derived from DSGE models are employed and the structural link between the theory and the data explicitly made. Section 6 describes problems which may distort the interpretation of structural VAR results. Time aggregation, omission of variables and shocks and non-fundamentalness should always be in the back of the mind of applied researchers when conducting policy analyses with VAR. Section 7 proposes a way to validate a class of DSGE models using structural

VARs. Log-linearized DSGE models have a restricted VAR representation. When a researcher is confident in the theory, a set of quantitative restrictions can be considered, in which case the methods described in chapters 5 to 7 could be used. When theory only provides qualitative implications or when its exact details are doubtful, one can still validate a model conditioning on its qualitative implications. Since DSGE models provide a wealth of robust sign restrictions, one can take the ideas of section 5 one step further, and use them to identify structural shocks. Model evaluation then consists in examining the qualitative (and quantitative) features of the dynamic responses to identified structural shocks. In this sense, VAR identified with sign restrictions offer a natural setting to validate incompletely specified (and possibly false) DSGE models.

4.1 The Wold theorem

The use of VAR models can be justified in many ways. Here we employ the Wold representation theorem as major building block. While the theory of Hilbert spaces is needed to make the arguments sound, we keep the presentation simple and invite the reader to consult Rozanov (1967) or Brockwell and Davis (1991) for precise statements.

The Wold theorem decomposes any $m\times 1$ vector stochastic process $y_t^\dagger$ into two orthogonal components: one linearly predictable and one linearly unpredictable (linearly regular). To show what the theorem involves let $\mathcal{F}_t$ be the time t information set; $\mathcal{F}_t = \mathcal{F}_{t-1} \oplus \mathcal{E}_t$, where $\mathcal{F}_{t-1}$ contains time $t-1$ information and $\mathcal{E}_t$ the news at t. Here $\mathcal{E}_t$ is orthogonal to $\mathcal{F}_{t-1}$ (written $\mathcal{E}_t \perp \mathcal{F}_{t-1}$) and $\oplus$ indicates direct sum, that is $\mathcal{F}_t = \{y_{t-1}^\dagger + e_t, y_{t-1}^\dagger \in \mathcal{F}_{t-1},\ e_t \in \mathcal{E}_t\}$.

Exercise 4.1 *Show that $\mathcal{E}_t \perp \mathcal{F}_{t-1}$ implies $\mathcal{E}_t \perp \mathcal{E}_{t-1}$ so that $\mathcal{E}_{t-j}$ is orthogonal to $\mathcal{E}_{t-j'}$, $j' < j$.*

Since the decomposition of $\mathcal{F}_t$ can be repeated for each t, iterating backwards we have

$$\mathcal{F}_t = \mathcal{F}_{t-1} \oplus \mathcal{E}_t = \ldots = \mathcal{F}_{-\infty} \oplus \sum_{j=0}^{\infty} \mathcal{E}_{t-j} \tag{4.1}$$

where $\mathcal{F}_{-\infty} = \bigcap_j \mathcal{F}_{t-j}$. Since $y_t^\dagger$ is known at time t (this condition is sometimes referred as adaptability of $y_t^\dagger$ to $\mathcal{F}_t$), we can write $y_t^\dagger \equiv E[y_t^\dagger | \mathcal{F}_t]$ where $E[.|\mathcal{F}_t]$ is the conditional expectations operator. Orthogonality of the news with past information then implies:

$$y_t^\dagger = E[y_t^\dagger|\mathcal{F}_t] = E[y_t^\dagger|\mathcal{F}_{-\infty} \oplus \sum_j \mathcal{E}_{t-j}] = E[y_t^\dagger|\mathcal{F}_{-\infty}] + \sum_{j=0}^{\infty} E[y_t^\dagger|\mathcal{E}_{t-j}] \tag{4.2}$$

We make two assumptions. First, we consider linear representations, that is, we substitute the expectations operator with a linear projection operator. Then (4.2) becomes

$$y_t^\dagger = a_t y_{-\infty} + \sum_{j=0}^{\infty} D_{jt} e_{t-j} \tag{4.3}$$

where $e_{t-j} \in \mathcal{E}_{t-j}$ and $y_{-\infty} \in \mathcal{F}_{-\infty}$. The sequence $\{e_t\}_{t=0}^{\infty}$, defined by $e_t = y_t^\dagger - E[y_t^\dagger|\mathcal{F}_{t-1}]$, is a white noise process (i.e. $E(e_t) = 0$; $E(e_t e_{t-j}') = \Sigma_t$ if $j = 0$ and zero otherwise). Second, we assume that $a_t = a$; $D_{jt} = D_j$; $\forall t$. This implies

$$y_t^\dagger = a y_{-\infty} + \sum_{j=0}^{\infty} D_j e_{t-j} \tag{4.4}$$

Exercise 4.2 *Show that if* $y_t^\dagger$ *is covariance stationary,* $a_t = a,\ D_{jt} = D_j$.

The term $ay_{-\infty}$ on the right hand side of (4.4) is the linearly deterministic component of $y_t^\dagger$ and can be perfectly predicted given the infinite past. The term $\sum_j D_j e_{t-j}$ is the linearly regular component, that is, the component produced by the news at each t. We say that $y_t^\dagger$ is deterministic if and only if $y_t^\dagger \in \mathcal{F}_{-\infty}$ and regular if and only if $\mathcal{F}_{-\infty} = \{0\}$.

Three important points need to be highlighted. First, for (4.2) to hold, no assumptions about $y_t^\dagger$ are required: we only need that new information is orthogonal to the existing one. Second, both linearity and stationary are unnecessary for the theorem to hold. For example, if stationarity is not assumed there will still be a linearly regular and a linearly deterministic component even though each will have time varying coefficients (see (4.3)). Third, if we insist on requiring covariance stationary, preliminary transformations of $y_t^\dagger$ may be needed to produce the representation (4.4).

The Wold theorem is a powerful tool but is too generic to guide empirical analysis. To impose some more structure, we assume first that the data is a mean zero process, possibly after deseasonalization (with deterministic periodic functions), removal of constants, etc. and let $y_t = y^\dagger - ay_{-\infty}$. Using the lag operator we write $\sum_{j=0}^{\infty} D_j e_{t-j} = \sum_j D_j \ell^j e_t = D(\ell) e_t$ so that $y_t = D(\ell)e_t$ is the MA representation for y_t where D_j is a $m \times m$ matrix of rank m, for each j. MA representations are not unique: in fact, for any nonsingular matrix $\mathcal{H}(\ell)$ satisfying $\mathcal{H}(\ell)\mathcal{H}(\ell^{-1})' = I$ such that $\mathcal{H}(z)$ has no singularities for $|z| \leq 1$, where $\mathcal{H}(\ell^{-1})'$ is the transpose (and possibly complex conjugate) of $\mathcal{H}(\ell)$, we can write $y_t = \tilde{D}(\ell)\tilde{e}_t$ with $\tilde{D}(\ell) = D(\ell)\mathcal{H}(\ell)$, $\tilde{e}_t = \mathcal{H}(\ell^{-1})' e_t$.

Exercise 4.3 *Show that* $E(\tilde{e}_t \tilde{e}_{t-j}') = E(e_t e_{t-j}')$. *Conclude that if* e_t *is covariance stationary, the two representation produce equivalent autocovariance functions for* y_t.

Matrices like $\mathcal{H}(\ell)$ are called Blaschke factors and are of the form $\mathcal{H}(\ell) = \prod_{i=1}^m \varrho_i \mathcal{H}^\dagger(d_i, \ell)$ where d_i are the roots of $D(\ell)$, $|d_i| < 1$, $\varrho_i \varrho_i' = I$ and, for each i, $\mathcal{H}^\dagger(d_i, \ell)$ is given by:

$$\mathcal{H}^\dagger(d_i, \ell) = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \frac{\ell - d_i}{1 - d_i^{-1}\ell} & \ldots & 0 \\ 0 & 0 & \ldots & 1 \end{bmatrix} \tag{4.5}$$

Exercise 4.4 *Suppose* $\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} (1+4\ell) & 0 \\ 0 & (1+10\ell) \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$. *Find the Blaashke factors of* $D(\ell)$. *Construct two alternative moving average representations for* y_t.

Example 4.1 *Consider* $y_{1t} = e_t - 0.5e_{t-1}$ *and* $y_{2t} = \tilde{e}_t - 2\tilde{e}_{t-1}$. *It is easy to verify that the roots of* $D(z)$ *are 2 in the first case, and 0.5 in the second. Since the roots are one the inverse of the other, the two processes span the same information space as long as the variance of innovations is appropriately adjusted. In fact, using the covariance generating function to have* $CGF_{y_1}(z) = (1-0.5z)(1-0.5z^{-1})\sigma_1^2$ *and* $CGF_{y_2}(z) = (1-2z)(1-2z^{-1})\sigma_2^2 = (1-0.5z)(1-0.5z^{-1})(4\sigma_2^2)$. *Hence, if* $\sigma_1^2 = 4\sigma_2^2$ *the CGF of the two processes is the same.*

Exercise 4.5 *Let* $y_{1t} = e_t - 4e_{t-1},\ e_t \sim (0, \sigma^2)$. *Set* $y_{2t} = (1-0.25\ell)^{-1}y_{1t}$. *Show that the CGF(z) of* y_{2t} *is a constant for all z. Show that* $y_{2t} = \tilde{e}_t - 0.25\tilde{e}_{t-1}$ *where* $\tilde{e}_t \sim (0, 16\sigma^2)$ *is equivalent to* y_{1t} *in terms of the covariance generating function.*

Among the class of equivalent MA representations, it is typical to choose the "fundamental" one. The following two definitions are equivalent.

Definition 4.1 *(Fundamentalness)*
1) A MA is fundamental if $det(D_0 E(e_t e_t')D_0') > det(D_j E(e_{t-j}e_{t-j}')D_j'),\ \forall j \neq 0$.
2) A MA is fundamental if the roots of $D(z)$ *are all greater than one in modulus.*

The roots of $D(z)$ are related to the eigenvalues of the companion matrix of the system (see section 3). Fundamental representations, also termed Wold representations, could also be identified by the requirement that the completion of the space spanned by linear combinations of the y_t's has the same information as the completion of the space spanned by linear combinations of e_t's. In this sense Wold representations are invertible: knowing y_t is the same as knowing e_t.

As it is shown in the next example, construction of a fundamental representation requires "flipping" all roots that are less than one in absolute value.

Example 4.2 *Suppose* $y_t = \begin{bmatrix} 1.0 & 0 \\ 0.2 & 0.9 \end{bmatrix} e_t + \begin{bmatrix} 2.0 & 0 \\ 0 & 0.7 \end{bmatrix} e_{t-1}$ *where* $e_t \sim$ *iid* $(0, I)$. *Here* $det(D_0) = 0.9 < det(D_1) = 1.4$ *so the representation is not fundamental. To find a fundamental one we compute the roots of* $D_0 + D_1 z = 0$; *their absolute values are 0.5 and 1.26 (these are the diagonal elements of* $-D_1^{-1}D_0$*). The problematic root is 0.5 which we flip to 1.0/0.5=2.0. The fundamental MA is then* $y_t = \begin{bmatrix} 1.0 & 0 \\ 0.2 & 0.9 \end{bmatrix} e_t + \begin{bmatrix} 0.5 & 0 \\ 0 & 0.7 \end{bmatrix} e_{t-1}$.

Exercise 4.6 *Determine which of the following polynomial produces fundamental representations when applied to a white noise innovation:(i)* $D(\ell) = 1 + 2\ell + 3\ell^2 + 4\ell^3$, *(ii)* $D(\ell) = 1 + 2\ell + 3\ell^2 + 2\ell^3 + \ell^4$, *(iii)* $D(\ell) = I + \begin{bmatrix} .8 & -.7 \\ .7 & .8 \end{bmatrix} \ell$, *(iv)* $D(\ell) = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix} \ell + \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \ell^2$.

Exercise 4.7 *Show that* $y_t = e_t + \begin{bmatrix} 1.0 & 0 \\ 0 & 0.8 \end{bmatrix} e_{t-1}$ *where* $var(e_t) = \begin{bmatrix} 2.0 & 1.0 \\ 1.0 & 1.0 \end{bmatrix}$ *and* $y_t = e_t + \begin{bmatrix} 0.9091 & 0.1909 \\ 0 & 0.8 \end{bmatrix} e_{t-1}$ *where* $var(e_t) = \begin{bmatrix} 2.21 & 1.0 \\ 1.0 & 1.0 \end{bmatrix}$ *generate the same ACF for* y_t. *Which representation is fundamental?*

Exercise 4.8 *Let* $\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} (1+4\ell) & 1+0.5\ell \\ 0 & (1+5\ell) \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}$ *where* $e_t = (e_{1t}, e_{2t})$ *has unitary variance. Is the space spanned by linear combinations of the* y_t *and* e_t *the same? If the MA is not fundamental, find a fundamental one.*

While it is typical to use Wold representations in applied work, there are economic models that do not generate a fundamental format. Two are presented in the next examples.

Example 4.3 *Consider a RBC model where households maximize* $E_0 \sum_t \beta^t(\ln(c_t) - \vartheta_N N_t)$ *subject to* $c_t + inv_t \leq GDP_t;\ K_{t+1} = (1-\delta)K_t + inv_t;\ c_t \geq 0;\ inv_t \geq 0; 0 \leq N_t \leq 1$ *where* $0 < \beta < 1$ *and* δ, ϑ_n *are parameters and assume that the production function is* $GDP_t = k_t^{1-\eta} N_t^{\eta} \zeta_t$ *where* $\ln \zeta_t = \ln \zeta_{t-1} + 0.1\epsilon_{1t} + 0.2\epsilon_{1t-1} + 0.4\epsilon_{1t-2} + 0.2\epsilon_{1t-3} + 0.1\epsilon_{1t-4}$. *Such a diffusion of technological innovations is appropriate when e.g., only the most advanced sector employs the new technology (say, a new computer chips) and it takes some time for the innovation to spread to the economy. If* $\epsilon_{1t} = 1,\ \epsilon_{1t+\tau} = 0, \forall \tau \neq 0$ ζ_t *looks like in figure 1. Clearly, a process with this shape does not satisfy the restrictions given in definition 4.1.*

Example 4.4 *Consider a model where fiscal shocks drive economic fluctuations. Typically, fiscal policy changes take time to have effects: between the programming, the legislation and the implementation of, say, a change in income tax rates several months may elapse. If agents are rational they may react to tax changes before the policy is implemented and, conversely, no behavioral changes may be visible when the changes actually take place. Since the information contained in tax changes may have a different timing than the information contained, say, in the income process, fiscal shocks may produce non-Wold representations.*

Whenever economic theory requires non-fundamental MAs, one could use Blaaske factors to flip the representations provided by standard packages, as e.g. in Lippi and Reichlin (1994). In what follows we will consider only fundamental structures and take $y_t = D(\ell)e_t$ be such a representation.

The "innovations" e_t play an important role in VAR analyses. Since $E(e_t|\mathcal{F}_{t-1}) = 0$ and $E(e_t e_t'|\mathcal{F}_{t-1}) = \Sigma_e$, e_t are serially uncorrelated but contemporaneously correlated. This means that we cannot attach a "name" to the disturbances. To do so we need an orthogonal representation for the innovations. Let Σ_e be the covariance matrix of e_t, let $\Sigma_e = \mathcal{P}\mathcal{V}\mathcal{P}' = \tilde{\mathcal{P}}\tilde{\mathcal{P}}'$ where $\mathcal{V}$ is a diagonal matrix and $\tilde{\mathcal{P}} = \mathcal{P}\mathcal{V}^{0.5}$. Then $y_t = D(\ell)e_t$ is equivalent to

$$y_t = \tilde{D}(\ell)\tilde{e}_t \tag{4.6}$$

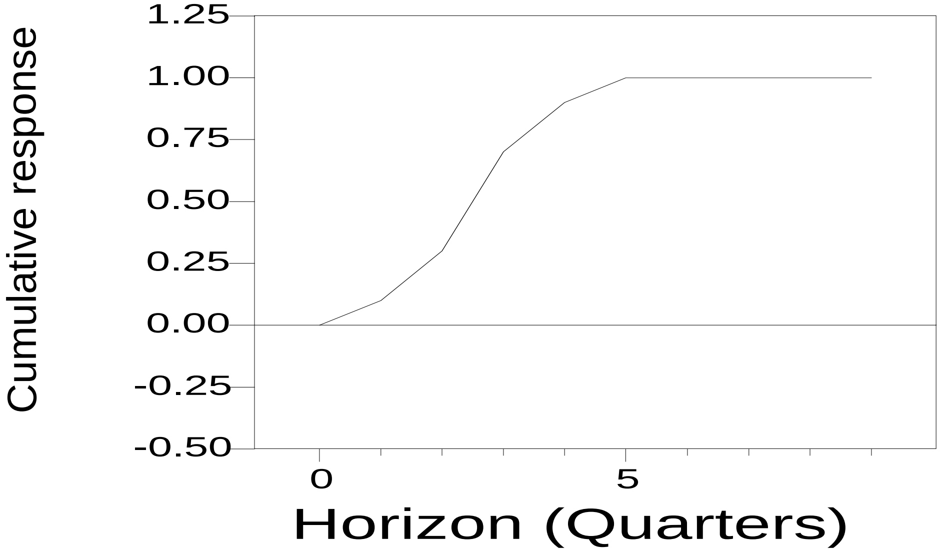

Figure 4.1: Non fundamental technological progress

for $\tilde{D}(\ell) = D(\ell)\tilde{\mathcal{P}}$ and $\tilde{e}_t = \tilde{\mathcal{P}}^{-1}e_t$. There are many ways of generating (4.6). One is a Choleski factorization, i.e. $\mathcal{V} = I$ and $\mathcal{P}$ is a lower triangular matrix. Another is obtained when $\mathcal{P}$ contains the eigenvectors and $\mathcal{V}$ the eigenvalues of Σ_e.

Example 4.5 *If e_t is a 2×1 vector with correlated entries, orthogonal innovations are $\tilde{e}_{1t} = e_{1t} - be_{2t}$ and $\tilde{e}_{2t} = e_{2t}$ where $b = \frac{cov(e_{1t}e_{2t})}{var(e_{2t})}$ and $var(\tilde{e}_{1t}) = \sigma_1^2 - b^2\sigma_2^2$, $var(\tilde{e}_{2t}) = \sigma_2^2$.*

It is important to stress that orthogonalization devices are void of economic content: they only transform the MA representation in a form which is more useful when tracing out the effect of a particular shock. To attach economic interpretations to the representation, these orthogonalizations ought to be linked to economic theory. Note also that while with the Choleski decomposition $\mathcal{P}$ has zero restrictions placed on the upper triangular part, no such restrictions are present when an eigenvalue-eigenvector decomposition is performed.

As mentioned, when the polynomial $D(z)$ has all its roots greater than one in modulus (and this condition holds if, e.g., $\sum_{j=0}^{\infty} D_j^2 < \infty$ (see Rozanov (1967)) the MA representation is invertible and we can express e_t as a linear combination of current and past y_t's, i.e. $[A_0 - A(\ell)]y_t = e_t$ where $[A_0 - A(\ell)] = (D(\ell))^{-1}$. Moving lagged y_t's on the right hand side and setting $A_0 = I$ a vector autoregressive (VAR) representation is obtained

$$y_t = A(\ell)y_{t-1} + e_t \tag{4.7}$$

In general, $A(\ell)$ will be of infinite length for any reasonable specification of $D(\ell)$.

There is an important relationship between the concept of invertibility and the one of stability of the system which we highlight next.

Definition 4.2 *(Stability) A VAR(1) is stable if* $det(I_m - Az) \neq 0, \forall |z| \leq 1$ *and a VAR(q) is stable if* $det(I_m - A_1 z - \ldots - A_q z^q) \neq 0 \ \forall |z| \leq 1$.

Definition 4.2 implies that all eigenvalues of A have modulus less or equal than 1 (or that the matrix A has no roots inside or on the complex unit circle). Hence, if y_t has an invertible MA representation, it also has a stable VAR structure. Therefore, one could start from stable processes to motivate VAR analyses (as, e.g. it is done in Lutkepohl (1991)). Our derivation shows the primitive restrictions needed to obtain stable VARs.

Example 4.6 *Suppose* $y_t = \begin{bmatrix} 0.5 & 0.1 \\ 0.0 & 0.2 \end{bmatrix} y_{t-1} + e_t$. *Here* $det(I_2 - Az) = (1 - 0.5z)(1 - 0.2z) = 0$ *and* $|z_1| = 2 > 1, |z_2| = 5 > 1$. *Hence, the system is stable.*

Exercise 4.9 *Check if* $y_t = \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.2 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.6 \end{bmatrix} y_{t-2} + e_t$ *is stable or not.*

To summarize, any vector of time series can be represented with a constant coefficient VAR(∞) under linearity, stationarity and invertibility. Hence, one can interchangeably think of data or the VAR for the data. Also, with a finite stretch of data only a VAR(q), q finite, can be used. For a VAR(q) to approximate any y_t sufficiently well, we need D_j to converge to zero repidly as j increases.

