6 Multivariate models

Learning Outcomes

In this chapter, you will learn how to

- Compare and contrast single equation and systems-based approaches to building models
- Discuss the cause, consequence and solution to simultaneous equations bias
- Derive the reduced form equations from a structural model
- Describe several methods for estimating simultaneous equations models
- Explain the relative advantages and disadvantages of VAR modelling
- Determine whether an equation from a system is identified
- Estimate optimal lag lengths, impulse responses and variance decompositions
- Conduct Granger causality tests
- Construct simultaneous equations models and VARs in EViews

6.1 Motivations

All of the structural models that have been considered thus far have been single equations models of the form

$$y = X\beta + u \tag{6.1}$$

One of the assumptions of the classical linear regression model (CLRM) is that the explanatory variables are *non-stochastic*, or fixed in repeated samples. There are various ways of stating this condition, some of which are slightly more or less strict, but all of which have the same broad implication. It could also be stated that all of the variables contained in the X matrix are assumed to be *exogenous* – that is, their values are determined outside that equation. This is a rather simplistic working definition of exogeneity, although several alternatives are possible; this issue will be revisited later in the chapter. Another way to state this is that the model is 'conditioned on' the variables in X.

As stated in chapter 2, the X matrix is assumed not to have a probability distribution. Note also that causality in this model runs from X to y, and not vice versa, i.e. that changes in the values of the explanatory variables cause changes in the values of y, but that changes in the value of y will not impact upon the explanatory variables. On the other hand, y is an *endogenous* variable – that is, its value is determined by (6.1).

The purpose of the first part of this chapter is to investigate one of the important circumstances under which the assumption presented above will be violated. The impact on the OLS estimator of such a violation will then be considered.

To illustrate a situation in which such a phenomenon may arise, consider the following two equations that describe a possible model for the total aggregate (country-wide) supply of new houses (or any other physical asset).

$$Q_{dt} = \alpha + \beta P_t + \gamma S_t + u_t \tag{6.2}$$
$$Q_{st} = \lambda + \mu P_t + \kappa T_t + v_t \tag{6.3}$$
$$Q_{dt} = Q_{st} \tag{6.4}$$

where

Q_{dt} = quantity of new houses demanded at time t
Q_{st} = quantity of new houses supplied (built) at time t
P_t = (average) price of new houses prevailing at time t
S_t = price of a substitute (e.g. older houses)
T_t = some variable embodying the state of housebuilding technology, u_t and v_t are error terms.

Equation (6.2) is an equation for modelling the demand for new houses, and (6.3) models the supply of new houses. (6.4) is an equilibrium condition for there to be no excess demand (people willing and able to buy new houses but cannot) and no excess supply (constructed houses that remain empty owing to lack of demand).

Assuming that the market always clears, that is, that the market is always in equilibrium, and dropping the time subscripts for simplicity, (6.2)–(6.4) can be written

$$Q = \alpha + \beta P + \gamma S + u \tag{6.5}$$
$$Q = \lambda + \mu P + \kappa T + v \tag{6.6}$$

Equations (6.5) and (6.6) together comprise a simultaneous structural form of the model, or a set of structural equations. These are the equations incorporating the variables that economic or financial theory suggests should be related to one another in a relationship of this form. The point is that price and quantity are determined simultaneously (price affects quantity and quantity affects price). Thus, in order to sell more houses, everything else equal, the builder will have to lower the price. Equally, in order to obtain a higher price for each house, the builder should construct and expect to sell fewer houses. P and Q are endogenous variables, while S and T are exogenous.

A set of reduced form equations corresponding to (6.5) and (6.6) can be obtained by solving (6.5) and (6.6) for P and for Q (separately). There will be a reduced form equation for each endogenous variable in the system.

Solving for Q

$$\alpha + \beta P + \gamma S + u = \lambda + \mu P + \kappa T + v \tag{6.7}$$

Solving for P

$$\frac{Q}{\beta} - \frac{\alpha}{\beta} - \frac{\gamma S}{\beta} - \frac{u}{\beta} = \frac{Q}{\mu} - \frac{\lambda}{\mu} - \frac{\kappa T}{\mu} - \frac{v}{\mu} \tag{6.8}$$

Rearranging (6.7)

$$\beta P - \mu P = \lambda - \alpha + \kappa T - \gamma S + v - u \tag{6.9}$$
$$(\beta - \mu)P = (\lambda - \alpha) + \kappa T - \gamma S + (v - u) \tag{6.10}$$
$$P = \frac{\lambda - \alpha}{\beta - \mu} + \frac{\kappa}{\beta - \mu}T - \frac{\gamma}{\beta - \mu}S + \frac{v - u}{\beta - \mu} \tag{6.11}$$