Exercise 4.10 *Consider* $y_t = e_t + 0.9e_{t-1}$ *and* $y_t = e_t + 0.3e_{t-1}$. *Compute the AR representations. What lag length is needed to approximate the two processes? What if* $y_t = e_t + e_{t-1}$*?*

Two concepts which are of some use are in applied work are those of Granger non-causality and Sims (econometric) exogeneity. It is important to stress that they refer to the ability of one variable to predict another one and do not imply any sort of economic causality (e.g. the government takes an action, the exchange rate will move). Let (y_{1t}, y_{2t}) be a partition of a covariance stationary y_t with fundamental innovations e_{1t} and e_{2t}; let Σ_e be diagonal and let $D_{i,i'}(\ell)$ be the i, i' block of $D(\ell)$.

Definition 4.3 *(Granger causality)* y_{2t} *fails to Granger cause* y_{1t} *if and only if* $D_{12}(\ell) = 0$.

Definition 4.4 *(Sims Exogeneity) We can write* $y_{2t} = \mathsf{Q}(\ell) y_{1t} + \epsilon_{2t}$ *with* $E_t[\epsilon_{2t} y_{1t-\tau}] = 0, \ \forall \tau \geq 0$ *and* $\mathsf{Q}(\ell) = \mathsf{Q}_0 + \mathsf{Q}_1 \ell + \ldots$ *if and only if* y_{2t} *fails to Granger cause* y_{1t} *and* $D_{21}(\ell) \neq 0$.

Exercise 4.11 *Show what Granger non-causality of* y_{2t} *for* y_{1t} *implies in a trivariate VAR.*

We conclude examining cases where the data deviates from the setup considered so far.

Exercise 4.12 *(i) Suppose that* $y_t = D(\ell)e_t$ *where* $D(\ell) = (1-\ell)D^{\dagger}(\ell)$*. Derive a VAR for* y_t*. Show that if* $D^{\dagger}(\ell) = 1$*, there is no convergent VAR representation for* y_t*.*
(ii) Suppose that $y_t^{\dagger} = a_0 + a_1 t + D(\ell)e_t$ *if* $t \leq \bar{T}$ *and* $y_t^{\dagger} = a_0 + a_2 t + D(\ell)e_t$ *if* $t > \bar{T}$*. How would you derive a VAR representation for* y_t*?*
(iii) Suppose that $y_t = D(\ell)e_t$ *and* $var(e_t) \propto y_{t-1}^2$ *Find a VAR for* y_t*.*
(iv) Suppose that $y_t = D(\ell)e_t$*,* $var(e_t) = b\ var(e_{t-1}) + \sigma^2$*. Find a VAR for* y_t*.*

4.2 Specification

In section 4.1 we showed that a constant coefficient VAR is a good approximation to any vector of time series. Here we examine how to verify the restrictions needed for the approximation to hold. The model we consider is (4.7) where $A(\ell) = A_1\ell + \ldots + A_q\ell^q$, y_t is a $m \times 1$ vector, and $e_t \sim (0, \Sigma_e)$. VARs with econometrically exogenous variables can be obtained via restrictions on $A(\ell)$ as indicated in definition 4.4. We let $\mathsf{A}_1' = (A_1', \ldots, A_q')'$ be a $(mq \times m)$ matrix and set $\alpha = vec(\mathsf{A}_1)$ where $vec(\mathsf{A}_1)$ stacks the columns of A_1 (so α is a $m^2q \times 1$ vector).

4.2.1 Lag Length 1

There are several methods to select the lag length of a VAR. The simplest is based on a likelihood ratio (LR) test. . Here the model with a smaller number of lags is treated as a restricted version of a larger dimensional model. Since the two models are nested, under the null that the restricted model is correct, differences in the likelihoods should be small. Let $R(\alpha) = 0$ be a set of restrictions and $\mathcal{L}(\alpha, \Sigma_e)$ the likelihood function. Then:

$$LR = 2[\ln \mathcal{L}(\alpha^{un}, \Sigma_e^{un}) - \ln \mathcal{L}(\alpha^{re}, \Sigma_e^{re})] \tag{4.8}$$

$$= (R(\alpha^{un}))'[\frac{\partial R}{\partial \alpha^{un}}(\Sigma_e^{re} \otimes (X'X)^{-1})(\frac{\partial R}{\partial \alpha^{un}})']^{-1}(R(\alpha^{un})) \tag{4.9}$$

$$= T(\ln |\Sigma_e^{re}| - \ln |\Sigma_e^{un}|) \stackrel{D}{\to} \chi^2(\nu) \tag{4.10}$$

where $X_t = (y_{t-1}', \ldots, y_{t-q}')'$, and $X' = (X_0, \ldots, X_{T-1})$ is a $mq \times T$ matrix and ν the number of restrictions. (4.8)-(4.9)-(4.10) are equivalent formulations of the likelihood ratio test. The first is the standard one. (4.9) is obtained maximizing the likelihood function with respect to α subject to $R(\alpha) = 0$. (4.10) is convenient for computing actual test values and to compare LR results with those of other testing procedures.

Exercise 4.13 *Derive (4.9) using a Lagrangian multiplier approach.*

Four important features of LR tests need to be highlighted. First, a LR test is valid when y_t is stationary and ergodic and if the residuals are white noise under the null. Second, it can be computed without explicit distributional assumptions on the y_t's. What is required is that e_t is a sequence of independent white noises with bounded fourth moments and that T is sufficiently large - in which case $\alpha^{un}, \Sigma_e^{un}, \alpha^{re}, \Sigma_e^{re}$ are pseudo maximum likelihood estimators. Third, a likelihood ratio test is biased against the null in small samples. Hence it is common to use $LR^c = (T - qm)(\ln|\Sigma^{re}| - \ln|\Sigma^{un}|)$ where qm is the number of estimated parameters in each equation of the unrestricted system. Finally, one should remember that the distribution of the LR test is only asymptotically valid. That is, significance levels only approximate probabilities of Type I errors.

In practice, an estimate of q is obtained sequentially as the next algorithm shows:

Algorithm 4.1

1) *Choose an upper bound* $\bar{q}$.

2) *Test a VAR*($\bar{q}-1$) *against VAR*($\bar{q}$) *using a LR test. If the null hypothesis is not rejected*

3) *Test a VAR*($\bar{q}-2$) *against VAR*($\bar{q}-1$) *using an LR test. Continue until rejection.*

Clearly, $\bar{q}$ depends on the frequency of the data. For annual data $\bar{q} = 3$; for quarterly data $\bar{q} = 8$; and for monthly data $\bar{q} = 18$ are typical choices. Note that with a sequential approach each null hypothesis is tested conditional on all the previous ones being true and that the chosen q crucially depends on the significance level. Furthermore, when a sequential procedure is used it is important to distinguish between the significance level of individual tests and the significance level of the procedure as a whole - in fact, rejection of a VAR($\bar{q} - j$) implies that all VAR($\bar{q} - j'$) will also be rejected, $\forall j' > j$.

Example 4.7 *Choose as a significance level* 0.05 *and set* $\bar{q} = 6$. *Then a likelihood ratio test for q=5 vs. q=6 has significance level* $1 - 0.95 = 0.05$. *Conditional on choosing q=5, a test for q=4 vs. q=5 has a significance level* $1 - (0.95)^2 = 0.17$ *and the significance level at the* $j - th$ *stage is* $1 - (1 - .05)^j$. *Hence, if we expect the model to have three or four lags, we better adjust the significance level so that at the second or third stage of the testing, the significance is around* 0.05.

Exercise 4.14 *A LR test restricts each equation to have the same number of lags. Is it possible to choose different lag lengths in different equations? How would you do this in a bivariate VAR?*

While popular, LR tests are unsatisfactory lag selection approaches when the VAR is used for forecasting. This is because LR tests look at the in-sample fit of models (see equation 4.10). When forecasting one would like to have lag selection methods which minimize the (out-of-sample) forecast error. Let $y_{t+\tau} - y_t(\tau)$ be the τ-step ahead forecast error based on time t information and let $\Sigma_y(\tau) = E[y_{t+\tau} - y_t(\tau)][y_{t+\tau} - y_t(\tau)]'$ be its mean square error (MSE). When $\tau = 1$, $\Sigma_y(1) \approx \frac{T+mq}{T}\Sigma_e$ where Σ_e is the variance covariance matrix of the innovations (see e.g. Lutkepohl (1991, p.88)). The next three information criteria choose lag length using transformations of $\Sigma_y(1)$.

- Akaike Information criterion (AIC) : $\min_q AIC(q) = \ln|\Sigma_y(1)|(q) + \frac{2qm^2}{T}$.

- Hannan and Quinn criterion (HQC): $\min_q HQC(q) = \ln|\Sigma_y(1)|(q) + \frac{2qm^2}{T}\ln(\ln T)$.

- Schwarz criterion (SWC): $\min_q SC(q) = \ln|\Sigma_y(1)|(q) + \frac{2qm^2}{T}\ln T$.

All criteria add a penalty to the one-step ahead MSE which depends on the sample size T, the number of variables m and the number of lags q. While for large T penalty differences are unimportant, this is not the case when T is small, as shown in table 4.1.

Criterion	T=40, m=4			T=80, m=4			T=120, m=4			T=120, m=4		
	q=2	q=4	q=6	q=2	q=4	q=6	q=2	q=4	q=6	q=2	q=4	q=6
AIC	0.4	3.2	4.8	0.8	1.6	2.4	0.53	1.06	1.6	0.32	0.64	0.96
HQC	0.52	4.17	6.26	1.18	2.36	3.54	0.83	1.67	2.50	0.53	1.06	1.6
SWC	2.95	5.9	8.85	1.75	3.5	5.25	1.27	2.55	3.83	0.84	1.69	2.52

Table 4.1: Penalties of Akaike, Hannan and Quinn and Schwarz criteria

In general, for $T \geq 20$ SWC and HQC will always choose smaller models than AIC.

The three criteria have different asymptotic properties. AIC is inconsistent (in fact, it overestimates the true order with positive probability) while HQC and SWC are consistent and when $m > 1$, they are both strongly consistent (i.e. they will choose the correct model almost surely). Intuitively, AIC is inconsistent because the penalty function used does not simultaneously goes to infinity as $T \to \infty$ and to zero when scaled by T. Consistency however, it is not the only yardstick to use since consistent methods may have poor small sample properties. Ivanov and Kilian (2001) extensively study the small sample properties of these three criteria using a variety of data generating processes and data frequencies and found that HQC is best for quarterly and monthly data, both when y_t is covariance stationary and when it is a near-unit root process.

Example 4.8 *Consider a quarterly VAR model for the Euro area for the sample 1980:1-1999:4 (T=80); restrict $m = 4$ and use output, prices, interest rates and money (M3) as variables. A constant is eliminated previous to the search. We set $\bar{q} = 7$. Table 4.2 reports the sequential p-values of basic and modified LR tests (first two columns) and the values of the AIC, HQC, SWC criteria (other three columns).*

Different tests select somewhat different lag length. The LR tests select 7 lags but the p-values are non-monotonic and it matters what $\bar{q}$ is. For example, if $\bar{q} = 6$, LR^c selects two lags. Nonmonotonicity appears also for the other three criteria. In general, SWC, which uses the harshest penalty, has a minimum at 1; HQC and AIC have a minimum at 2. Based on these outcomes, we tentatively select a VAR(2).

4.2.2 Lag Length 2

The Wold theorem implies, among other things, that VAR residuals must be white noise. A LR test can therefore be interpreted as a diagnostic to check whether residuals satisfy

Hypothesis	*LR*	LR^c	*AIC*	*HQC*	*SWC*
q=6 vs. q=7	*2.9314e-05*	*0.0447*	*-7.5560*	*-6.3350*	*-4.4828*
q=5 vs. q=6	*3.6400e-04*	*0.1171*	*-7.4139*	*-6.3942*	*-4.8514*
q=4 vs. q=5	*0.0509*	*0.5833*	*-7.4940*	*-6.6758*	*-5.4378*
q=3 vs. q=4	*0.0182*	*0.4374*	*-7.5225*	*-6.9056*	*-5.9726*
q=2 vs. q=3	*0.0919*	*0.6770*	*-7.6350*	*-7.2196*	*-6.5914*
q=1 vs. q=2	*3.0242e-07*	*6.8182e-03*	*-7.2266*	*-7.0126*	*-6.6893*

Table 4.2: Lag length of a VAR

this property. Similarly, AIC, HQC and SWC can be seen as trading-off the white noise assumption on the residuals with the best possible out-of-sample forecasting performance.

Another class of tests to lag selection directly examines the properties of VAR residuals. Let $ACRF_e(\tau)^{i,i'}$ denote the cross correlation of e_{it} and $e_{i't}$ at lag $\tau = \ldots, -1, 0, 1, \ldots$. Then , under the null of white noise $ACRF_e(\tau)^{i,i'} = \frac{ACF_e(\tau)^{i,i'}}{\sqrt{ACF_e(0)^{i,i} ACF_e(0)^{i',i'}}} \rightarrow \mathsf{N}(0, \frac{1}{T})$ for each τ (see e.g. Lutkepohl (1991, p.141).

Exercise 4.15 *Design a test for the joint hypothesis that* $ACRF_e(\tau) = 0 \ \forall i, i', \tau$ *fixed.*

Care must be exercised in implementing white noise tests sequentially - say, starting from an upper $\bar{q}$, checking if the residual are white noise and, if they are, decrease $\bar{q}$ by one value at the time until the null hypothesis is rejected. Since serial correlation is present in incorrectly specified VARs, one must choose a $\bar{q}$ for which the null hypothesis is satisfied.

Exercise 4.16 *Provide a test statistic for the null that* $ACRF_e(\tau)^{i,i'} = 0, \ \forall \tau$ *which is robust to the presence of heteroschedasticity in VAR residuals.*

In implementing white noise tests, one should remember that since VAR residuals are estimated, the asymptotic covariance matrix of the ACRF must include parameter uncertainty. Contrary to what one would expect, the covariance matrix of the estimated residuals is smaller than the one based on the true ones (see e.g. Lutkepohl (1991, p.142-148)). Hence, $\frac{1}{T}$ is conservative in the sense that the null hypothesis will be rejected less often than indicated by the significance level.

Portmanteau or Q-tests for the whiteness of the residuals can also be used to choose the lag length of a VAR. Both Portmanteau and Q-tests are designed to verify the null that $ACRF_e^{\tau} = (ACRF_e(1), \ldots, ACRF_e(\tau)) = 0$, (the alternative is $ACRF_e^{\tau} \neq 0$). The Portmanteau statistic is $PS(\tau) = T \sum_{i=1}^{\tau} tr(ACF(i)' ACF(0)^{-1'} ACF(i) ACF(0)^{-1}) \xrightarrow{D} \chi^2(m^2(\tau - q))$ for $\tau > q$ under the null. The Q-statistic is $QS(\tau) = T(T+2) \sum_{i=1}^{\tau} \frac{1}{T-i}$ $tr(ACF(i)' ACF(0)^{-1'} ACF(i) ACF(0)^{-1})$. For large T, it has the same asymptotic distribution as $PS(\tau)$.

Exercise 4.17 *Use US quarterly data from 1960:1 to 2002:4 to optimally select the lag length of a VAR with output, prices, nominal interest rate and money. Use modified LR,* *AIC, HQC, SWC and white noise tests. Does it make a difference if the sample is 1970-2003 or 1980-2003? How do you interpret differences across tests and/or samples?*

4.2.3 Nonlinearities and nonnormalities

So far we have focused on linear specifications. Since time aggregation washes most of the nonlinearities out, the focus is hardly restrictive, at least for quarterly data. However, with monthly data nonlinearities could be important (especially if financial data is used). Furthermore, time variations in the coefficients (see chapter 10), outliers or structural breaks may also generate (in a reduced form sense) nonlinearities and nonnormalities in the residuals of a constant coefficient VAR. Hence, one wants methods to detect departures from nonlinearities and nonnormalities if they exist.

In deriving the MA representation we have used linear projections. Since omitted nonlinear terms will end up in the error term, the same ideas employed in testing for white noise residuals can be used to check if nonlinear effects are present.

Two ways of formally testing for nonlinearities are the following: i) run a regression of estimated VAR residuals on nonlinear functions of the lagged dependent variables and examine the significance of estimated coefficients adjusting standard errors for the fact that e_t is proxied by estimated residuals. ii) Directly insert high order terms in the VAR and examine their significance. Graphical techniques, e.g. a scatter plot of estimated residuals against nonlinear functions of the regressors, could also be used as diagnostics.

There is also an indirect approach to check for nonlinearities which builds on the idea that whenever nonlinear terms are important, the moments of the residuals have a special structure. In particular, their distribution will be non-normal, even in large samples.

Testing for nonnormalities is simple: a normal white noise process with unit variance has zero skewness (third moment) and kurthosis (the forth moment) equal to 3. Hence, an asymptotic test for nonnormalities is as follows. Let $\hat{e}_t = y_t - \sum_j \hat{A}_j y_{t-j}$; $\Sigma_e = \frac{1}{T-1}\sum_t \hat{e}_t\hat{e}_t'$; $\tilde{e}_t = \tilde{\mathcal{P}}^{-1}\hat{e}_t$; $\tilde{\mathcal{P}}\tilde{\mathcal{P}}' = \Sigma_e$ where $\hat{A}_j$ is an estimator of A_j. Define $\mathsf{S}_{1i} = \frac{1}{T}\sum_t \tilde{e}_{it}^3$; $\mathsf{S}_{2i} = \frac{1}{T}\sum_t \tilde{e}_{it}^4, i = 1,\ldots,m$, $\mathsf{S}_j = (\mathsf{S}_{j1},\ldots,\mathsf{S}_{jm})', j = 1,2$ and let 3_m be a $m \times 1$ vector with 3 in each entry. Then $\sqrt{T}\begin{bmatrix} \mathsf{S}_1 \\ \mathsf{S}_2 - 3_m \end{bmatrix} \xrightarrow{D} \mathsf{N}(0, \begin{bmatrix} 6 \times I_m & 0 \\ 0 & 24 \times I_m \end{bmatrix})$.

4.2.4 Stationarity

Covariance stationarity is crucial to derive a VAR representation with constant coefficients. However, a time varying MA representation for a nonstationary y_t always exists if the other assumptions used in the Wold theorem hold. If $\sum_j D_{jt}^2 < \infty$ for all t, a non-stationary VAR representation can be derived. Hence, time varying coefficient VAR models, which we examine in chapter 10, are the natural alternative to covariance stationary structures.

While covariance stationarity is unnecessary, it is a convenient property to have when estimating VAR models. Also, although models with smooth changes in the coefficients may be the natural extensions of covariance stationary models, the literature has focused on a more extreme form of nonstationarity: unit root processes. Unit root models are less natural for two reasons: they imply drastically different dynamic properties; classical statistics has difficulties in testing this null hypothesis in the presence of a near-unit root alternative (see e.g. Watson (1995)). Despite these problems, contrasting stationary vs. unit root behavior has become a rule, the common wisdom being that macroeconomic time series are characterized by near-unit root behavior, i.e. they are in the grey area where the tests have low power. Hence, it will take a long time for a randomly perturbed series to revert back to the original (steady) state.

Unit root tests are somewhat tangential to the scope of the book. Favero (2001) provides an excellent review of this literature. Hence, we limit attention to the implications that nonstationary (or near nonstationary) has for the specification of the VAR, for the estimation of the parameters and for the identification of structural shocks.

If a test has detected one or more unit roots, how should one proceed in specifying a VAR? Suppose we are confident in the testing results and that all variables are either stationary or integrated, but no cointegration is detected. Then one would difference unit-root variables until covariance stationary is obtained and estimate the VAR using transformed variables. For example, if all variables are I(1), a VAR in growth rates is appropriate.

Specification is simple also when there are some cointegrating relationships. For example, both prices and money may display unit root behavior but real balances may be stationary. In this case, one typically transforms the VAR into a vector error correction model (VECM) and either imposes the cointegrating relationships (using the theoretical or the estimated restrictions) or jointly estimates short run and long run coefficients from the data. VECMs are preferable here to differenced VARs because the latters throw away information about the long run properties of the data. Plugging-in estimates of the long run relationships is justified since estimates of the long run relationships are super-consistent, i.e. they asymptotically converge at the rate T (estimates of short run relationships converge asymptotically at the rate $T^{0.5}$). Since a VECM is a reparametrization of the VAR in levels, the latter is appropriate if all variables are cointegrated, even though some (or all) of its components are not covariance stationary.

Despite two decades of work in the area, unit root tests still have poor small sample properties. Furthermore, barring exceptional circumstances, neither explosive nor unit root behavior has been observed in long stretches of OECD macroeconomic data. Both reasons may cast doubts about the non-stationarities detected and the usefulness of such tests.

When doubts about the tests exist, one can indirectly check the reasonableness of the stationarity assumption by studying estimated residuals. In fact, if y_t is nonstationary and no cointegration emerges, the estimated residuals are likely to display nonstationary path. Hence a plot of the VAR residuals may indicate a problem if it exists. Practical experience suggests that VAR residuals show breaks and outliers but they rarely display unit root type behavior. Hence, a level VAR could be appropriate even when y_t looks nonstationary. It is also important to remember that the properties of y_t are important in testing hypotheses about the coefficients since classical distribution theory is different when unit roots are present. Consistent estimates of VAR coefficients obtain with classical methods even when unit roots are present (see Sims, Stock and Watson (1990)).

A final argument against the use of specification tests for stationarity comes from a Bayesian perspective. In Bayesian analysis the posterior distribution of the quantities of interest is all that matters. While Bayesian and classical analyses have many common aspects, they dramatically differ when unit roots are present. In particular, while the classical asymptotic distribution of coefficients estimates under unit roots is nonstandard, the posterior distribution is unchanged. Therefore, if one takes a Bayesian perspective to testing, no adjustment for nonstationarity is required.

Finally, one should remember that pretesting has consequences for the distribution of parameters estimates since incorrect choices produce inconsistent estimates of the quantities of interest. To minimize pretesting problems, we recommend to start assuming covariance stationarity and deviate from it only if the data overwhelmingly suggests the opposite.

4.2.5 Breaks

While exact unit root behavior is unlikely to be relevant in macroeconomics, changes in the intercept, in the dynamics or in the covariance matrix of a vector of time series are quite common. A time series with breaks is neither stationary nor covariance stationary. To avoid problems, applied researchers typically focus attention on subsamples which are (assumed to be) homogenous. However, this is not always possible: the break may occur at the end of the sample (e.g. creation of the Euro); there maybe several of them; or they may be linked to expansions and contractions and it may be unwise to throw away runs with these characteristics.

While structural breaks with dramatic changing dynamics may sometimes occur (e.g. breakdown or unification of a country), it is more often the case that time series display slowly evolving features with no abrupt changes at one specific point - a pattern which would be more consistent with a time varying coefficient specifications. Nevertheless, it may be useful to have tools to test for structural breaks if visual inspection suggests that such a pattern may be present. If the break date is known, Chow tests can be used. Let Σ_e^{re} be the covariance matrix of the VAR residuals with no breaks and $\Sigma_e^{un} = \Sigma_e^{un}(1,\bar{t}) + \Sigma_e^{un}(\bar{t}+1,T)$ is the covariance matrix when a break is allowed at $\bar{t}$. Then $CS(\bar{t}) = \frac{|\Sigma_e^{re}|-|\Sigma_e^{un}|)/\nu}{|\Sigma_e^{un}|/T-\nu} \sim F(\nu, T-\nu)$ where ν is the number of regressors in the model. When $\bar{t}$ is unknown but suspected to occur within an interval, one could run Chow tests for all $\bar{t} \in [t_1, t_2]$, take $\max_{\bar{t}} CT(\bar{t})$ and compare it with a modified F-distribution (critical values are e.g. in Stock and Watson (2002, p. 111)).

An alternative testing approach can be obtained by noting that if no break occurs the τ-steps ahead forecast error of $y_{t+\tau}$, $e_t(\tau) = y_{t+\tau} - y_t(\tau)$, should be similar to sample residuals. Then, under the null of no breaks at forecasting horizon τ, τ large $e_t(\tau) \xrightarrow{D} \mathsf{N}(0, \Sigma_e(\tau))$.

Exercise 4.18 *Show that an appropriate statistic to check for breaks over τ forecasting horizons is $FT(\tau) = e_t \Sigma_e^{-1} e_t \xrightarrow{D} \chi^2(\tau)$ under the null of no breaks, T large, where $e_t = (e_t(1), \dots e_t(\tau))$. (The alternative here is that the DGP for y_t differs before and after t).*

As usual these tests may be biased in small samples. A small sample version of the forecasting test is obtained using $\Sigma_e^c(\tau) = \Sigma_e(\tau) + \frac{1}{T}E[\frac{\partial y_t(\tau)}{\partial \alpha'}\Sigma_\alpha \frac{\partial y_t(\tau)}{\partial \alpha}']$ in place of $\Sigma_e(\tau)$.

4.3 Alternative Representations of VAR(q)

There are two alternative representations for a $VAR(q)$ which are easier to manipulate than (4.7) and are of use when deriving estimators of the unknown parameters of the model.

4.3.1 Companion form representation

The companion form representation transforms a VAR(q) model in a larger scale VAR(1) model and it is useful when one needs to compute moments or derive parameter estimates.

Let $\mathsf{Y}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \dots \\ y_{t-q+1} \end{bmatrix}$; $\mathsf{E}_t = [\begin{bmatrix} e_t \\ 0 \\ \dots \end{bmatrix} 0$; $\mathsf{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_q \\ I_m & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & I_m & 0 \end{bmatrix}$. Then (4.7) is

$$\mathsf{Y}_t = \mathsf{A}\mathsf{Y}_{t-1} + \mathsf{E}_t \quad \mathsf{E}_t \sim (0, \Sigma_E) \tag{4.11}$$

where $\mathsf{Y}_t, \mathsf{E}_t$ are $mq \times 1$ vectors and A is $mq \times mq$ matrix.

Example 4.9 *Consider a bivariate VAR(2) model. Here* $\mathsf{Y}_t = [y_t, y_{t-1}]'$ $\mathsf{E}_t = [e_t, 0]'$, *are a* 4×1 *vectors, and* $\mathsf{A} = \begin{bmatrix} A_1 & A_2 \\ I_2 & 0 \end{bmatrix}$ *is a* 4×4 *matrix.*

Moments of y_t can be immediately calculated from (4.11).

Example 4.10 *The unconditional mean of* y_t *can be computed using* $E(\mathsf{Y}_t) = [(I - \mathsf{A}\ell)^{-1}]$ $E(\mathsf{E}_t) = 0$ *and a selection matrix which picks the first* m *elements out of* $E(\mathsf{Y}_t)$. *To calculate the unconditional variance notice that, because of covariance stationarity*

$$\begin{aligned} E[(\mathsf{Y}_t - E(\mathsf{Y}_t))(\mathsf{Y}_t - E(\mathsf{Y}_t))'] &= \mathsf{A}E_t[(\mathsf{Y}_{t-1} - E(\mathsf{Y}_{t-1}))(\mathsf{Y}_{t-1} - E(\mathsf{Y}_{t-1})']\mathsf{A}' + \Sigma_E \\ \Sigma_Y &= \mathsf{A}\Sigma_Y\mathsf{A}' + \Sigma_E \end{aligned} \tag{4.12}$$

To solve (4.12) for Σ_Y *we will make use of the following result.*

Result 4.1 *If* T, V, R *are conformable matrices,* $vec(TVR) = (R' \otimes T)vec(V)$.

Then $vec(\Sigma_Y) = [I_{mq} - (\mathsf{A} \otimes \mathsf{A})]^{-1} vec(\Sigma_E)$ *where* I_{mq} *is a* $mq \times mq$ *identity matrix.*

Unconditional covariances and correlations can also be easily computed. In fact

$$\begin{aligned} ACF_Y(\tau) &= E[(\mathsf{Y}_t - E(\mathsf{Y}_t))(\mathsf{Y}_{t-\tau} - E(\mathsf{Y}_{t-\tau})'] \\ &= \mathsf{A}E_t[(\mathsf{Y}_{t-1} - E(\mathsf{Y}_{t-1}))(\mathsf{Y}_{t-\tau} - E(\mathsf{Y}_{t-\tau}))'] + E[\mathsf{E}_t(\mathsf{Y}_{t-\tau} - E(\mathsf{Y}_{t-\tau}))'] \\ &= \mathsf{A}ACF_Y(\tau - 1) = \mathsf{A}^\tau\Sigma_Y \quad \tau = 1, 2, \dots \end{aligned} \tag{4.13}$$

The companion form could also be used to obtain the spectral density matrix of y_t. Let $ACF_E(\tau) = cov(\mathsf{E}_t, \mathsf{E}_{t-\tau})$. Then the spectral density of E_t is $\mathcal{S}_E(\omega) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} e^{-i\omega\tau}$ $ACF_E(\tau)$ and $vec[\mathcal{S}_Y(\omega)] = [I(\omega) - \mathsf{A}(\omega)\mathsf{A}(-\omega)']vec[\mathcal{S}_E(\omega)]$ where $I(\omega) = \sum_j e^{-i\omega j} I$, $\mathsf{A}(\omega) = \sum_j e^{-i\omega j}\mathsf{A}^j$ and $\mathsf{A}(-\omega)'$ is the complex conjugate of $\mathsf{A}(\omega)$.

Exercise 4.19 *Suppose a VAR(2) has been fitted to unemployment and inflation data and* $\hat{A}_1 = \begin{bmatrix} 0.95 & 0.23 \\ 0.21 & 0.88 \end{bmatrix}$, $\hat{A}_2 = \begin{bmatrix} -0.05 & 0.13 \\ -0.11 & 0.03 \end{bmatrix}$ *and* $\hat{\Sigma}_e = \begin{bmatrix} 0.05 & 0.01 \\ 0.01 & 0.06 \end{bmatrix}$ *have been obtained. Calculate the spectral density matrix of* y_t. *What is the value of* $\mathcal{S}_Y(\omega = 0)$*?*

A companion form representation has also computational advantages when deriving estimators of the unknown parameters of the model. We first consider estimators obtained when no constraints (lag restrictions, zero restrictions, etc.) are imposed on the VAR; when $y_{-q+1}, \ldots y_0$ are fixed and e_t are normally distributed with covariance matrix Σ_e.

Given the VAR structure, $(y_t|y_{t-1}, \ldots, y_0, y_{-1}, \ldots, y_{-q+1}) \sim \mathsf{N}(\mathsf{A}_1\mathsf{Y}_{t-1}, \Sigma_e)$ where A_1 is a $m \times mq$ matrix containing the first m rows of A. The density of y_t is $f(y_t|y_{t-1}, \ldots, \mathsf{A}_1, \Sigma_e) = (2\pi)^{0.5m}|\Sigma_e|^{-0.5}exp[-0.5(y_t - \mathsf{A}_1\mathsf{Y}_{t-1})'\Sigma_e^{-1}(y_t - \mathsf{A}_1\mathsf{Y}_{t-1})]$. Hence $f(y_t, y_{t-1}, \ldots, |\mathsf{A}_1, \Sigma_e) = \prod_{t=1}^{T} f(y_t|y_{t-1}, \ldots, \mathsf{A}_1, \Sigma_e)$ and the log likelihood is

$$\mathcal{L}(\mathsf{A}_1, \Sigma_e|y_t) = -\frac{T}{2}(m\log(2\pi) - \log|\Sigma_e|) - \frac{1}{2}\sum_t (y_t - \mathsf{A}_1\mathsf{Y}_{t-1})'\Sigma_e^{-1}(y_t - \mathsf{A}_1\mathsf{Y}_{t-1})] \quad (4.14)$$

Taking the first order conditions with respect to $vec(\mathsf{A}_1)$ leads to

$$\mathsf{A}'_{1,ML} = [\sum_{t=1}^{T}\mathsf{Y}_{t-1}\mathsf{Y}'_{t-1}]^{-1}[\sum_{t=1}^{T}\mathsf{Y}_{t-1}y'_t] = \mathsf{A}'_{1,OLS} \quad (4.15)$$

Hence, when no restrictions are imposed, ML and OLS estimators of the first m rows of the companion matrix A coincide. Note that an estimator of the j-th row of A_1 (an $1 \times mq$ vector) is $\mathsf{A}'_{1j} = [\sum_t \mathsf{Y}_{t-1}\mathsf{Y}'_{t-1}]^{-1}[\sum_{t=1}^{T}\mathsf{Y}_{t-1}y_{jt}]$.

Exercise 4.20 *Provide conditions for* $\mathsf{A}_{1,ML}$ *to be consistent. Is it efficient?*

Exercise 4.21 *Show that if there are no restrictions on the VAR, OLS estimation of the parameters, equation by equation, is consistent and efficient.*

The result of exercise 4.21 is important: as long as all variables appear with the same lags in every equation, single equation OLS estimation is sufficient. Intuitively, such a VAR is a seemingly unrelated regression (SUR) model and for such models single equation and system wide methods are equally efficient (see e.g. Hamilton (1994, p.315)).

Using $\mathsf{A}_{1,ML}$ into the log likelihood we obtain $\ln\mathcal{L}(\Sigma_e|y_t) = -\frac{Tm}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma_e|) - \frac{1}{2}\sum_{t=1}^{T} e'_{t,ML}\Sigma_e^{-1}e_{t,ML}$ where $e_{t,ML} = (y_t - \mathsf{A}_{1,ML}\mathsf{Y}_{t-1})$. Taking the first order conditions with respect to $vech(\Sigma_e)$, where $vech(\Sigma_e)$ vectorizes the symmetric matrix Σ_e, and using the fact that $\frac{\partial(b'\mathcal{Q}b)}{\partial\mathcal{Q}} = b'b$; $\frac{\partial \log|\mathcal{Q}|}{\partial\mathcal{Q}} = (\mathcal{Q}')^{-1}$ we have $\frac{T}{2}\Sigma'_e - \frac{1}{2}\sum_{t=1}^{T} e_{t,ML}e'_{t,ML} = 0$ or

$$\Sigma'_{ML} = \frac{1}{T}\sum_{t=1}^{T} e_{t,ML}e'_{t,ML} \tag{4.16}$$

and the ML estimate of the (i,i') element of Σ_e is $\sigma_{i,i'} = \frac{1}{T}\sum_{t=1}^{T} e_{it,ML}e'_{i't,ML}$.

Exercise 4.22 *Show that Σ_{ML} is biased but consistent.*

4.3.2 Simultaneous equations format

Two other useful transformations of a VAR are obtained using the format of a simultaneous equations system. The first is obtained setting $x_t = [y_{t-1}, y_{t-2}, \ldots]$; $\mathsf{X} = [x_1, \ldots, x_T]'$ (a $T \times mq$ matrix), $\mathsf{Y} = [y_1, \ldots, y_T]'$ (a $T \times m$ matrix) and letting $\mathsf{A} = [A'_1, \ldots A'_q]' = \mathsf{A}'_1$ be a $mq \times m$ matrix to have

$$\mathsf{Y} = \mathsf{X}\mathsf{A}' + \mathsf{E} \tag{4.17}$$

The second transformation is obtained from (4.17). The equation for variable i in fact is $\mathsf{Y}_i = \mathsf{X}\mathsf{A}_i + \mathsf{E}_i$. Stacking the columns of $\mathsf{Y}_i, \mathsf{E}_i$ into $mT \times 1$ vectors we have

$$y = (I_m \otimes \mathsf{X})\alpha + e \equiv X\alpha + e \tag{4.18}$$

Note that in (4.17) all variables are grouped together for each t; in (4.18) all time periods for one variable are grouped together. As shown in chapter 10, (4.18) is useful to decompose the likelihood function of a VAR(q) into the product of a normal density, conditional on the OLS estimates of the VAR parameters, and a Wishart density for Σ_e^{-1}.

Using these representations it is immediate to compute moments of y_t.

Example 4.11 *The unconditional mean of y_t is $E(\mathsf{Y}) = E(\mathsf{X})\mathsf{A}'$ or $E(y) = E(I_m \otimes \mathsf{X})\alpha$. The unconditional variance is $E[\mathsf{Y}] \equiv \Sigma_Y = E\{[\mathsf{X} - E(\mathsf{X})]\mathsf{A}' - \mathsf{E}\}^2$ or $\Sigma_Y = E\{[(I_m \otimes \mathsf{X}) - E(I_m \otimes \mathsf{X})]\alpha + e\}^2$.*

Exercise 4.23 *Using (4.18), assuming that $\Sigma_{xx} = p\lim \frac{X'X}{T}$ exists and is non-singular and $\frac{1}{\sqrt{T}}vec(Xe) \xrightarrow{D} \mathsf{N}(0, \Sigma_{xx} \otimes \Sigma_e)$ show: (i)$p\lim_{T\to\infty} \alpha_{OLS} = \alpha$; (ii) $\sqrt{T}(\alpha_{OLS} - \alpha) \xrightarrow{D} \mathsf{N}(0, \Sigma_{xx}^{-1} \otimes \Sigma_e)$; (iii) $\Sigma_{e,OLS} = \frac{(y-X\alpha)(y-X\alpha)'}{T-mq}$ is such that $p\lim \sqrt{T}(\Sigma_{e,OLS} - \frac{ee'}{T}) = 0$.*

Estimators of the VAR parameters can also be obtained via the Yule-Walker equations. From (4.7) we have that $E[(y_t - E(y_t))(y_{t-\tau} - E(y_{t-\tau}))] = A(\ell)E[(y_{t-1} - E(y_{t-1}))(y_{t-\tau} - E(y_{t-\tau}))] + E[e_t(y_{t-\tau} - E(y_{t-\tau}))]$ for all $\tau \geq 0$. Hence, letting $ACF_y(\tau) = E[(y_t - E(y_t))(y_{t-\tau} - E(y_{t-\tau}))]$ we have

$$ACF_y(\tau) = A_1 ACF_y(\tau-1) + A_2 ACF_y(\tau-2) + \ldots + A_q ACF_y(\tau-q) \tag{4.19}$$

Example 4.12 *If $q=1$ (4.19) reduces to $ACF_y(\tau) = A_1 ACF_y(\tau-1)$. Given estimates of A_1 and Σ_e, we have that $ACF_y(0) \equiv \Sigma_y = A_1 \Sigma_y A_1' + \Sigma_e$ so $vec(\Sigma_y) = (I - A_1 \otimes A_1) vec(\Sigma_e)$ and $ACF_y(1) = A_1 ACF_y(0)$, $ACF_y(2) = A_1 ACF_y(1)$, etc.*

Equation (4.19) can also be more compactly written as $ACF_y = \mathsf{A}_1 ACF_y^*$ where $ACF_y = [ACF_y(1), \ldots ACF_y(q)]$; and $ACF_y^* = \begin{bmatrix} ACF_y(0) & \ldots & ACF_y(q-1) \\ \ldots & \ldots & \ldots \\ ACF_y(-q+1) & \ldots & ACF_y(0) \end{bmatrix}$. Then an estimate of A_1 is $\mathsf{A}_{1,YW} = ACF_y(ACF_y^*)^{-1}$.

Exercise 4.24 *Show that $\mathsf{A}_{1,YW} = \mathsf{A}_{1,ML}$. Conclude that Yule-Walker and ML estimators have the same asymptotic properties.*

Exercise 4.25 *Show how to modify the Yule-Walker estimator when $E(y_t)$ is unknown. Show that the resulting estimator is asymptotically equivalent to $\mathsf{A}_{1,YW}$.*

It is interesting to study what happens when a VAR is estimated under some restrictions (exogeneity, cointegration, lag elimination, etc.). Suppose restrictions are of the form $\alpha = R\theta + r$ where R is $mk \times k_1$ matrix of rank k_1; r is a $mk \times 1$ vector; θ a $k_1 \times 1$ vector.

Example 4.13 *i) Consider the restriction $A_q = 0$. Here $k_1 = m^2(q-1)$, $r = 0$, and $R = [I_{k_1}, 0]$*
ii) Suppose that y_{2t} is exogenous for y_{1t} in a bivariate VAR(2). Here $R = blockdiag[R_1, R_2]$ where $R_i,\ i = 1, 2$ is upper triangular.

Using (4.18) we have $y = (I_m \otimes \mathsf{X})\alpha + e = (I_m \otimes \mathsf{X})(R\theta + r) + e$ or $y - (I_m \otimes \mathsf{X})r = (I_m \otimes \mathsf{X})R\theta + e$. Since $\frac{\partial \ln \mathcal{L}}{\partial \theta} = R \frac{\partial \ln \mathcal{L}}{\partial \alpha}$ then

$$\theta_{ML} = [R'(\Sigma_e^{-1} \otimes \mathsf{X}'\mathsf{X})R]^{-1} R' [\Sigma_e^{-1} \otimes \mathsf{X}](y - (I_m \otimes \mathsf{X})r) \quad (4.20)$$

$$\alpha_{ML} = R\,\theta_{ML} + r \quad (4.21)$$

$$\Sigma_e = \frac{1}{T} \sum_t e_{ML} e_{ML}' \quad (4.22)$$

Exercise 4.26 *Verify that when a VAR is estimated under some restrictions:*
i) ML estimates are different from OLS estimates.
ii) ML estimates are consistent and efficient if the restrictions are true but inconsistent if the restrictions are false.
iii) OLS is consistent when stationarity is incorrectly assumed but t-tests are incorrect.
iv) OLS is inconsistent if lag restrictions are incorrect.