Multiplying (6.8) through by $\beta\mu$ and rearranging

$$\mu Q - \mu\alpha - \mu\gamma S - \mu u = \beta Q - \beta\lambda - \beta\kappa T - \beta v \tag{6.12}$$
$$\mu Q - \beta Q = \mu\alpha - \beta\lambda - \beta\kappa T + \mu\gamma S + \mu u - \beta v \tag{6.13}$$
$$(\mu - \beta)Q = (\mu\alpha - \beta\lambda) - \beta\kappa T + \mu\gamma S + (\mu u - \beta v) \tag{6.14}$$
$$Q = \frac{\mu\alpha - \beta\lambda}{\mu - \beta} - \frac{\beta\kappa}{\mu - \beta}T + \frac{\mu\gamma}{\mu - \beta}S + \frac{\mu u - \beta v}{\mu - \beta} \tag{6.15}$$

(6.11) and (6.15) are the reduced form equations for P and Q. They are the equations that result from solving the simultaneous structural equations given by (6.5) and (6.6). Notice that these reduced form equations have only exogenous variables on the RHS.

6.2 Simultaneous equations bias

It would not be possible to estimate (6.5) and (6.6) validly using OLS, as they are clearly related to one another since they both contain P and Q, and OLS would require them to be estimated separately. But what would have happened if a researcher had estimated them separately using OLS? Both equations depend on P. One of the CLRM assumptions was that X and u are independent (where X is a matrix containing all the variables on the RHS of the equation), and given also the assumption that $\mathrm{E}(u) = 0$, then $\mathrm{E}(X'u) = 0$, i.e. the errors are uncorrelated with the explanatory variables. But it is clear from (6.11) that P is related to the errors in (6.5) and (6.6) – i.e. it is *stochastic*. So this assumption has been violated.

What would be the consequences for the OLS estimator, $\hat{\beta}$ if the simultaneity were ignored? Recall that

$$\hat{\beta} = (X'X)^{-1}X'y \tag{6.16}$$

and that

$$y = X\beta + u \tag{6.17}$$

Replacing y in (6.16) with the RHS of (6.17)

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u) \tag{6.18}$$

so that

$$\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \tag{6.19}$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'u \tag{6.20}$$

Taking expectations,

$$E(\hat{\beta}) = E(\beta) + E((X'X)^{-1}X'u) \tag{6.21}$$

$$E(\hat{\beta}) = \beta + E((X'X)^{-1}X'u) \tag{6.22}$$

If the Xs are non-stochastic (i.e. if the assumption had not been violated), $\mathrm{E}[(X'X)^{-1}X'u] = (X'X)^{-1}X'\mathrm{E}[u] = 0$, which would be the case in a single equation system, so that $E(\hat{\beta}) = \beta$ in (6.22). The implication is that the OLS estimator, $\hat{\beta}$, would be unbiased.

But, if the equation is part of a system, then $\mathrm{E}[(X'X)^{-1}X'u] \neq 0$, in general, so that the last term in (6.22) will not drop out, and so it can be concluded that application of OLS to structural equations which are part of a simultaneous system will lead to biased coefficient estimates. This is known as *simultaneity bias* or *simultaneous equations bias*.

Is the OLS estimator still consistent, even though it is biased? No, in fact, the estimator is inconsistent as well, so that the coefficient estimates would still be biased even if an infinite amount of data were available, although proving this would require a level of algebra beyond the scope of this book.

6.3 So how can simultaneous equations models be validly estimated?

Taking (6.11) and (6.15), i.e. the reduced form equations, they can be rewritten as

$$P = \pi_{10} + \pi_{11} T + \pi_{12} S + \varepsilon_1 \tag{6.23}$$
$$Q = \pi_{20} + \pi_{21} T + \pi_{22} S + \varepsilon_2 \tag{6.24}$$

where the π coefficients in the reduced form are simply combinations of the original coefficients, so that

$$\pi_{10} = \frac{\lambda - \alpha}{\beta - \mu}, \quad \pi_{11} = \frac{\kappa}{\beta - \mu}, \quad \pi_{12} = \frac{-\gamma}{\beta - \mu}, \quad \varepsilon_1 = \frac{v - u}{\beta - \mu},$$
$$\pi_{20} = \frac{\mu\alpha - \beta\lambda}{\mu - \beta}, \quad \pi_{21} = \frac{-\beta\kappa}{\mu - \beta}, \quad \pi_{22} = \frac{\mu\gamma}{\mu - \beta}, \quad \varepsilon_2 = \frac{\mu u - \beta v}{\mu - \beta}$$

Equations (6.23) and (6.24) can be estimated using OLS since all the RHS variables are exogenous, so the usual requirements for consistency and unbiasedness of the OLS estimator will hold (provided that there are no other misspecifications). Estimates of the π_{ij} coefficients would thus be obtained. But, the values of the π coefficients are probably not of much interest; what was wanted were the original parameters in the structural equations – $\alpha, \beta, \gamma, \lambda, \mu, \kappa$. The latter are the parameters whose values determine how the variables are related to one another according to financial or economic theory.