4.4 Reporting VAR results

It is rare to report estimated VAR coefficients. Since the number of parameters is large presenting all of them is cumbersome. Furthermore, they are poorly estimated: except for the first own lag, in general, they are all insignificant. It is therefore typical to report functions of the VAR coefficients which summarize information better, have some economic meaning and, hopefully, are more precisely estimated. Among the many possible functions, three are typically used: impulse responses, variance and historical decompositions. Impulse responses trace out the MA of the system, i.e. they describe how $y_{it+\tau}$ responds to a shock in $e_{i't}$; the variance decomposition measures the contribution of $e_{i't}$ to the variability of $y_{it+\tau}$; the historical decomposition describes the contribution of shock $e_{i't}$ to the deviations of $y_{it+\tau}$ from its baseline forecasted path.

4.4.1 Impulse responses

There are three ways to calculate impulse responses which roughly correspond to recursive, nonrecursive (companion form) and forecast revision approaches. In the recursive approach, the impulse response matrix at horizon τ is $D_\tau = \sum_{j=1}^{max[\tau,q]} D_{\tau-j}A_j$ where $D_0 = I,\ A_j = 0\ \forall\ \tau \geq q$. Clearly, a consistent estimate is obtained if a consistent $\hat{A}_j$ is used in place of A_j.

Example 4.14 *Consider a VAR(2) with $y_t = A_0 + A_1 y_{t-1} + A_2 y_{t-2} + e_t$. Then the response matrices are: $D_0 = I, \quad D_1 = D_0 A_1, \quad D_2 = D_1 A_1 + D_0 A_2,\ \ldots, D_\tau = D_{\tau-1}A_1 + D_{\tau-2}A_2$.*

Calculation of meaningful impulse responses requires orthogonal disturbances. Let $\tilde{\mathcal{P}}$ be a square matrix such that $\tilde{\mathcal{P}}\tilde{\mathcal{P}}' = \Sigma_e$. Then the impulse response matrix to orthogonal shocks $\tilde{e}_t = \tilde{\mathcal{P}}^{-1}e_t$ at horizon τ is $\tilde{D}_\tau = D_\tau \tilde{\mathcal{P}}$.

Exercise 4.27 *Provide the first 5 elements of the MA representation of a bivariate VAR(3) with orthogonal shocks.*

When the VAR is in a companion form, we can compute impulse responses in a different way. Using (4.11) and repeatedly substituting for $\mathsf{Y}_{t-\tau},\ \tau = 1, 2, \ldots$ we have:

$$\mathsf{Y}_t = \mathsf{A}^t \mathsf{Y}_0 + \sum_{\tau=0}^{t-1} \mathsf{A}^\tau \mathsf{E}_{t-\tau} \tag{4.23}$$

$$= \mathsf{A}^t \mathsf{Y}_0 + \sum_{\tau=0}^{t-1} \tilde{\mathsf{A}}^\tau \tilde{\mathsf{E}}_{t-\tau} \tag{4.24}$$

where $\tilde{\mathsf{A}}^\tau = \mathsf{A}^\tau \tilde{\mathsf{P}},\ \tilde{\mathsf{E}}_{t-\tau} = \tilde{\mathsf{P}}^{-1}\mathsf{E}_t,\ \tilde{\mathsf{P}}\tilde{\mathsf{P}}' = \Sigma_E$. (4.23) is used with non-orthogonal residuals, (4.24) with orthogonal ones. The first m rows of A^τ provide the required responses.

Exercise 4.28 *Using the companion form of a bivariate VAR(2) show the first 4 elements of A^τ.*

A final way to compute impulse responses uses forecast revisions of future y_ts. We will use the companion form representation to illustrate the point but the argument goes through with any representation. Let $\mathsf{Y}_t(\tau) = \mathsf{A}^\tau \mathsf{Y}_t$ and $\mathsf{Y}_{t-1}(\tau) = \mathsf{A}^{\tau+1}\mathsf{Y}_{t-1}$ be the τ-steps and $\tau+1$-steps ahead forecast of Y_t. Hence the forecast revision is

$$Rev_t(\tau) = \mathsf{Y}_t(\tau) - \mathsf{Y}_{t-1}(\tau) \quad = \quad \mathsf{A}^\tau[\mathsf{Y}_t - \mathsf{A}\mathsf{Y}_{t-1}] = \mathsf{A}^\tau \mathsf{E}_t \qquad (4.25)$$

Example 4.15 *Suppose we shock the i'-th component of* e_t *once at time* t, *i.e.* $e_{i't} = 1;\ e_{i'\tau} = 0,\ \tau > t;\ e_{it} = 0\,\forall i \neq i', \forall t$. *Then* $Rev_{t,i'}(1) = \mathsf{A}_{i',.}$; $Rev_{t,i'}(2) = \mathsf{A}^2_{i',.}$; $Rev_{t,i'}(\tau) = \mathsf{A}^\tau_{i',.}$ *where* $\mathsf{A}_{i',.}$ *is the i-th column of* A. *Therefore, the response of* $y_{i,t+\tau}$ *to a shock in* $e_{i't}$ *can be read off the* τ*-step ahead forecast revisions.*

Example 4.16 *At times cumulative multipliers are required. For example, in examining the effects of fiscal disturbances on output one may want to measure the cumulative displacement produced by a shock up to horizon* τ. *Alternatively, in examining the relationship between money growth and inflation one may want to know whether an increase in the former translates in an increase in the latter in the long run of the same amount. In the first case one computes* $\sum_{j=0}^{\tau} D_j$, *in the second* $\lim_{j\to\infty}\sum_{j=0}^{\tau} D_j$.

4.4.2 Variance decomposition

To derive the variance decomposition we use (4.7). The τ-step ahead forecast error is $y_{t+\tau} - y_t(\tau) = \sum_{j=0}^{\tau-1}\tilde{D}_j\tilde{e}_{t+\tau-j}$ where $D_0 = I$ and $\tilde{e}_t = \tilde{\mathcal{P}}^{-1}e_t = \tilde{\mathcal{P}}_1^{-1}e_{1t} + \ldots + \tilde{\mathcal{P}}_m^{-1}e_{mt}$ are orthogonal disturbances. Hence $\Sigma_{\tilde{e}} = \tilde{\mathcal{P}}_1^{-1}\tilde{\mathcal{P}}_1^{-1'}\Sigma_e + \ldots + \tilde{\mathcal{P}}_m^{-1}\tilde{\mathcal{P}}_m^{-1'}\Sigma_e$. The MSE of the forecast is

$$\begin{aligned} MSE(\tau) &= E[y_{t+\tau} - y_t(\tau)]^2 = \Sigma_e + D_1\Sigma_e D_1' + \ldots + D_{\tau-1}\Sigma_e D_{\tau-1}' \\ &= \sum_{i=1}^{m}\Sigma_{\tilde{e}}(\tilde{\mathcal{P}}_i^{-1}\tilde{\mathcal{P}}_i^{-1'} + \tilde{D}_1\tilde{\mathcal{P}}_i^{-1}\tilde{\mathcal{P}}_i^{-1'}\tilde{D}_1' + \ldots + \tilde{D}_{\tau-1}\tilde{\mathcal{P}}_i^{-1}\tilde{\mathcal{P}}_i^{-1'}\tilde{D}_{\tau-1}') \end{aligned} \qquad (4.26)$$

Hence the percentage of the variance in $y_{i,t+\tau}$ due to $e_{i',t}$

$$VD_{i,i'}(\tau) = \frac{\Sigma_{\tilde{e}}(\tilde{\mathcal{P}}_{i'}^{-1}\tilde{\mathcal{P}}_{i'}^{-1'} + \tilde{D}_{1i}\tilde{\mathcal{P}}_{i'}^{-1}\tilde{\mathcal{P}}_{i'}^{-1'}\tilde{D}_{1i}' + \ldots + \tilde{D}_{\tau-1,i}\tilde{\mathcal{P}}_{i'}^{-1}\tilde{\mathcal{P}}_{i'}^{-1'}\tilde{D}_{\tau-1,i}')}{MSE(\tau)} \qquad (4.27)$$

A compact way to rewrite (4.27)is $VD(\tau) = \Sigma_{D_\tau}^{-1}\sum_{j=0}^{\tau-1} D_j \bigodot D_j$ where $\Sigma_{D_\tau} = diag[\Sigma_{D_{\tau,11}}, \ldots, \Sigma_{D_{\tau,mm}}] = \sum_{j=0}^{\tau-1} D_j D_j'$ and where $D_j \bigodot D_j$ is a matrix with $D_j^{i,i'} * D_j^{i,i'}$ in the i, i' position ($\bigodot$ is called Hadamman product (see e.g. Mittnick and Zadrozky(1993)).

4.4.3 Historical decomposition

Let $e_{i,t}(\tau) = y_{i,t+\tau} - y_{i,t}(\tau)$ be the τ-steps ahead forecast error in the i-th variable of the VAR. The historical decomposition of $e_{i,t}(\tau)$ can be calculated using

$$e_{i,t}(\tau) = \sum_{i'=1}^{m}\tilde{D}^{i'}(\ell)\tilde{e}_{i't+\tau} \qquad (4.28)$$

Example 4.17 *Consider a bivariate VAR(1). At horizon τ we have $y_{t+\tau} = Ay_{t+\tau-1} + e_{t+\tau} = \ldots = A^\tau y_t + \sum_{j=0}^{\tau-1} A^j e_{t+\tau-j}$ so that $e_t(\tau) = \sum_{j=0}^{\tau-1} A^j e_{t+\tau-j} = A(\ell)e_{t+\tau}$. Hence, deviations from the baseline forecasts of the first variable from t to $t+\tau$ due to, say, structural supply shocks are $\tilde{A}_{11}(\ell)\tilde{e}_{1,t+\tau}$ and to, say, structural demand shocks are $\tilde{A}_{12}(\ell)\tilde{e}_{2,t+\tau}$.*

From (4.27) and (4.28) it is immediate to notice that the ingredients needed to compute impulse responses, variance and historical decompositions are the same. Therefore, these statistics package the same information in a different way.

Exercise 4.29 *Using the estimate obtained in exercise 4.19, compute the variance and the historical decomposition for the two variables at horizons 1,2 and 3.*

4.4.4 Distribution of Impulse Responses

To assess the statistical (and the economic) significance of the effect of certain shocks, we need standard errors. As we have shown, impulse responses, variance and historical decompositions are complicated functions of the estimated VAR coefficients and of the covariance matrix of the shocks. Therefore, even when the distribution of the latters is known, it is not easy to find their distribution. In this subsection we describe three approaches to compute standard errors: one based on asymptotic theory and two based on resampling methods. All procedures are easy to implement when orthogonal shocks are generated by Choleski factorizations, i.e. if $\tilde{P}$ is lower triangular and need minor modification when the system is not contemporaneously recursive (but just-identified). In the other cases, resampling methods have a slight computational hedge.

Since impulse responses, variance and historical decompositions all use the same information we only discuss how to compute standard errors for impulse responses. The reader will be asked to derive the corresponding expressions for the other two statistics.

•The δ-method

The method pioneered by (Lutkepohl (1991)) and Mittnick and Zadrozky (1993) uses asymptotic approximations and works as follows. Suppose that $\alpha \xrightarrow{D} \mathsf{N}(0, \Sigma_\alpha)$. Then any differentiable function $f(\alpha)$ will have asymptotically the distribution $\mathsf{N}(0, \frac{\partial f}{\partial \alpha}\Sigma_\alpha \frac{\partial f}{\partial \alpha}')$ provided that $\frac{\partial f}{\partial \alpha} \neq 0$. Since impulse responses are differentiable functions of the VAR parameters and of the covariance matrix, their asymptotic distribution can be easily obtained.

Let $\mathsf{S} = [I, 0, \ldots, 0]$ be a $m \times mq$ selection matrix so that $y_t = \mathsf{S}\mathsf{Y}_t$ and $\mathsf{E}_t = \mathsf{S}'e_t$, consider the revision of the forecast at step τ and let

$$rev_t(\tau) = \mathsf{S}Rev_t(\tau) = \mathsf{S}[\mathsf{Y}_t(\tau) - \mathsf{Y}_{t-1}(\tau)] = \mathsf{S}[\mathsf{A}^\tau \mathsf{S}'\mathsf{E}_t] \equiv \psi_\tau e_t \tag{4.29}$$

We want the asymptotic distribution of the $m \times m$ matrix ψ_τ. Taking total differentials

$$d\psi_\tau = \mathsf{S}[Id\mathsf{A}\mathsf{A}^{\tau-1} + \mathsf{A}d\mathsf{A}\mathsf{A}^{\tau-2} + \ldots + \mathsf{A}^{\tau-1}d\mathsf{A}I]\mathsf{S}' \tag{4.30}$$

Since $\text{var}(\mathsf{Y}_{t+\tau}) = \mathsf{A}^\tau var(\mathsf{E}_{t+k})(\mathsf{A}^\tau)'$, using the fact that $dZ = \begin{bmatrix} dZ_1 \\ 0 \end{bmatrix} = \mathsf{S}' dZ_1$ and result 4.1, we have that $vec(\mathsf{SA}^j(d\mathsf{A})\mathsf{A}^{\tau-(j+1)}\mathsf{S}') = vec(\mathsf{SA}^j(\mathsf{S}'d\mathsf{A}_1)\mathsf{A}^{\tau-(j+1)}\mathsf{S}') = [\mathsf{S}(\mathsf{A}^{\tau-(j+1)})' \otimes \mathsf{SA}^j\mathsf{S}']vec(d\mathsf{A}_1) = [\mathsf{S}(\mathsf{A}^{\tau-(j+1)})' \otimes \psi_j]vec(d\mathsf{A}_1)$. Hence

$$\frac{vec(d\psi_\tau)}{vec(d\mathsf{A}_1)} = \sum_{j=0}^{\tau-1}[\mathsf{S}(\mathsf{A}')^{\tau-(j+1)} \otimes \psi_j] \equiv \frac{\partial vec(\psi_\tau)}{\partial vec(\mathsf{A}_1)} \tag{4.31}$$

Given (4.31), it is immediate to find the distribution of ψ_τ.

Exercise 4.30 *Derive the asymptotic distribution of ψ_τ.*

The above formulas, which use the companion form, may be computationally cumbersome when m or q are large. In these cases, the following recursive formula may be useful

$$\frac{\partial D_\tau}{\partial \alpha} = \sum_{j=1}^{max[\tau,q]} [(D'_{\tau-j} \otimes I_m)\frac{\partial A_j}{\partial \alpha} + (I_m \otimes A_j)\frac{\partial D_{\tau-j}}{\partial \alpha})] \tag{4.32}$$

Exercise 4.31 *Derive the distribution of $VD(\tau)$ for orthogonal shocks.*

Standard error bands computed with the δ-method have three problems. First, they tend to have poor properties in experimental designs featuring small scale VARs and samples of 100-120 observations. Second, the asymptotic coverage is also poor when near unit roots or near singularities are present. Third, since estimated VAR coefficients have large standard errors, impulse responses have large standard errors as well resulting, in many cases, in insignificant responses at all horizons. For these reasons, methods which employ the small sample properties of the VAR coefficients might be preferred.

Exercise 4.32 *Derive the asymptotic distribution of the τ-th term of a historical decomposition.*

- Bootstrap methods

Bootstrap standard errors, first employed in VARs by Runkle (1987), are easy to compute. Using equation (4.7) one proceeds as follows:

Algorithm 4.2

1) *Obtain $A(\ell)_{OLS}$ and $e_{t,OLS} = y_t - A(\ell)_{OLS}y_{t-1}$.*

2) *Obtain $e^l_{t,OLS}$ via bootstrap and construct $y^l_t = A(\ell)_{OLS}y^l_{t-1} + e^l_{t,OLS}, l = 1, 2, \ldots, L$.*

3) *Estimate $A(\ell)^l_{OLS}$ using data constructed in 2). Compute $D^l_j, (\tilde{D}^l_j),\ j = 1, \ldots, \tau$.*

4) *Report percentiles of the distribution of* $D_j, (\tilde{D}_j)$ *(i.e. 16-84% or 2.5-97.5%), or the simulated mean and the standard deviation of* $D_j, (\tilde{D}_j),\ j = 1, \ldots, \tau$.

Algorithm 4.2 is easily modified to produce confidence bands for other statistics.

Example 4.18 *To compute standard error bands for the variance decomposition one would insert the calculation of* $VD_{i,i'}(\tau)^l$ *as suggested in (4.27) after step 3) of algorithm 4.2.* $VD_{i,i'}(\tau)^l$ *is the percentage of the variance of* $y_{i,t}$ *explained by* $e_{i',t}$ *at horizon* τ *in replication* l. *Then in 4) order* $VD_{i,i'}(\tau)^l$ *and report percentiles or the first two moments.*

Few remarks are in order. First, bootstrapping is appropriate when e_t is a white noise with constant variance. Therefore, the approach yields poor standard error band estimates when the lag length of the VAR is misspecified or when heteroschedasticy is present.

Since conditional heteroschedasticity is less likely to emerge with low frequency data, one possible solution is to time aggregate the data before a VAR is run and standard errors are computed.

Second, estimates of the VAR coefficients are typically biased downward in small samples. For example, in a VAR(1) with the largest root around 0.95, a downward bias of about 30 percent is to be expected even when $T = 80 - 100$. Biasedness of $A(\ell)$ is a problem because in step 2) we are generating biased y_t series. Hence, the resulting distribution is likely to be centered around an incorrect estimate of the true VAR coefficients.

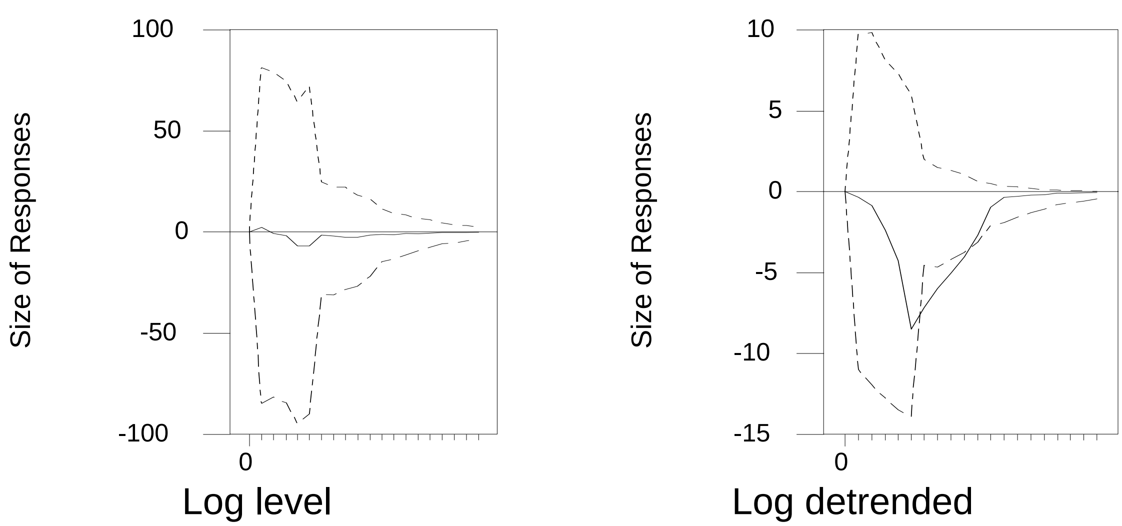

Figure 4.2: Bootstrap responses

Third, the bootstrap distribution of $D_j(\tilde{D}_j)$ is not scale invariant. In particular, units matter. This implies that standard error bands may not include, point estimates of the impulse responses. Such a problem emerges e.g. in figure 4.2, where we report one standard error bands for log output (right panel) or log linearly detrended level of output (left panel) in response to an orthogonal price shock in a bivariate VAR(4) system together with the point estimate of the responses. Clearly, the size and the shape of the band depend on the units. Furthermore, in the right panel there are horizons where the point estimate is outside the computed standard error band.

Finally, while it is common to report the mean and construct confidence bands using numerical standard deviations (across replications), this approach is unsatisfactory since it assumes symmetric distributions. Since simulated distributions of impulse responses tend to be highly skewed when $T < 100$, we recommend the use of simulated distribution percentiles in constructing confidence intervals (i.e extract the relevant band directly from the ordered replications at each horizon).

To solve the biasedness and the lack of scale invariance, Kilian (1998) has suggested a bootstrap-after-the-bootstrap procedure. The approach can be summarized as follows:

Algorithm 4.3

1) *Given $A(\ell)_{OLS}$, obtain $e^l_{t,OLS}$ and construct $y^l_t = A(\ell)_{OLS} y^l_{t-1} + e^l_{t,OLS}, l = 1, 2, \ldots, L$.*

2) *Estimate $A(\ell)^l_{OLS}$ for each l. If the bias is approximately constant in a neighbor of $A(\ell)_{OLS}$, $Bias(\ell) = E[A(\ell)_{OLS} - A(\ell)] \approx E[A(\ell)^l_{OLS} - A(\ell)_{OLS}]$.*

3) *Calculate the largest root of the system. If it is greater or equal than one, set $\tilde{A}(\ell) = A(\ell)_{OLS}$ - here the bias is irrelevant since estimates are superconsistent. Otherwise set $\tilde{A}(\ell) = A(\ell)_{OLS} - Bias(\ell)_{OLS}$, where $Bias(\ell)_{OLS} = \frac{1}{L}\sum_{l=1}^{L}[A(\ell)^l_{OLS} - A(\ell)_{OLS}]$.*

4) *Repeat 1)-3) of algorithm 4.2, L_1 times using $\tilde{A}(\ell)$ in place of $A(\ell)_{OLS}$.*

Kilian shows that the procedure of eliminating the bias, assuming that it is constant in a neighborhood of $A(\ell)_{OLS}$, has an asymptotic justification and that the bias correction becomes negligible asymptotically. It also shows that such an approach has a better small sample coverage properties than a simple bootstrap. However, when the bias is not constant in the neighborhood of $A(\ell)_{OLS}$, the properties of the bands produced from algorithm 4.3 may still be poor.

- Monte Carlo methods

Monte Carlo methods will be described in details in the last three chapters of this book. Here we describe a simple approach which allows the computation of standard error bands using the simultaneous equation representation of an unrestricted VAR(q).

As mentioned, the likelihood function of a VAR(q), $\mathcal{L}(\alpha, \Sigma_e|y_t)$, can be decomposed into a Normal portion for α, conditional on Σ_e, and a Wishart portion for Σ_e^{-1}. Assuming that no prior information for α, Σ_e is available, i.e. $g(\alpha, \Sigma_e) \propto |\Sigma_e|^{\frac{-(m+1)}{2}}$, the posterior distribution (which is proportional to the product of the likelihood and the prior) will have a form which is identical to the likelihood. Furthermore, the posterior for (α, Σ_e) will be proportional to the product of the posterior of $(\alpha|\Sigma_e, y_t)$ and of $(\Sigma_e|y_t)$. As detailed in chapter 10, the posterior for Σ_e^{-1} has a Wishart form with $T - mq$ degrees of freedom .

The posterior of $(\alpha|\Sigma_e,\ y_t)$ is normal centered at α_{OLS} with variance equals to $var(\alpha_{OLS})$. Hence, standard error bands for impulse responses can be constructed as follows:

Algorithm 4.4

1) *Generate* $T-mq$ *iid draws for* e_t^{-1} *from a* $\mathsf{N}(0,(\mathsf{Y}-\mathsf{X}\mathsf{A}_{OLS})'(\mathsf{Y}-\mathsf{X}A_{OLS}))$. *Compute* $\Sigma_e^l = (\frac{1}{T-mq}\sum_{t=1}^{T-mq}(e_t^{-1} - \frac{1}{T-mq}\sum_{t=1}^{T-mq} e_t^{-1})^2)^{-1}$.

2) *Draw* $\alpha^l = \alpha_{OLS} + \epsilon_t^l$, *where* $\epsilon_t^l \sim \mathsf{N}(0,\Sigma_e^l)$. *Compute* $D_j^l(\tilde{D}_j^l)$, $j=1,\ldots\tau$.

3) *Repeat 1)-2) L times and report percentiles.*

Three features of algorithm 4.4 are important. First, the posterior distribution is exact and conditional on the OLS estimator - which summarizes the information contained in the data. Therefore, biasedness of $A(\ell)_{OLS}$ is not an issue. Second, given the exact small sample nature of the posterior distribution, standard error bands are likely to be skewed and, possibly, leptokurtic. Therefore, bands extracted from percentiles are preferable to 1 or 2 standard error bands around mean. Third, algorithm 4.4 is appropriate only for just identified systems (both of semi-structural or of structural types). When the VAR system is overidentified, the technique described in section 3 of chapter 10 should be used.

Exercise 4.33 *Show how to use algorithm 4.4 to compute confidence bands for variance and historical decompositions.*

All three approaches we have described produce standard error bands estimates which are correlated. This is because responses at each step are correlated (see e.g. the recursive computation of impulse responses). Hence, plots connecting the points at each horizon are likely to misrepresent the true uncertainty. Sims and Zha (1999) propose a transformation which eliminates this correlation. Their approach relies on the following result.

Result 4.2 *If* $\tilde{D}_1,\ldots,\tilde{D}_\tau$ *are normally distributed with covariance matrix* $\Sigma_{\tilde{D}}$, *the best coordinate system is given by the projection on the principal components of* $\Sigma_{\tilde{D}}$.

Intuitively, we need to orthogonalize the covariance matrix of the impulse responses to break down the correlation of its elements. To implement such an orthogonalization, for structural coefficients, steps 1) to 3) of algorithm 4.4 remain unchanged, but we need to add the following two steps

4) Let the $\tau\times\tau$ covariance matrix of $\tilde{D}$ be decomposed as $\mathcal{P}_{\tilde{D}}\mathcal{V}_{\tilde{D}}\mathcal{P}'_{\tilde{D}} = \Sigma_{\tilde{D}}$, where $\mathcal{V}_{\tilde{D}} = diag\{v_j\}$ $\mathcal{P}_{\tilde{D}} = col\{pp_{.,\ j}\}, j=1,\ldots,\tau,$ $\mathcal{P}_{\tilde{D}}\mathcal{P}'_{\tilde{D}} = I$.

5) For each (i,i') report $\tilde{D}^*(i,i') \pm \sum_{j=1}^{\tau}\varrho_j pp_{.,j}$, where $\tilde{D}^*(i,i')$ is the mean of $\tilde{D}(i,i')$ and $\varrho_j = pp_{j,.}\tilde{D}(i,i')$.

In practice, it is often sufficient to use the largest eigenvalue of $\Sigma_{\tilde{D}}$ to have a good idea of the existing uncertainty. Then standard error bands are $\tilde{D}^*(i,i') \pm pp_{.,j}\sqrt{v_{\sup}}$ (symmetric) and $[(\tilde{D}^*(i,i') - \varrho_{sup,.16}; \tilde{D}^*(i,i') + \varrho_{sup,.84}]$ (asymmetric), where $\varrho_{sup,.r}$ is the r-th percentile of ϱ_j computed using the largest eigenvalue of $\Sigma_{\tilde{D}}$ and $v_{\sup} = \sup_j v_j$

Exercise 4.34 *Show how to apply the Sims and Zha approach to orthogonalize standard error bands computed with the δ-method.*

Given that the asymptotic approach has poor small sample properties, which of the two resampling methods should one prefer? A-priori the choice is difficult: the bootstrap method does not require distributional assumptions but it requires homoschedasticity. Also, unless Kilian method is used, bands may have little meaning. The MC approach works even with heteroschedasticity but normality of the errors or a large sample are required. The question is therefore empirical. Sims and Zha (1999) show that, in specific experiments, the MC approach outperforms the bootstrap approach but not uniformly so.

4.4.5 Generalized Impulse Responses

This subsection discusses the computation of impulse responses for nonlinear structures. Since VARs with time varying coefficients fit well into this class, it is worthwhile to study how impulse responses for these models can be constructed. The discussion here is basic; more details are in Gallant, Tauchen and Rossi (1993) and Koop, Pesaran and Potter (1995).

In linear models impulse responses do not depend on the sign or the size of shocks nor on their history. This simplifies the computations but prevents researchers from studying interesting economic questions such as: do shocks which occur in a recession produce different dynamics than those in an expansion? Are large shocks different than small ones? In nonlinear models, responses do depend on the sign, the size and the history of the shocks up to the point where they are computed.

Let $\mathcal{F}_{t-1}$ be the history of y_{t-1} up to $t-1$. In general, $y_{t+\tau}$ depends on $\mathcal{F}_{t-1}$, the parameters α of the model and the innovations $e_{t+j}, j = 0, \ldots, \tau$. Let $Rev(\tau, \mathcal{F}_{t-1}, \alpha, e^*) = E(y_{t+\tau}|\alpha,\ \mathcal{F}_{t-1}, e_t = e^*, e_{t+j} = 0, j \geq 1) - E(y_{t+\tau}|\alpha, \mathcal{F}_{t-1}, e_{t+j} = 0, j \geq 0)$.

Example 4.19 *Consider $y_t = Ay_{t-1} + e_t$, let $\tau = 2$ and assume $|A| < 1$. Then $E(y_{t+2}|A, \mathcal{F}_{t-1}, e_{t+j} = 0,\ j \geq 0) = A^3 y_{t-1}$ and $E(y_{t+2}|A, \mathcal{F}_{t-1}, e_t = e^*,\ e_{t+j} = 0,\ j \geq 1) = A^3 y_{t-1} + A^2 e^*$ and $Rev_y(\tau, \mathcal{F}_{t-1}, A, e^*) = A^2 e^*$ which is independent of history and of the size of the shock (hence set $e^* = 1$ or $e^* = \sigma_e$) and symmetric in the sign of e^* (hence set $e^* > 0$).*

Exercise 4.35 *Consider the model $\Delta y_t = A\Delta y_{t-1} + e_t$; $|A| < 1$. Calculate the impulse response function at a generic τ. Show it is independent of the history and that the size of e^* scales the whole impulse response function. Consider an ARIMA$(d_1, 1, d_2)$: $D_1(\ell)\Delta y_t = D_2(\ell)e_t$. Show that $Rev(\tau, \mathcal{F}_{t-1}, D_2(\ell), D_1(\ell), e^*)$ is history and size independent.*

Example 4.20 *Consider the model $\Delta y_t = A_1\Delta y_{t-1} + A_2\Delta y_{t-1}\mathcal{I}_{[\Delta y_{t-1} \geq 0]} + e_t$, where $\mathcal{I}_{[\Delta y_{t-1} \geq 0]} = 1$ if $\Delta y_{t-1} \geq 0$ and zero otherwise. Let $0 < A = A_1 + A_2 < 1$. Then, for $e_t = e^*$*

$Rev(\tau, \Delta y_{t-1}, A, e^*) = \frac{1-A^{\tau+1}}{1-A} e^*$ if $\Delta y_{t-1} \geq 0$ and $Rev(\tau, \Delta y_{t-1}, A, e^*) = \frac{1-A_1^{\tau+1}}{1-A_1} e^*$ if $\Delta y_{t-1} < 0$. Here $Rev(\tau, \Delta y_{t-1}, A, e^*)$ depends on the history of Δy_{t-1}.

Exercise 4.36 *Consider the logistic map $\tilde{y}_t = a\tilde{y}_{t-1}(1 - \tilde{y}_{t-1}) + v_t$ where $0 \leq a \leq 4$. This model can be transformed into a nonlinear AR(1) model: $y_t = A_1 y_{t-1} - A_2 y_{t-1}^2 + e_t$ for $A_2 \neq 0$, $-2 \leq A_1 \leq 2$, $A_1 = 2 - a$, $e_t = \frac{2-A_1}{A_2} v_t$, $y_t = \frac{A_1 - 1}{A_2} + \frac{2-A_1}{A_2}\tilde{y}_t$. Simulate the impulse response function. Does the sign and the size of e^* matter?*

In impulse responses computed from linear models $e_{t+j} = 0$, $\forall j \geq 1$. This is inappropriate in nonlinear models since it may violate bounds for e_t. In exercise 4.36 the bounds occur because the logistic map is unstable if y_{t-1} passes a threshold. These bounds depend on the realizations of $v_{t-\tau}$ and therefore vary over time. Also, when parameters are estimated, we either need to condition on a particular α (e.g. α_{OLS}) or integrate α out to compute forecast revisions. Generalized impulse (GI) responses are designed to meet all these requirements: in fact we condition on the size, the sign, the history of the shocks and, if required, on a particular estimate of α and integrate out all future shocks.

Definition 4.5 *Generalized impulse responses conditional on a shock e_t, a history $\mathcal{F}_{t-1}$ and a vector α are $GI_y(\tau, \mathcal{F}_{t-1}, \alpha, e_t) = E(y_{t+\tau}|\alpha, e_t,\ \mathcal{F}_{t-1}) - E(y_{t+\tau}|\alpha,\ \mathcal{F}_{t-1})$.*

Responses produced by definition 4.5 have three important properties. First $E(GI_y) = 0$. Second, $E(GI_y|\mathcal{F}_{t-1}) = 0$. Third, $E(GI_y|e_t) = E(y_{t+\tau}|e_t) - E(y_{t+\tau})$.

Example 4.21 *Three interesting cases where definition 4.5 is useful are the following:*

- *(Impulse responses in recession): GI conditional only on a history $\mathcal{F}_{t-1}$ in a region: $GI_y(\tau, \mathcal{F}_{t-1} \in \mathcal{F}_1, \alpha, e_t) = E(y_{t+\tau}|\alpha,\ \mathcal{F}_{t-1} \in \mathcal{F}_1, e_t) - E(y_{t+\tau}|\alpha,\ \mathcal{F}_{t-1} \in \mathcal{F}_1)$.*
- *(Impulse responses on average over histories): We have two options. GI conditional only on α: $GI_y(\tau, \alpha, e_t) = E(y_{t+\tau}|\alpha, e_t) - E(y_{t+\tau}|\alpha)$ and GI unconditional on α : $GI_y(\tau, e_t) = E(y_{t+\tau}|e_t) - E(y_{t+\tau})$.*
- *(Impulse responses if oil prices go above 40 dollars a barrel) GI conditional on a shock in a region: $GI_y(\tau, \mathcal{F}_{t-1}, \alpha, e_t) = E(y_{t+\tau}|\mathcal{F}_{t-1}, \alpha, e_t \in E_1) - E(y_{t+\tau}|\mathcal{F}_{t-1}, \alpha)$*

Definition 4.5 conditions on a particular value of α. In some situations we may want to treat parameters as random variables. This is important in applications where symmetric shocks may have asymmetric impact on y_t depending on the value of α. Alternatively, we may want to average α out of GI. As an alternative to definition 4.5 one could use:

Definition 4.6 *Generalized impulse responses, conditional on a shock e_t and a history $\mathcal{F}_{t-1}$, are $GI_y(\tau, \mathcal{F}_{t-1}, e_t) = E(y_{t+\tau}|\mathcal{F}_{t-1}, e_t) - E(y_{t+\tau}|\mathcal{F}_{t-1})$.*

Exercise 4.37 *Extend definitions 4.5-4.6 to condition on the size and the sign of e_t.*

In practice, GI are computed numerically using Monte Carlo methods. We show how to do this conditional a on history and a set of parameters in the next algorithm.

Algorithm 4.5

1) *Fix* $y_{t-1} = \hat{y}_{t-1}, \ldots, y_{t-\tau} = \hat{y}_{t-\tau};\ \alpha = \hat{\alpha}$.

2) *Draw* $e^l_{t+j}, j = 0, 1, \ldots,$ *from* $\mathsf{N}(0, \Sigma_e)$, $l = 1, \ldots L$ *and compute* $GI^l = (y^l_{t+\tau} | \hat{y}_{t-1}, \ldots \hat{y}_{t-\tau}, \hat{\alpha},\ e_t,\ e^l_{t+j},\ j > 1) - (y^l_{t+\tau} | \hat{y}_{t-1}, \ldots \hat{y}_{t-\tau}, \hat{\alpha},\ e_t = 0,\ e^l_{t+j},\ j > 1)$.

3) *Compute* $GI = \frac{1}{L}\sum_{l=1}^{L} GI^l$, $E(GI^l - GI)^2$ *and/or the percentiles of the distribution.*

Note that in algorithm 4.5 the history $(y_{t-1}, \ldots, y_{t-\tau})$ could be a recession or expansion and $\hat{\alpha}$ an OLS or a posterior estimator. In practice, when the model is multivariate we need to orthogonalize the shocks so as to be able to measure the effect of a shock. When e_t is normal, its response to a shock in $e_{i't}$ is $E(e_t | e_{i't} = e^*_{i'}) = E(e_t e_{i't})\sigma_{i'}^{-2} e^*_{i'}$ where $\sigma_i^2 = E(e_{i't})^2$ and this can be inserted in step 2) of algorithm 4.5 to compute GI. For example, for a linear VAR $GI(\tau, \mathcal{F}_{t-1}, e_{it}) = (\frac{A_\tau E(e_t, e_{i't})}{\sigma_i})\frac{e^*}{\sigma_i}$ and the generalized impulse of variable i equals $\mathsf{S}_i GI(\tau, \mathcal{F}_{t-1}, e_{it})$ where S_i is a selection vector with one in the i-th position and zero everywhere else. Here the term $\frac{e^*}{\sigma_i}$ is a scale factor and the first term measures the effect of a one standard error shock in the $i' - th$ variable. Note also, that $(\frac{A_\tau E(e_t, e_{i't})}{\sigma_i})$ corresponds to the effect obtained when the variables are assumed to have a Wold causal chain. Hence meaningful interpretations are possible only if the orthogonalization is derived from relevant economic restrictions.

Exercise 4.38 *Describe a Monte Carlo method to compute GI without conditioning on a particular history or a particular* α.

Example 4.22 *Consider the model* $\Delta y_t = A_1 \Delta y_{t-1} + A_2 \Delta y_{t-1} \mathcal{I}_{[\Delta y_{t-1} \geq 0]} + e_t$, *where* $\mathcal{I}_{[\Delta y_{t-1} \geq 0]}$ *is an indicator function. Then:*

- *GI responses allowing for randomness in* e_t *can be computed by fixing* $y_{t-1},\ A_1, A_2$ *and drawing* $e^l_{t+j},\ j \geq 0,\ l = 1, \ldots L$.
- *GI responses allowing for randomness in history can be computed fixing* $e_{t+j}\ j \geq 0, A_1, A_2$ *and drawing* y^l_{t-1}.
- *GI responses allowing for randomness in the parameters can be computed fixing* $y_{t-1},\ e_{t+j},\ j \geq 0$ *and drawing* $A^l_1,\ A^l_2$ *from some distribution (e.g. the asymptotic one).*
- *GI responses allowing for randomness in the size of* e_t *can be computed fixing* $y_{t-1},\ A_1, A_2,$ $e_{t+j}\ j > 1$ *and keeping those* e^l_t *that satisfy* $e^l_t \geq e^*$ *or* $e^l_t < e^*$. *If the process is multivariate apply the above to e.g.* e_{1t}, *after averaging over draws of* $(e_{2t}, \ldots, e_{mt})$.

Exercise 4.39 *Consider a bivariate model with inflation* π *and unemployment* UN, $y_t = A_1 y_{t-1} + A_2 y_{t-1} \mathcal{I}_{[\pi \geq 0]} + e_t$ *where* $\mathcal{I}_{[\pi > 0]}$ *is an indicator function. Calculate GI at steps 1 to 3 for an orthogonal shock in* π *when* $\pi \geq 0$ *and when* $\pi < 0$. *Does the size of* e_t *matter?*

Exercise 4.40 *Consider a switching bivariate AR(1) model with money and output:*

$$\Delta y_t = \begin{cases} \alpha_{01} + \alpha_{11}\Delta y_{t-1} + e_{1t} & \text{if} \quad \Delta y_{t-1} \leq \Delta \bar{y}, \quad e_{1t} \sim \mathsf{N}(0, \sigma_1^2) \\ \alpha_{02} + \alpha_{12}\Delta y_{t-1} + e_{2t} & \text{if} \quad \Delta y_{t-1} > \Delta \bar{y}, \quad e_{2t} \sim \mathsf{N}(0, \sigma_2^2) \end{cases}$$

Fix the size of the shock and the parameters and compute GI as a function of history. Fix the size of shocks and the history and compute GI as function of the parameters.

We defer further discussion on the computation of impulse responses for a particular type of non-linear model to chapter 10.

4.5 Identification

So far in this chapter, economic theory has played no role. Projections methods are used to derive the Wold theorem; statistical and numerical analysis are used to estimate the parameters and the distributions of interesting functions of the parameters. Since VARs are reduced form models it is impossible to structurally interpret the dynamics induced by their disturbances unless economic theory comes into play. As seen in chapter 2, Markovian DSGE models when approximated linearly or log linearly around the steady state typically deliver VAR(1) solutions. The reduced form parameters are complicated functions of the structural ones and the resulting set of extensive cross equations restrictions could be used to disentangle the latters if one is willing to take the model seriously as the process generating the data. When doubts about the quality of the model exists, one can still conduct inference as long as a subset of the model restrictions are credible or uncontroversial. Typical restrictions employed in the literature include constraints on the short run or long run impact of certain shocks on variables or informational delays (e.g. output is not contemporaneously observed by Central Banks when deciding interest rates). As we will argue later on, these restrictions are rarely produced by DSGE models. Restrictions involving lag responses or the dynamics are generally ignored being perceived as non robust or controversial.

To conduct structural analyses, one therefore starts from an unrestricted VAR(q) where all variables appear with the same lags in each equation, estimates the parameters of the VAR by OLS, imposes a minimal set of "structural" restrictions, possibly consistent with a variety of behavioral theories, and constructs impulse responses, historical decomposition, etc. to structural shocks. In this sense, VARs are at the antipodes of maximum likelihood or generalized method of moments approaches: the majority of the theoretical restrictions are disregarded; there is no interest in estimating preference and technology parameters; and only a structural interpretation of the shocks is sought.

We first examine identification in stationary and non-stationary VAR using zero-type (or constant-type) restrictions. Afterward, we discuss identification via sign restrictions.

4.5.1 Stationary VARs

Let the reduced form VAR be

$$y_t = A(\ell)y_{t-1} + e_t \quad e_t \sim \ iid\ (0, \Sigma_e) \tag{4.33}$$

We assume that associated with (4.33) there is a structural model of the form

$$y_t = \mathcal{A}(\ell)y_{t-1} + \mathcal{A}_0\epsilon_t \quad \epsilon_t \sim \ iid\ (0, \Sigma_\epsilon = diag\{\sigma^2_{\epsilon_i}\}) \tag{4.34}$$

Equation (4.34) generically defines a class of models but it is easy to show that it is non-empty. For example, many of the log-linearized DSGE models of chapter 2, produce solutions like (4.34) with $\mathcal{A}(\ell) = \mathcal{A}(\theta)$ and $\mathcal{A}_0 = \mathcal{A}_0(\theta)$ where θ are structural parameters. Matching contemporaneous coefficients in (4.33) and (4.34) implies $e_t = \mathcal{A}_0\epsilon_t$ or

$$\mathcal{A}_0\Sigma_\epsilon\mathcal{A}_0' = \Sigma_e \tag{4.35}$$

To compute responses to structural shocks we can proceed in two steps. First, we can estimate $A(\ell)$ and Σ_e from (4.33) using the techniques described in section 3. Second, from (4.35) and given identification restrictions, we estimate Σ_ϵ, free parameters of $\mathcal{A}_0$ and the structural dynamics $\mathcal{A}(\ell)$. This two-step approach resembles the indirect least square (ILS) technique used in a system of (static) structural equations (see Hamilton (1994, p. 244)). The main difference lies in the fact that here restrictions are imposed on the covariance matrix of reduced form residuals and not on the lags of the VAR or on the exogenous variables. This is convenient: had we imposed restrictions on the lags of the VAR, joint estimation of $A(\ell)$, Σ_e, Σ_ϵ and of free parameters of $\mathcal{A}_0$ would be required.

As in simultaneous equation systems there are necessary and sufficient conditions that need to be satisfied for identification. An order condition can be calculated as follows. On the left hand side of (4.35) there are m^2 free parameters, while given the symmetry of Σ_e, the right hand side has only $(m(m+1)/2)$ free parameters. Hence, to go from reduced form to structural shocks we need, at least, $m(m-1)/2$ restrictions (with more restrictions structural shocks are overidentified).

Example 4.23 *Consider a trivariate model with hours, productivity, and interest rates. Suppose that $\mathcal{A}_0$ is lower triangular, that is, suppose that shocks to hours enter contemporaneously in the productivity and interest rate equations and that productivity shocks enter only contemporaneously in the interest rate equation. This obtains, e.g. if interest rate shocks take time to produce effects and if hours are predetermined with respect to productivity. If structural shocks are independent, $\mathcal{A}_0$ has $m(m-1)/2 = 3$ zeros restrictions. Hence, the order condition is satisfied.*

Example 4.24 *Consider VAR with includes output, prices, nominal interest rates and money, $y_t = [GDP_t, p_t, i_t, M_t]$. Suppose that a class of models suggests that output contemporaneously reacts only to its own shocks; that prices respond contemporaneously to output and money shocks; that interest rates respond contemporaneously only to money shocks, while money contemporaneously responds to all shocks. Then* $\mathcal{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a^0_{12} & 1 & 0 & a^0_{22} \\ 0 & 0 & 1 & a^0_{31} \\ a^0_{41} & a^0_{42} & a^0_{43} & 1 \end{bmatrix}$. *Since there are six (zero) restrictions, structural shocks are identifiable.*

Exercise 4.41 *Suppose we have extraneous information which allows us to pin down some of the parameters of $\mathcal{A}_0$. For example, suppose in a trivariate system with output, hours and taxes, we can obtain estimates of the elasticity of hours with respect to taxes. How many restrictions do you need to identify the shocks? Does it make a difference if zero or constant restriction is used?*

Exercise 4.42 *Specify and estimate a bivariate VAR using Euro area GDP and M3 growth. Using the restriction that output growth is not contemporaneously affected by money growth shocks, trace out impulse responses and evaluate the claim that money has no medium-long run effect on output. Repeat the exercise assuming that the contemporaneous effect of money growth on output growth is in the interval [-0.5, 1,5] (do this in increments of 0.1 each). What can you say about the medium-long run effect of money growth on output growth in general?*

There is one additional (rank) condition one should typically check: i.e. $rank(\Sigma_e) = rank(\mathcal{A}_0\Sigma_\epsilon\mathcal{A}_0)$ (see Hamilton (1994) for a formal derivation). Intuitively, this restriction rules out that any column of $\mathcal{A}_0$ can be expressed as linear combination of the others. While the rank condition is typically important in large scale SES, it is almost automatically satisfied in small scale VAR identified with economic theory restrictions. When other types of restrictions are employed, the condition should be always checked.

Rank and order conditions are only valid for "local identification". That is, the system may not be identified even though $m(m-1)/2$ restrictions are imposed. This requires experimenting with different initial conditions when estimating the parameters of $\mathcal{A}_0$.

Example 4.25 *Suppose $\Sigma_\epsilon = I$ and that $\mathcal{A}_0^1 = \begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix}$. It is immediate to verify that the likelihood obtained with these two matrices and any positive definite Σ_e is equivalent to the one obtained with the same Σ_ϵ and Σ_e and $\mathcal{A}_0^2 = \begin{bmatrix} 5 & 0.8 \\ 0 & 0.6 \end{bmatrix}$. Clearly the two decompositions have different economic interpretations. Depending on the initial conditions, the maximum can be reached at $\mathcal{A}_0^2$ or $\mathcal{A}_0^1$.*

To estimate the free parameters in (4.35) one typically has two options. The first is to write down the likelihood function of (4.35) (conditional on Σ_e), that is

$$\ln\mathcal{L} = 2\ln|\mathcal{A}_0| + \ln|\Sigma_\epsilon| + trace(\Sigma_\epsilon^{-1}\mathcal{A}_0^{-1}\Sigma_e\mathcal{A}_0^{-1\prime}) \tag{4.36}$$

Maximizing (4.36) with respect to Σ_ϵ and concentrating it out we obtain $2\ln|\mathcal{A}_0| + \sum_{i=1}^m \ln(\mathcal{A}_0^{-1}\Sigma_e\mathcal{A}_0^{-1\prime})_{ii}$. An estimate of the parameters can be found maximizing this expression with respect to the free entries of $\mathcal{A}_0$. Since the concentrated likelihood is nonstandard, maximization is typically difficult. Therefore, it is advisable to get some estimates with a simple method (e.g. a simplex algorithm) and then use these as initial conditions in other algorithms (see chapter 6) to find a global maximum.

A likelihood approach is general and works with both just-identified and overidentified systems. For a just identified system one could also use instrumental variables, as suggested, e.g. by Shapiro and Watson (1988). We describe in a example how this can be done.

Example 4.26 *Consider a bivariate VAR model with inflation and unemployment. Suppose that theory tells us that the structural system (4.34) is*

$$\begin{bmatrix} \pi_t \\ UN_t \end{bmatrix} = \begin{bmatrix} \mathcal{A}_{11}(\ell) & \mathcal{A}_{12}(\ell) \\ \mathcal{A}_{21}(\ell) & \mathcal{A}_{22}(\ell) \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ UN_{t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \alpha_{01} & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}$$

Since $\epsilon_{1t} = e_{1t}$ is predetermined with respect to ϵ_{2t} it can used as an instrument to estimate α_{01}. Therefore choosing as a vector of instruments $z_t = [e_{1t}, e_{1t-1}, \ldots, e_{2t-1}, \ldots]$ joint estimates of the α and $\mathcal{A}(\ell)$ can be obtained, applying the IV techniques described in chapter 5.

4.5.2 Nonstationary VARs

The identification process in non-stationary VAR models is similar but additional identification restrictions are available. Furthermore, the presence of cointegration constraints may change the nature of the order condition.

Let the MA representations of the VAR and of the structure be

$$\Delta y_t = D(\ell)e_t = D(1)e_t + D^*(\ell)\Delta e_t \tag{4.37}$$
$$\Delta y_t = \mathcal{D}(\ell)\mathcal{A}_0\epsilon_t = \mathcal{D}(\ell)(1)\mathcal{A}_0\epsilon_t + \mathcal{D}^*(\ell)\mathcal{A}_0\Delta\epsilon_t \tag{4.38}$$

where $D^*(\ell) \equiv \frac{D(\ell)-D(1)}{1-\ell}$, $\mathcal{D}^*(\ell) \equiv \frac{\mathcal{D}(\ell)-\mathcal{D}(1)}{1-\ell}$ and $\Delta = (1-\ell)$.

In (4.37)-(4.38) we have rewritten the system in two ways: the first is a standard MA; the second exploits the multivariate BN decomposition (see chapter 3). Matching coefficients we have $\mathcal{D}(\ell)\mathcal{A}_0\epsilon_t = D(\ell)e_t$. Separating permanent and transitory components and using in the latter case only contemporaneous restrictions we have

$$\mathcal{D}(1)\mathcal{A}_0\epsilon_t = D(1)e_t \tag{4.39}$$
$$\mathcal{A}_0\Delta\epsilon_t = \Delta e_t \tag{4.40}$$

When y_t is stationary, $\mathcal{D}(1) = D(1) = 0$, (4.39) is vacuous and only (4.40) is available. However, if y_t is integrated the restrictions linking the permanent components of the reduced and of the structural form could also be used for identification. (4.39) is the basis, e.g., for the Blanchard and Quah's decomposition discussed in chapter 3. To obtain estimates of structural parameters we need the same order and rank restrictions. However, the $m(m-1)/2$ constraints could be placed either on (4.39) or (4.40) or both. In this latter case iterative approaches are needed to estimate the free parameters of $\mathcal{A}_0$ and the structural shocks ϵ_t.

Example 4.27 *In a bivariate VAR system imposing (4.39) is simple since only one restriction is needed. Suppose that $\mathcal{D}(1)^{12} = 0$ (i.e. ϵ_{2t} has no long run effect on y_{1t}). If $\Sigma_\epsilon = I$ the three elements of $\mathcal{D}(1)\mathcal{A}_0\Sigma_\epsilon\mathcal{A}_0'\mathcal{D}(1)'$ can be obtained from the Choleski factorization of $D(1)\Sigma_e D(1)'$.*

Exercise 4.43 *Consider the model of example 4.24 and assume that all variables are integrated. Suppose we impose the same 6 restrictions via the long run multipliers $\mathcal{D}(1)\mathcal{A}_0$. Describe how to undertake maximum likelihood estimation of the free parameters.*

Exercise 4.44 *(Gali) Consider a structural model of the form*

$$y_t = \alpha_0 + \epsilon_t^S - \alpha_1(i_t - E_t \Delta p_{t+1}) + \epsilon_t^{IS} \quad (4.41)$$
$$M_t - p_t = \alpha_2 y_t - \alpha_3 i_t + \epsilon_t^{MD} \quad (4.42)$$
$$\Delta M_t = \epsilon_t^{MS} \quad (4.43)$$
$$\Delta p_t = \Delta p_{t-1} + \alpha_4(y_t - \epsilon_t^S) \quad (4.44)$$

where ϵ_t^S is a supply shock; ϵ_t^{IS} is an IS shock; ϵ_t^{MS} is a money supply shock and ϵ_t^{MD} is a money demand shock, GDP_t is output, P_t prices, i_t the nominal interest rate and M_t money. Identify these shocks from a VAR with $(\Delta GDP_t,\ \Delta i_t,\ i_t - \Delta p_t,\ \Delta M_t - \Delta p_t)$ using Euro area data and the following restrictions: (i) only supply shocks have long run effects on output, (ii) money demand and money supply shocks have no contemporaneous effects on ΔGDP, (iii) money demand shocks have no contemporaneous effect on the real interest rate. Trace out the effects of a money supply shock on interest rates and output.

When some of the variables of the system are cointegrated, the number of permanent structural shocks is lower than m. Therefore, if long run restrictions are used, one only needs $(m-m_1)(m-m_1-1)/2$ constraints to identify all m shocks where m_1 is the number of common trends (rank of $\mathcal{D}(1) = m - m_1$).

Example 4.28 *As shown in exercise 3.4 of chapter 3, a RBC model driven by integrated technology shocks implies that all variables are integrated but $\frac{C_t}{GDP_t}$ and $\frac{Inv_t}{GDP_t}$ are stationary. Consider a trivariate VAR with $\Delta gdp_t, c_t - gdp_t, inv_t - gdp_t$ where lower case letters indicate logarithms of the variables. Since the system has two cointegrating vectors, there is one permanent shocks and two transitory ones and $(1,1,1)'\epsilon_t = D(1)e_t$ identifies the permanent shock. If all structural shocks are orthogonal we need one extra restriction to identify the two transitory disturbances - for example, we could assume a Choleski structure.*

Exercise 4.45 *(Shapiro and Watson) Consider a bivariate system $\Delta y_t = D(\ell)e_t$ where $e_t \sim (0, \Sigma_e)$ and let the structural model be $\Delta y_t = \mathcal{D}(\ell)\epsilon_t$ where $\epsilon_t \sim (0, I)$ and $\mathcal{D}(1)$ is lower triangular. Show that $D(1) = \mathcal{D}(1)\Sigma_e^{0.5}$ is lower triangular. Show that to estimate $\mathcal{D}(1)$ and $\mathcal{D}_0$ one could normalize the system $\Delta y_t^* = \Sigma_e^{-0.5}\Delta y_t$ and run a regression of $\Delta y{*}_{1t}$ on q lags of Δy_{1t}^* and the current and $q-1$ lags of Δy_{2t}^* and a regression of $\Delta y{*}_{2t}$ on q lags of Δy_{1t}^* and Δy_{2t}^*, instrumenting current values with $\Delta y_{t-j}^*, j = 1, 2, \ldots$.*

4.5.3 Alternative identification schemes

The identification of structural shocks is, in general, a highly controversial enterprise because researchers imposing different identifying assumptions may reach different conclusions about interesting economic questions (e.g. the sources of business cycle fluctuations). However, an embarrassing uniformity has emerged over the last 10 years since identifying restrictions have become largely conventional and unrelated to the class of DSGE models described in section 2. Criticisms to the nature of identification process have repeatedly appeared in the literature. For example, Cooley and LeRoy (1985) criticize Choleski decompositions because contemporaneous recursive structures are hard to obtain in general equilibrium models. Faust and Leeper (1997) argue that long run restrictions are unsatisfactory as they may exclude structures which generate perfectly reasonable short run dynamics but fail to satisfy long run constraints by infinitesimal amounts. Cooley and Dwyer (1998) indicate that long restrictions may also incompletely disentangle permanent and transitory disturbances. Canova and Pina (2004) show that standard DSGE models almost never provide the zero restrictions employed to identify monetary disturbances in structural VAR systems and that misspecification of the features of the underlying economy can be substantial.

Figure 4.3 shows the extent of the problem when a working capital model, similar to the one presented in chapter 2, with either a partial accommodative (PA) or a Taylor type (FB) rule for monetary policy is used to generate data and monetary shocks are identified in the VAR for simulated data either with a Choleski scheme (CEE), with variables in the order $(GDP_t, p_t, i_t, \frac{M_t}{p_t})$ or via an overidentified structure (SZ) where i_t responds only to $\frac{M_t}{p_t}$. The straight line is the response produced by the model, the dotted ones one standard error bands produced by the VAR. Note that a Choleski system correctly recognizes the policy input when a Taylor rule is used, while the overidentified model correctly characterizes the policy rule in the partial accommodative case. Misspecification is pervasive even when one correctly selects the inputs of the monetary policy rule. For example, a Choleski scheme fails to capture the persistent response of real balances to interest rate increases and produces perverse output responses (first box, first column) while a price puzzle is produced (second row, first and third boxes).

We would like to stress that the patterns presented in figure 4.3 are not obtained because the model is unrealistic or the parametrization "crazy". As shown in Canova and Pina (2004) a sticky price, sticky wage model, parametrized in a standard way, produces similar outcomes. The problem is that a large class of DSGE structures do not display the zero restrictions imposed by the two identification schemes (in particular, that output and prices have a Wold causal structure and do not respond instantaneously to policy shocks). Therefore, misspecification results even when the policy rule is correctly identified.

To produce a more solid bridge between DSGE models and VARs, a new set of identification approaches have emerged. Although justified with different arguments, the procedures of Faust (1998), Uhlig (2003) and Canova and De Nicoló (2002) have one feature in common: they do not use zero-type of restrictions. Instead, they achieve identification restricting the sign (and/or shape) of structural responses. Restrictions of this type are often used by applied researchers informally: for example, monetary shocks which do not generate a liquidity effect (e.g. opposite comovements in interest rate and money) are typically discarded and the zero restrictions reshuffled in the hope to produce the required outcome. One advantage of these approaches is to make restrictions of this type explicit.

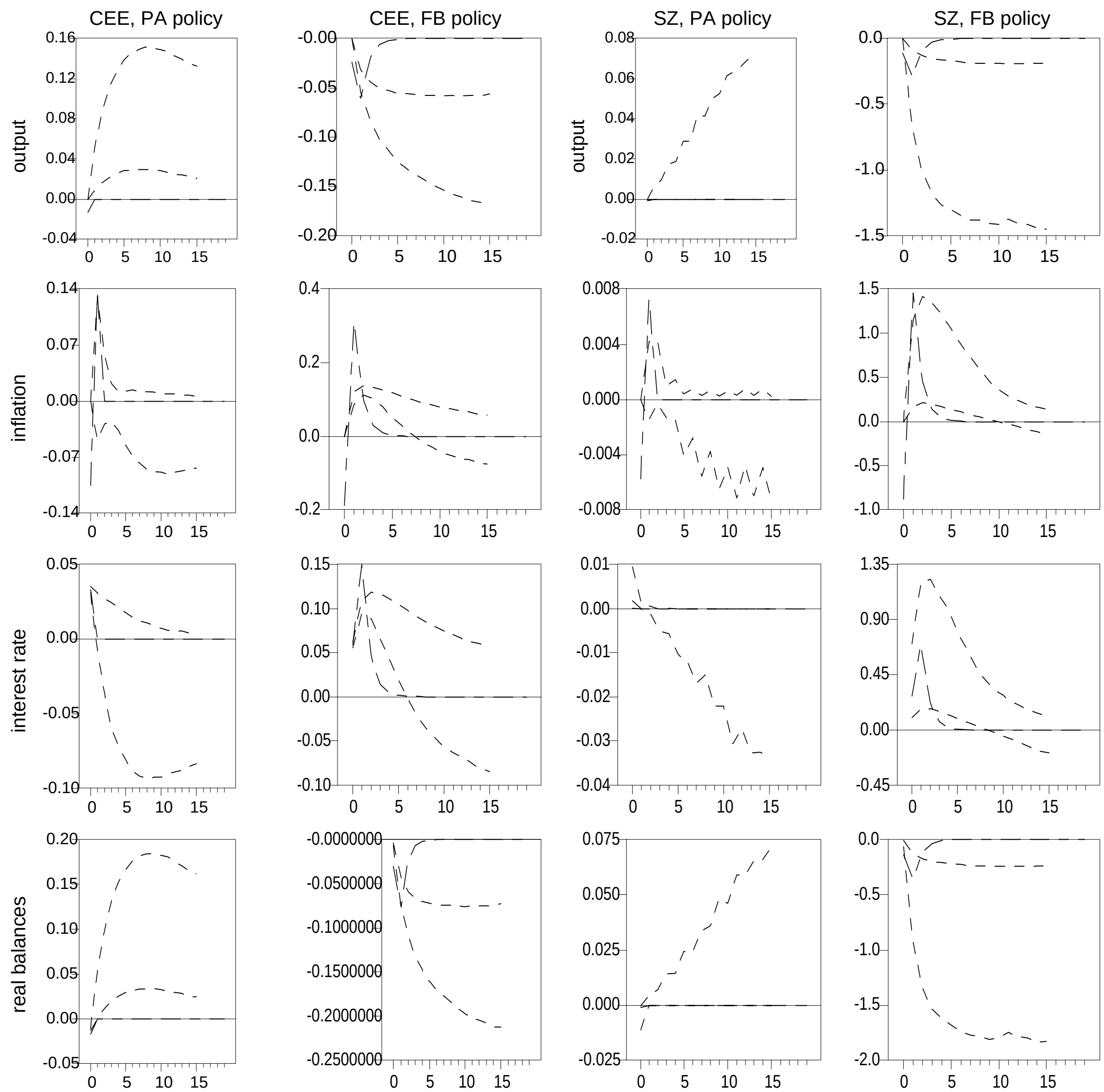

Figure 4.3: Impulse responses to monetary shocks, Working capital model

Sign restrictions are enticing. While (log)-linearized versions of DSGE models seldomly deliver the $m(m-1)/2$ set of zero restrictions needed to recover m structural shocks, they contain a large number of sign restrictions usable for identification purposes.

Example 4.29 *(Technology shocks) All RBC-type models examined in chapter 2 have the feature that positive technology disturbances increase output, consumption, and investment either instantaneously or with a short lag, while prices and interest rates decline as the aggregate supply curve shifts to the right. Therefore, such a class of models suggests that technology disturbances can be identified via the restriction that in response to positive shocks real variables increase and prices decrease, either contemporaneously or with a lag.*

Example 4.30 *(Monetary shocks) Several of the models of chapter 2 have the feature that policy driven increases in the nominal interest rates reduce real balances instantaneously and induce a fall in prices. Hence, contemporaneous (and lagged) comovements of real balances, prices and nominal interest rates can be used to identify monetary disturbances.*

The restrictions of examples 4.29 and 4.30 could be imposed on two or more variables, at one or more horizons. In other words, we can "weakly" or "strongly" identify the shocks. To maintain comparability with other structural VARs, weak forms of identification will be typically preferred. However, one should be aware that restrictions which are too weak may be unable to distinguish shocks with somewhat similar features, i.e. labor supply and technology shocks.

It is relatively complicated to impose sign restrictions directly on the coefficients of the VAR, as this requires maximum likelihood estimation of the full system under inequality constraints. However, it is relatively easy to do it ex-post on impulse responses. For example, as in Canova De Nicolo'(2002), one could estimate $A(\ell)$ and Σ_e from the data using OLS and orthogonalize the reduced form shocks using, e.g. an eigenvalue-eigenvector decomposition, $\Sigma_e = \mathcal{P}\mathcal{V}\mathcal{P}' = \tilde{\mathcal{P}}\tilde{\mathcal{P}}'$ where $\mathcal{P}$ is a matrix of eigenvectors and $\mathcal{V}$ is a diagonal matrix of eigenvalues. This decomposition does not have any economic content, but produces uncorrelated shocks without employing zero restrictions. For each of the orthogonalized shocks one can check if the identifying restrictions are satisfied. If there is one such a shock, the process terminates. If there is more than one shock satisfying the restrictions, one may want to increase the number of restrictions (either across variables or across leads and lags) until one candidate remains or take an average. Practical experience suggests that contemporaneous and/or one lag restrictions suffice to produce a unique set of shocks.

If no shock satisfies the restrictions, the non-uniqueness of the MA representation can be used to provide alternative structural shocks. In fact, for any $\mathcal{H}$ with $\mathcal{H}\mathcal{H}' = I, \ \Sigma_e = \tilde{\mathcal{P}}\tilde{\mathcal{P}}' = \tilde{\mathcal{P}}\mathcal{H}\mathcal{H}'\tilde{\mathcal{P}}'$. Hence, one can construct a new decomposition using $\tilde{\mathcal{P}}\mathcal{H}$ and examine if the shocks produce the required pattern.

The only remaining practical question is how to choose $\mathcal{H}$ and how to systematically explore the space of MA representations, which is infinite dimensional, if this is of interest. Canova and de Nicoló choose $\mathcal{H} = \mathcal{H}(\omega), \ \omega \in (0, 2\pi)$ and search the space of $\mathcal{H}$ by varying ω on a grid. Here $\mathcal{H}$ are matrices which rotate the columns of $\mathcal{P}$ by an angle ω.

Example 4.31 *Consider a bivariate system with unemployment and inflation and suppose that a basic eigenvector-eigenvalue decomposition has not produced a shock which produced contemporaneously negative comovements in inflation and unemployment. Set* $\mathcal{H}(\omega) = \begin{bmatrix} cos(\omega) & -sin(\omega) \\ sin(\omega) & cos(\omega) \end{bmatrix}$. *Then we can trace out all possible MA representations for the bivariate system, varying* $\omega \in (0, 2\pi)$.

In larger scale systems, rotation matrices are more complex.

Exercise 4.46 *Consider a four variable VAR. How many matrices rotating two or pairs of two columns exist? How would you explore the space of rotations simultaneously flipping the first and the second column together with the third and the fourth?*

When m is of medium size, the matrix $\mathcal{H}$ has the following form

$$\mathcal{H}_{i,i'}(\omega) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \cos(\omega) & \dots & -\sin(\omega) & 0 \\ \vdots & \vdots & \vdots & 1 & \vdots & \vdots \\ 0 & 0 & \sin(\omega) & \dots & \cos(\omega) & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$0 < \omega \leq 2\pi$ where the index (i, i') indicates that columns i and i' are rotated by the angle ω. Let $\mathcal{Z}(\mathcal{H}_{i,i'}(\omega)$ be the space of ortonormal rotation matrices where, given ω, each i, i' element has proability $\frac{2}{m(m-1)}$. Then the following search algorithm could be used to explore the space of identifications.

Algorithm 4.6

1) *Draw* ω^l *from* $(0, 2\pi)$. *Draw* $\mathcal{H}_{i,i'}(\omega^l)$ *from* $\mathcal{Z}(\mathcal{H}_{i,i'}(\omega^l)$.

2) *Used* $\mathcal{H}(i,i')(\omega^l)$ *to compute* ϵ_t *and* $\mathcal{A}(\ell)$. *Check whether restrictions are satisfied in response to* $\epsilon_{it}, i = 1, \dots m$. *If they are keep the draw, if they are not, drop the draw.*

3) *Repeat 1) and 2) unless L draws satisfying the restrictions are found. Report percentile response bands.*

Note that, by continuity, it is typical to find an interval (ω_1, ω_2) which produces a shock with the required characteristics. Since within this interval the dynamics produced by structural shocks are similar, one can average statistics for all the shocks in the interval or choose, say, the shock corresponding to the median point of the interval or keep all of them, as we have done in algorithm 4.6. We have already discussed what to do if more than one ϵ_{it} satisfies the restrictions for a given ω^l and $\mathcal{H}_{i,i'}(\omega^l)$. At times one may find disjoint intervals where one or more shocks satisfy the restrictions. In this case it is a good idea to graphically inspect the outcome since responses may not be economically meaningful (for example, a shock may imply an output elasticity of 50). When visual inspection fails, increasing the number of restrictions is typically sufficient to eliminate "unreasonable" intervals.

Exercise 4.47 *Provide a Monte Carlo algorithm to construct standard error bands for structural impulse responses identified with sign restrictions which takes into account parameter uncertainty.*

Example 4.32 *Figure 4.4 presents the responses of industrial output, prices and M1 in the US in response to a monetary policy shock. In the right column are the 68% impulse response bands obtained requiring that a nominal interest rate increase must be accompanied by a liquidity effect - a contemporaneous decline in M1. In the left column are the 68% impulse response bands obtained with the Choleski system where the interest rate is assumed to contemporaneously react to industrial output and prices but not to money.*

Clearly, the standard identification has unpleasant outcomes: point estimates of money, output and prices are all positive after the shock even though the increase is not significant. With sign restrictions, output and prices significantly decline after a contractionary shock and they do so for about 5 months. Note that in both systems no measure of commodity prices is used.

4.6 Problems

While popular among applied researchers, VARs are not free of problems and a number of common pitfalls should be avoided when interpreting the results.

First, one should be aware of time aggregation problems. As Sargent and Hansen (1991), Marcet (1991) and others have shown time aggregation may make inference difficult. In fact, if agents take decisions every τ periods but an econometrician observes data only every $j\tau,\ j > 1$, the statistical model used by the econometrician (with data sampled at every $j\tau$) may have little to do with the one produced by agents' decisions. For example, the MA traced out by the econometrician is not necessarily the MA of the model sampled every j period, but a complex function of all MA coefficients from that point on to infinity.

Example 4.33 *Marcet (1991) showed that if agents' decisions are taken in continuous time, continuous and discrete time MA representations are related via* $D_j = [d \diamond v'_{-j}][v \diamond v_0]$ *where* d *is the moving average in continuous time,* $\diamond$ *indicates the convolution operator and* $v_j = d_j - b \times (d_j|D)$ *is the forecast error in predicting* d_j *using the information contained in the discrete time MA coefficients,* b *is a constant and* $j = 1, 2, \ldots, \tau$*. Hence a humped-shaped monthly response can easily be transformed into a smoothly declining quarterly response (see figure 4.5).*

One important special case obtains when agents' decisions generate a VAR(1) for the endogenous variables. In that case, the MA coefficients of, say, a quarterly model are

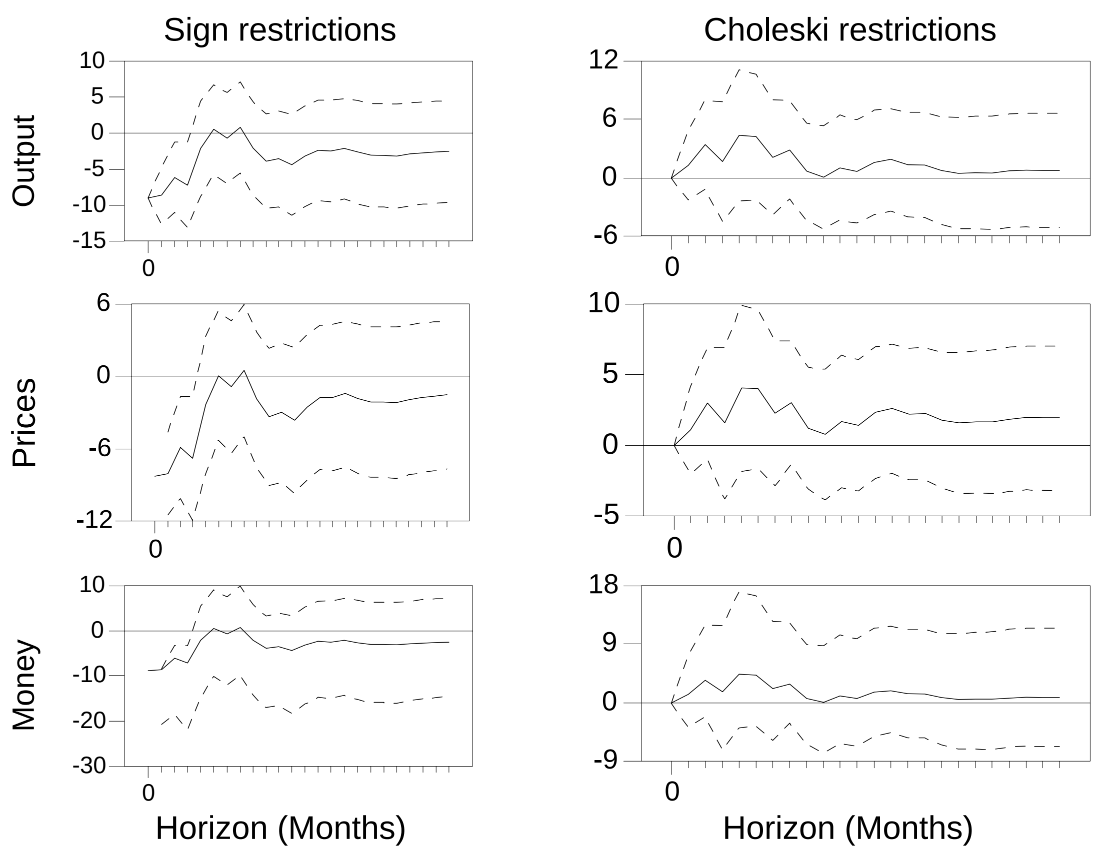

Figure 4.4: Responses to a US policy shock, 1964:1-2001:10

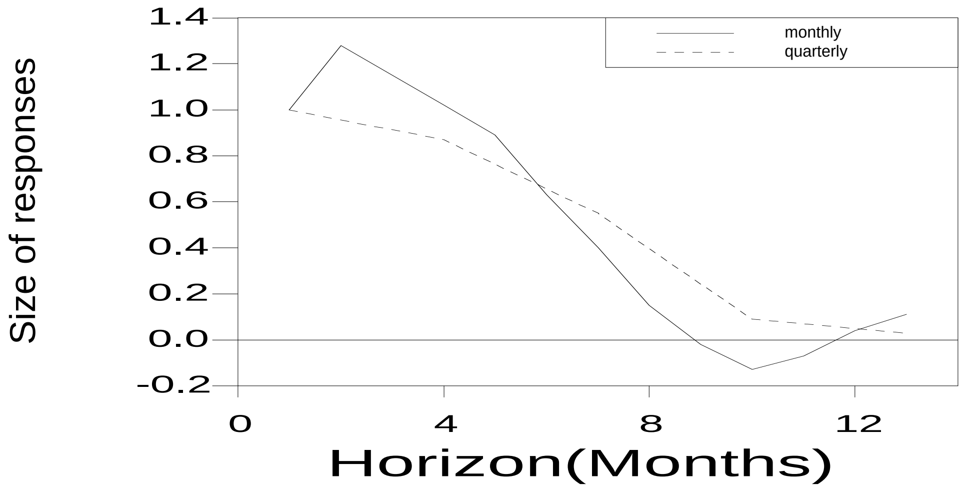

Figure 4.5: Quarterly and Monthly MA representations

the same as the quarterly sampled version of MA coefficients of a monthly model. While log-linear or quadratic approximate solutions to many DSGE models do deliver VAR(1) models, one should be aware that models with e.g. habit in consumption or quadratic costs of adjustment to investments, produce more complicated dynamics and therefore may face important aggregation problems.

Exercise 4.48 *Consider a RBC model perturbed by technology and government expenditure disturbances. Suppose that* $g_t = T_t$ *where* T_t *are lump-sum taxes and that the utility function depends on current and lagged leisure, i.e.* $U(c_t, N_t, N_{t-1}) = \ln c_t + (N_t - \gamma N_{t-1})^{\varphi_n}$.
i) Calculate the linearized decision rules after you have appropriately parametrized the model at quarterly and annual frequencies. Compare the MA coefficients of the annual model with the annual sampling of the MA of the quarterly model.
ii) Simulate consumption and output for the two specifications. Sample at annual frequencies the quarterly data and compare the autocovariance functions. Does aggregation hold?
iii) Set $\gamma = 0$ *and assume that both capital and its utilization enter in the production function as in exercise 2.10 of chapter 2. Repeat steps i)- ii) and comment on the results.*

Exercise 4.48 suggests that one way to detect possible aggregation problems is to run VARs at different frequencies and compare their ACF or their MA representations. If differences are detected, given the same amount of data, aggregation is likely to be a problem.

A second important problem has to with the dimensionality of the VAR. Small scale VAR models are typically preferred by applied researchers since parameter estimates are more precise (and impulse response bands are tighter) and because identification of the structural shocks is easier. However, small scale VARs are prone to misspecification. For example, there may be important omitted variables and shocks may be confounded or misaggregated. As Braun and Mittnik (1993), Cooley and Dwyer (1998), Canova and Pina (2004) have shown, important biases may result. To illustrate the effects of omitting variables we make use of the following result:

Result 4.3 *In a bivariate VAR(q):* $\begin{bmatrix} A_{11}(\ell) & A_{12}(\ell) \\ A_{21}(\ell) & A_{22}(\ell) \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$, *the univariate representation for* y_{1t} *is* $[A_{11}(\ell) - A_{12}(\ell)A_{22}(\ell)^{-1}A_{21}(\ell)]y_{1t} = e_{1t} - A_{12}(\ell)A_{22}(\ell)^{-1}e_{2t} \equiv \upsilon_t$

Example 4.34 *Suppose the true DGP has* $m = 4$ *variables but an investigator incorrectly estimates a bivariate VAR (there are three of these models). Using result 4.3 it is immediate to see that the system with, e.g., variables 1 and 3, has errors of the form* $\begin{bmatrix} \upsilon_{1t} \\ \upsilon_{2t} \end{bmatrix} \equiv \begin{bmatrix} e_{1t} \\ e_{3t} \end{bmatrix} - \mathsf{Q}_1(\ell)\mathsf{Q}_2^{-1}(\ell) \begin{bmatrix} e_{2t} \\ e_{4t} \end{bmatrix}$ *where* $\mathsf{Q}_1(\ell) = \begin{bmatrix} A_{12}(\ell) & A_{14}(\ell) \\ A_{32}(\ell) & A_{34}(\ell) \end{bmatrix}$ $\mathsf{Q}_2(\ell) = \begin{bmatrix} A_{22}(\ell) & A_{24}(\ell) \\ A_{42}(\ell) & A_{44}(\ell) \end{bmatrix}$.
From this one can verify that:

- *If the true system is a VAR(1), a model with* $m_1 < m$ *variables is a VAR(*∞*).*
- *If* e_t*'s are contemporaneously and serially uncorrelated,* υ_t*'s are typically contemporaneously and serially correlated.*
- *Two small scale VAR, both with* $m_1 < m$ *variables, may have different innovations.*
- υ_t *is a linear combination of current and past* e_t*. The timing of innovations is preserved if the* m_1 *included variables are Granger causally prior to the* $m - m_1$ *omitted ones (i.e. if* $\mathsf{Q}_1(\ell) = 0$*).*

Several implications one can draw from example 4.34. First, if relevant variables are omitted a long lag length is needed to whiten the residuals. While long lags do not always indicate misspecification (for example, if y_t is nearly non-stationary long lags are necessary to approximate its autocovariance function), care should be exercised in drawing inference in such models. Second, two researchers estimating small scale models with different variables may obtain different structural innovations, even if the same identification restrictions are used. Finally, innovation accounting exercises when variables are omitted may misrepresent the timing of the responses to structural shocks.

Exercise 4.49 *(Giordani) Consider a sticky price model composed of an output gap (*$gap_t = gdp_t - gdp_t^P$*) equation, a potential output (*gdp_t^P*) equation, a backward looking Phillips curve (normalized on* π_t*) and a Taylor rule of the type*

$$gap_{t+1} = a_1 gap_t - a_2(i_t - \pi_t) + \epsilon_{t+1}^{AD} \quad (4.45)$$

$$gdp_{t+1}^P = a_3 gdp_t^P + \epsilon_{t+1}^P \quad (4.46)$$

$$\pi_{t+1} = \pi_t + a_4 gdp_t^g + \epsilon_{t+1}^{CP} \quad (4.47)$$

$$i_t = a_5 \pi_t + a_6 gdp_t^g + \epsilon_{t+1}^{MP} \quad (4.48)$$

The last equation has an error term (monetary policy shock) since the central bank may not always follow the optimal solution to its minimization problem. Let $var(\epsilon^i_{t+1}) = \sigma^2_i$, $i =$ *AD, P, CP, MP and assume that the four shocks are uncorrelated with each other.*
(i) Argue that contractionary monetary policy shocks have one period lagged (negative) effects on output and two periods lagged (negative) effects on inflation. Show that monetary policy actions do not Granger cause gdp^P_t *for all* t.
(ii) Derive a VAR for $[gdp_t,\ gdp^P_t,\ \pi_t,\ i_t]$. *Display the matrix of impact coefficients.*
(iii) Derive a representation for a three variable system $[gdp_t,\ \pi_t,\ i_t]$ *(Careful: when you solve out potential output from the system the remaining variables do not follow a VAR any longer). Label the three associated shocks* $e_t = [e^{AD}_t, e^{CP}_t, e^{MP}_t]$ *and their covariance matrix* Σ_e. *Show the matrix of impact coefficients in this case.*
(iv) Show that $var(e^{AD}_t) > var(\epsilon^{AD}_t)$; $var(e^{MP}_t) > 0$ *even when* $\epsilon^{MP}_t = 0\ \forall t$ *and that* $corr(e^{MP}_t, \epsilon^P_t) < 0$. *Show that in a trivariate system, contractionary monetary policy shocks produce positive price responses (compare this with what you have in i))*
(v) Intuitively explain why the omission of potential output from the VAR causes problems.

It is worthwhile to look at omitted variable problems from another perspective. Suppose the structural MA for a partition with $m_1 < m$ variables of the true DGP is

$$y_t = D(\ell)\epsilon_t \tag{4.49}$$

where ϵ_t is an $m \times 1$ vector, so that $D(\ell)$ is $m_1 \times m$ matrix $\forall \ell$. Suppose a researcher specifies a VAR with $m_1 < m$ variables and obtains an MA of the form:

$$y_t = \tilde{D}(\ell)e_t \tag{4.50}$$

where e_t is an $m_1 \times 1$, and $\tilde{D}(\ell)$ is a $m_1 \times m_1$ matrix $\forall \ell$. Matching (4.49) and (4.50) one obtains $\tilde{D}(\ell)e_t = D(\ell)\epsilon_t$ or letting $D^{\ddagger}(\ell)$ be a $m_1 \times m$ matrix

$$D^{\ddagger}(\ell)\epsilon_t = e_t \tag{4.51}$$

As shown by Faust and Leeper (1997) (4.51) teaches us an important lesson. Assume that there are m^a shocks of one type and m^b shocks of another, $m^a + m^b = m$, and that $m_1 = 2$. Then $e_{it}, i = 1, 2$ recovers a linear combination of shocks of type $i' = a, b$ only if $D^{\ddagger}(\ell)$ is block diagonal and only correct current shocks if $D^{\ddagger}(\ell) = D^{\ddagger}, \forall \ell$ and block diagonal. In all other cases, true innovations are mixed up in estimated structural shocks.

Note that these problems have nothing to do with estimation or identification. Misspecification occurs because a VAR(q) is transformed in a VARMA(∞) whenever a variable is omitted and this occurs even when the MA representation of the small scale model is known.

Example 4.35 *Suppose the true structural model has* $m = 4$ *shocks, that there are two supply and two demand shocks, and that an investigator estimates a bivarate VAR. When would the two estimated structural shocks correctly aggregate shocks of the same type? Using*

(4.51) we have $\begin{bmatrix} D^{\ddagger}_{11}(\ell) & D^{\ddagger}_{12}(\ell) & D^{\ddagger}_{13}(\ell) & D^{\ddagger}_{14}(\ell) \\ D^{\ddagger}_{21}(\ell) & D^{\ddagger}_{22}(\ell) & D^{\ddagger}_{23}(\ell) & D^{\ddagger}_{24}(\ell) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \end{bmatrix} = \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$. *Hence,* e_{1t} *will recover only type 1 shocks if* $D^{\ddagger}_{13}(\ell) = D^{\ddagger}_{14}(\ell) = 0$ *and* e_{2t} *will recover type 2 shocks if* $D^{\ddagger}_{21}(\ell) = D^{\ddagger}_{22}(\ell) = 0$*. Furthermore,* e_{1t} *recovers current type 1 shocks if* $D^{\ddagger}_{13}(\ell) = D^{\ddagger}_{14}(\ell) = 0$ *and* $D^{\ddagger}_{ii'}(\ell) = D^{\ddagger}_{ii'}$, $\forall \ell$.

The conditions required for correct aggregation are therefore somewhat strong. As it is shown in the next example, they are not satisfied in at least one type of DSGE model. It is likely that such a problem also appears in other models macroeconomists currently use.

Example 4.36 *We simulate data from a version of the working capital economy of example 2.14 of chapter 2 with a permanent (technology) disturbance and temporary labor supply, monetary and government expenditure shocks. Monetary policy is characterized by a Taylor rule. Using output and employment data we estimate a bivariate VAR and extract a permanent and a transitory shock where the latter is identified by the requirement that it has no long run effects on output. Table 4.3 presents the estimated coefficients of a distributed lag regression of two of the theoretical shocks on the estimated ones. In parenthesis are t-statistics. The last column presents the p-value of a F-test excluding monetary disturbances from the first equation and technological disturbances from the second. Estimated supply shocks mix both current and lagged monetary and technology disturbances while for estimated demand shocks current and lagged monetary disturbances matter but only current technology disturbances are important. This pattern is independent of the sample size.*

	Technology Shocks			*Monetary Shocks*			*P-value*
	0	*-1*	*-2*	*0*	*-1*	*-2*	
Estimated	*1.20*	*0.10*	*0.04*	*0.62*	*-0.01*	*-0.11*	
Supply Shocks	*(80.75)*	*(6.71)*	*(3.05)*	*(45.73)*	*(-0.81)*	*(-8.22)*	*0.000*
Estimated	*-0.80*	*0.007*	*0.08*	*0.92*	*-0.48*	*-0.20*	
Demand Shocks	*(-15.27)*	*(0.13)*	*(1.59)*	*(19.16)*	*(-10.03)*	*(-4.11)*	*0.000*

Table 4.3: Regressions on simulated data

Exercise 4.50 *(Cooley and Dwyer) Simulate data from a CIA model where a representative agent maximizes* $E_0 \sum_t \beta^t [a \ln c_{1t} + (1-a) \ln c_{2t} - \vartheta_N N_t]$ *subject to* $p_t c_{1t} \leq M_t + (1+i_t)B_t + T_t - B_{t-1}$ *and* $c_{1t} + c_{2t} + inv_t + \frac{M_{t+1}}{p_t} + \frac{B_{t+1}}{p_t} \leq w_t N_t + r_t K_t + \frac{M_t}{p_t} + (1+i_t)\frac{B_t}{p_t} + \frac{T_t}{p_t}$ *where* $K_{t+1} = (1-\delta)K_t + inv_t$, $y_t = \zeta_t K_t^{1-\eta} N_t^{\eta}$, $\ln \zeta_t = \rho_\zeta \ln \zeta_{t-1} + \epsilon_{1t}$, $\ln M^s_{t+1} = \ln M^s_t + \ln M^g_t$, M^g_t *is a constant and* $\rho_\zeta = 0.99$ *(you are free to choose the other parameters, but motivate your choices). Consider a bivariate system with output and hours and verify that output has a unit root but hours does not. Using the restriction that demand shocks have no long run effects on output, plot output and hours responses in theory and in the VAR. Is there any feature of the theoretical economy which is distorted?*

In section 5 we have seen that for a just identified structural model, a two-step estimation approach is equivalent to a direct 2SLS approach on the structural system. Since structural shocks depend on the identification restrictions, we may have situations where a 2SLS approach produces "good' estimators, in the sense that they nicely correlate with the structural shocks they instrument for, and situations where they are bad. Cooley and Dwyer (1998) present an example where, by changing the identifying restrictions, the correlation of the instruments with the structural shocks go from high to very low, therefore resulting in instrumental variables failures (see chapter 5). Hence, if such a problem is suspected, a maximum likelihood approach should be preferred.

Finally, we would like to mention once again that there are several economic models which generate non-Wold decompositions, see e.g. Leeper (1991), Quah (1990), Hansen and Sargent (1991). Hence examining these models with Wold decompositions is meaningless. When a researcher suspects that this is a problem Blaascke factors should be used to construct non-fundamental structural MA representations. Results do depend on the representations used. For example, Lippi and Reichlin (1993) present a non-Wold version of Blanchard and Quah (1989)'s model which gives opposite conclusions regarding the relative importance of demand and supply shocks in generating business cycle fluctuations.

Exercise 4.51 *(Quah) Consider a three equations permanent income model*

$$\begin{aligned} c_t &= rWe_t \\ We_t &= sa_t + [(1+r)^{-1}\sum_j (1+r)^{-j} E_t GDP_{t+j} \\ sa_{t+1} &= (1+r)sa_t + GDP_t - c_t \end{aligned} \tag{4.52}$$

where c_t *is consumption,* We_t *is wealth,* r *is the (constant) real rate,* sa_t *are savings and* $\Delta GDP_t = D(\ell)\epsilon_t$ *is the labor income. Show that a bivariate representation for consumption and output is* $\begin{bmatrix} \Delta GDP_t \\ \Delta c_t \end{bmatrix} = \begin{bmatrix} A_1(\ell) & (1-\ell)A_0(\ell) \\ A_1(\beta) & (1-\beta)A_0(\beta) \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{0t} \end{bmatrix}$ *where* $\beta = (1+r)^{-1}$, e_{1t} *is a permanent shock and* e_{0t} *a transitory shock. Find* $A_1(\ell)$ *and* $A_0(\ell)$*. Show that if* $\Delta Y_t = \epsilon_t$*, the representation collapses to* $\begin{bmatrix} \Delta GDP_t \\ \Delta c_t \end{bmatrix} = \begin{bmatrix} 1 & (1-\ell) \\ 1 & (1-\beta) \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{0t} \end{bmatrix}$*. Show that the determinant of the matrix vanishes at* $\ell = \beta < 1$ *so that the MA representation for consumption and income is non-fundamental. Show that the fundamental MA is* $\begin{bmatrix} \Delta GDP_t \\ \Delta c_t \end{bmatrix} = b(\beta)^{-1} \begin{bmatrix} (2-\beta)(1-\frac{1-\beta}{2-\beta}\ell) & (1-\beta\ell) \\ 1+(1-\beta)^2 & 0 \end{bmatrix} \begin{bmatrix} \tilde{e}_{1t} \\ \tilde{e}_{0t} \end{bmatrix}$*,* $var(\tilde{e}_{0t})=var(\tilde{e}_{1t}) = 1$*.*

4.7 Validating DSGE models with VARs

VARs are extensively used to summarize those conditional and unconditional moments that "good" models should be able to replicate. Generally, informal comparisons between the models and the data are performed. At times, the model's statistics are compared with 68 or 95% bands for the statistics of the data (see e.g. Christiano Eichenbaum and Evans

(2001)). There conclusions about the quality of the model rest on whether model's statistics are inside or outside these bands for a number of variables. If parameter uncertainty is allowed for, comparison of posterior distributions is possible (see chapters 7 and 11).

However, DSGE theories can be more directly tested via VARs. For example, in Canova, Pagan and Finn (1994) theoretical cointegration restrictions coming from a RBC model driven by permanent technology shocks are imposed on a VAR and tested using standard statistical tools. Their point of view can be generalized and the applicability of their idea extended if qualitative implications, which are more robust than quantitative ones, are used to restrict the data and if restrictions are used for identification rather than for estimation.

DSGE models are misspecified in the sense that they are too simple to capture the complex probabilistic nature of the data. Hence, it may be senseless to compare their outcomes with the data: if one looks hard enough and data is abundant, statistically or economically large deviations can always be found. Both academic economists and policymakers use DSGE models to tell stories about how the economy responds to unexpected movements in exogenous variables. Hence, there may be substantial consensus in expecting output to decline after an unexpected interest rate increase but considerable uncertainty about the size of the impact and the timing of the output responses. The techniques described in chapter 5 to 7 have hard time to deal with this uncertainty. Estimation and testing with maximum likelihood requires the whole model to be the correct DGP (up to uncorrelated measurement errors), at least under the null. Generalized methods of moments and simulation estimators can be tailored to focus only on those aspects where misspecification could be smaller (e.g. the Euler equation, or the great ratios). However, estimation and validation still requires that these aspects of the model are quantitatively correct under the null. When one feels comfortable only with the qualitative implications of a model and is not willing to (quantitatively) entertain a part or the whole of it as a null hypothesis, the approach described in section 5.3 can be used to formally evaluate the fit of any model or the relative merit of two competitor models.

The method agrees with the minimalist identification philosophy underlying VARs. In fact, one can use some of the least controversial qualitative implications of a model to identify structural shocks in the data. Once shocks in data and the model are forced to have qualitatively similar features, the dynamic discrepancy between the two in the dimensions of interest can be easily examined. We summarize the main features of the approach in the next algorithm.

Algorithm 4.7

1) *Find qualitative, robust implications of a class of models.*

2) *Use (a subset of) these implications to identify shocks in the actual data. Stop validation if data does not conform to the qualitative robust restrictions of the model.*

3) *If theoretical restrictions have a data counterpart,* qualitatively *evaluate the model (use e.g. sign and shape of responses to shocks, the pattern of peak responses, etc.)*

4) *Validate* qualitatively *across models if more than one candidate is available.*

5) *If results in 3) and 4) are satisfactory, and policy analyses need to be performed, compare model and data* quantitatively.

6) *Repeat 2)-5) using other robust implications of the model(s), if needed.*

7) *If mismatch between theory and data is relevant, alter the model so as to maintain restrictions in 1) satisfied and repeat 3) and/or 5) to evaluate improvements. Otherwise, proceed to policy analyses.*

Few comments on algorithm 4.7 are in order. In 1) we require theoretical restrictions to be robust, that is independent of parametrization and/or of the functional forms of primitives. The idea is avoid restrictions which emerge only in special cases of the theory. In the second step we force certain shocks in the data and in the model to be qualitatively similar. In steps 2) to 7) evaluation is conducted at different levels: first, we examine whether the restrictions are satisfied in the data; second, we evaluate qualitative dynamic features of the model; finally, quantitative properties are considered. Qualitative evaluation should be considered a prerequisite to a quantitative one: many models can be discarded using the former alone. Also, to make the evaluation meaningful economic measures of discrepancy, as opposed to statistical ones, should be used.

The algorithm is simple, easily reproducible, and computationally affordable, particularly in comparison to ML or the Bayesian methods we discuss in Chapter 11; it can be used when models are very simplified descriptions of the actual data; and can be employed to evaluate one or more dimensions of the model. In this sense, it provides a flexible, limited information criteria which can be made more or less demanding, depending on the desires of the investigator. We illustrate the use of algorithm 4.7 in an example.

Example 4.37 *We take a working capital (WK) and a sticky price (SP) model, with the idea of studying the welfare costs of employing different monetary rules. We concentrate on the first step of the exercise, i.e. in examing which model is more appropriate to answer the policy question.*

Canova (2002) shows that these two models produce a number of robust sign restrictions in response to technology and monetary policy shocks. For example, in response to a policy disturbance the WK economy generates negative comovements of inflation and output, of inflation and real balances, and of inflation and the slope of the term structure and positive comovements of output and real balances. In the SP economy, the correlation between inflation and output is positive contemporaneously and for lags of output and negative for leads of output. The one between inflation and real balances is negative everywhere, the one of output and real balances is positive for lags of real balances and negative contemporaneously and for leads of real balances. Finally, the correlation of the slope of the term structure with inflation is negative everywhere. One could use some or all of these restrictions to characterize monetary shocks in the two models. Here we select restrictions on the contemporaneous cross correlation of output, inflation and the slope of the term structure for *the WK model and on the cross correlation of output, inflation and real balances in the SP model and impose them in a VAR composed of output, inflation, real balances, the slope of the term structure and labor productivity using US, UK and EURO data from 1980:1 to 1998:4.*

We find that WK sign restrictions fail to recover monetary shocks in the UK, while SP sign restrictions do not produce monetary shocks in the Euro land. That is to say, out of 10000 draws for ω *and* $\mathcal{H}_{i,i'}(\omega)$ *we are able to find less than 0.1% of teh cases where the restrictions are satisfied. Since no combination of reduced form residuals produces cross correlations for output, inflation and the slope (or real balances) with the required sign, both models are at odds with the dynamic comovements in response to monetary shocks in at least one data set. One may stop here and try to respecify the models, or proceed with the data sets where restrictions hold and evaluation can continue examining e.g. the dynamic responses of the two other VAR variables to identified monetary shocks.*

There are at least two reasons for why a comparison based on real balances (or the slope) and labor productivity may be informative of the quality of the model's approximation to the data. First, we would like to know if identified monetary shocks produce liquidity effects, a feature present in both models and a simple "test" often used to decide whether a particular identification schemes is meaningful or not (see e.g. Leeper and Gordon (1994)). Second, it is common to use the dynamics of labor productivity to discriminate between flexible price real business cycle and sticky price demand driven explanations of economic fluctuations (see Gali (1999)). Since the dynamics of labor productivity in response to contractionary monetary shocks are similar in the two models (since employment declines more than output, labor productivity increases), it is interesting to check if the identified data qualitative conforms to these predictions.

Figure 4.6 plots the responses of these two variables for each data set (straight lines) together with the responses obtained in the two models (dotted lines), scaled so that the variance of the monetary policy innovation is the same. Two conclusions can be drawn. First, the WK identification scheme cannot account for the sign and the shape of the responses of labor productivity in US and Euro area and generates monetary disturbances in the Euro area which lack liquidity effects. Second, with the SP identification scheme monetary shocks generate instantaneous responses of the slope of the term structure which have the wrong sign in the US and lack persistence with UK data.

Given that the two theories produce dynamics which are qualitatively at odds with the data, it is not surprising to find that quantitative predictions are also unsatisfactory. For example, the percentage of output variance accounted for by monetary shocks in US at the 24 step horizon is between 11 and 43% with the WK scheme and 3 and 34% with the SP scheme. In comparison, and regardless of the parametrization used, monetary disturbances account for 1% of output variance in both models. Hence, both models lack internal propagation.

Given the mismatch of the models and the data one should probably go back to the drawing board before answering any policy question. Canova (2001) shows that adding capacity utilization and/or labor hoarding to the models is not enough to enhance at least the qualitative match. Whether other frictions will change this outcome is an open question.

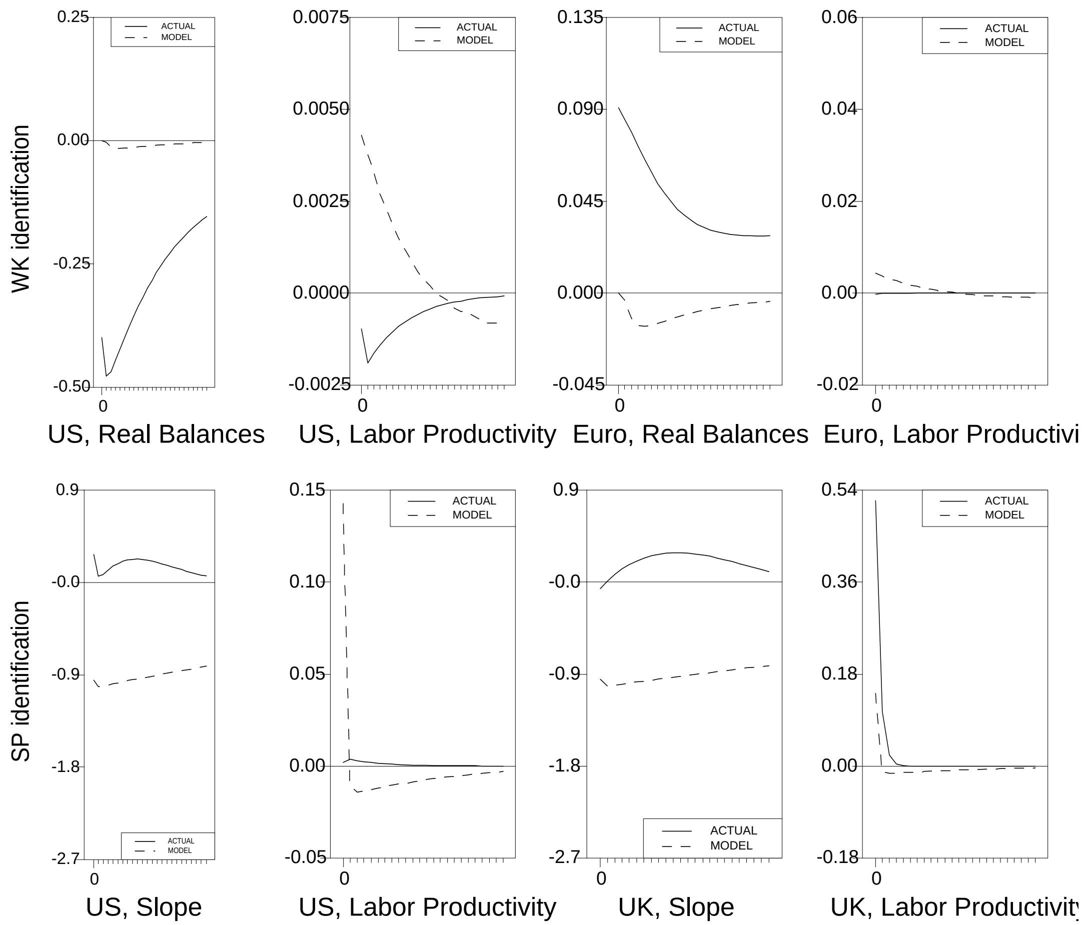

Figure 4.6: Responses to Monetary Shocks

Until a better match is found, it is probably unhealthy to try to answer any policy question with any of the two models.

Exercise 4.52 *(Dedola and Neri) Take a standard RBC model with habit persistence in consumption and highly persistent but stationary technology shocks. Examine whether robust sign restrictions for the correlation of output, hours and labor productivity exist when the extent of habit* (γ)*, the power of utility parameter* (φ)*, the share of hours in production* (η)*, the depreciation rate* (δ)*, and the persistence of technology shocks* (ρ_ζ) *are varied within reasonable ranges. Using a VAR with labor productivity, real wages, hours, investment, consumption and output examine whether the model fits the data, when robust sign restrictions are used to identify technology shocks in the data.*

Exercise 4.53 *(Pappa) In a sticky price model with monopolistic competitive firms anything that moves aggregate demand (e.g. goverment shocks) induces a shift in the labor demand curve and therefore induces positive comovements of hours and real wages. In a flexible price RBC model, on the other hand, goverment expenditure shocks shift both the aggregate supply and the aggregate demand curve. For many parametrizations movements in the former are larger than movements in the latter and therefore negative comovements of hours and real wages are generated. Using a VAR with labor productivity, hours, real wages, investment, consumption and output, verify whether a RBC style model fits the data better than a sticky price, monopolistic competitive model.*