6.4 Can the original coefficients be retrieved from the πs?

The short answer to this question is ‘sometimes’, depending upon whether the equations are identified. *Identification* is the issue of whether there is enough information in the reduced form equations to enable the structural form coefficients to be calculated. Consider the following demand

and supply equations

$$Q = \alpha + \beta P \quad \text{Supply equation} \tag{6.25}$$

$$Q = \lambda + \mu P \quad \text{Demand equation} \tag{6.26}$$

It is impossible to tell which equation is which, so that if one simply observed some quantities of a good sold and the price at which they were sold, it would not be possible to obtain the estimates of α, β, λ and μ. This arises since there is insufficient information from the equations to estimate 4 parameters. Only 2 parameters could be estimated here, although each would be some combination of demand and supply parameters, and so neither would be of any use. In this case, it would be stated that both equations are *unidentified* (or not identified or underidentified). Notice that this problem would not have arisen with (6.5) and (6.6) since they have different exogenous variables.

6.4.1 *What determines whether an equation is identified or not?*

Any one of three possible situations could arise, as shown in box 6.1.

How can it be determined whether an equation is identified or not? Broadly, the answer to this question depends upon how many and which variables are present in each structural equation. There are two conditions that could be examined to determine whether a given equation from a system is identified – the *order condition* and the *rank condition*:

- The *order condition* – is a necessary but not sufficient condition for an equation to be identified. That is, even if the order condition is satisfied, the equation might not be identified.
- The *rank condition* – is a necessary and sufficient condition for identification. The structural equations are specified in a matrix form and the rank of a coefficient matrix of all of the variables excluded from a

Box 6.1 Determining whether an equation is identified

(1) An equation is *unidentified*, such as (6.25) or (6.26). In the case of an unidentified equation, structural coefficients cannot be obtained from the reduced form estimates by any means.
(2) An equation is *exactly identified* (*just identified*), such as (6.5) or (6.6). In the case of a just identified equation, unique structural form coefficient estimates can be obtained by substitution from the reduced form equations.
(3) If an equation is *overidentified*, more than one set of structural coefficients can be obtained from the reduced form. An example of this will be presented later in this chapter.

particular equation is examined. An examination of the rank condition requires some technical algebra beyond the scope of this text.

Even though the order condition is not sufficient to ensure identification of an equation from a system, the rank condition will not be considered further here. For relatively simple systems of equations, the two rules would lead to the same conclusions. Also, in fact, most systems of equations in economics and finance are overidentified, so that underidentification is not a big issue in practice.

6.4.2 *Statement of the order condition*

There are a number of different ways of stating the order condition; that employed here is an intuitive one (taken from Ramanathan, 1995, p. 666, and slightly modified):

Let G denote the number of structural equations. An equation is just identified if the number of variables excluded from an equation is $G-1$, where 'excluded' means the number of all endogenous and exogenous variables that are not present in this particular equation. If more than $G-1$ are absent, it is over-identified. If less than $G-1$ are absent, it is not identified.

One obvious implication of this rule is that equations in a system can have differing degrees of identification, as illustrated by the following example.

Example 6.1

In the following system of equations, the Ys are endogenous, while the Xs are exogenous (with time subscripts suppressed). Determine whether each equation is overidentified, underidentified, or just identified.

$$Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_3 Y_3 + \alpha_4 X_1 + \alpha_5 X_2 + u_1 \tag{6.27}$$
$$Y_2 = \beta_0 + \beta_1 Y_3 + \beta_2 X_1 + u_2 \tag{6.28}$$
$$Y_3 = \gamma_0 + \gamma_1 Y_2 + u_3 \tag{6.29}$$

In this case, there are $G = 3$ equations and 3 endogenous variables. Thus, if the number of excluded variables is exactly 2, the equation is just identified. If the number of excluded variables is more than 2, the equation is overidentified. If the number of excluded variables is less than 2, the equation is not identified.

The variables that appear in one or more of the three equations are Y_1, Y_2, Y_3, X_1, X_2. Applying the order condition to (6.27)–(6.29):

- Equation (6.27): contains all variables, with none excluded, so that it is not identified
- Equation (6.28): has variables Y_1 and X_2 excluded, and so is just identified
- Equation (6.29): has variables Y_1, X_1, X_2 excluded, and so is overidentified

6.5 Simultaneous equations in finance

There are of course numerous situations in finance where a simultaneous equations framework is more relevant than a single equation model. Two illustrations from the market microstructure literature are presented later in this chapter, while another, drawn from the banking literature, will be discussed now.

There has recently been much debate internationally, but especially in the UK, concerning the effectiveness of competitive forces in the banking industry. Governments and regulators express concern at the increasing concentration in the industry, as evidenced by successive waves of merger activity, and at the enormous profits that many banks made in the late 1990s and early twenty-first century. They argue that such profits result from a lack of effective competition. However, many (most notably, of course, the banks themselves!) suggest that such profits are not the result of excessive concentration or anti-competitive practices, but rather partly arise owing to recent world prosperity at that phase of the business cycle (the 'profits won't last' argument) and partly owing to massive cost-cutting by the banks, given recent technological improvements. These debates have fuelled a resurgent interest in models of banking profitability and banking competition. One such model is employed by Shaffer and DiSalvo (1994) in the context of two banks operating in south central Pennsylvania. The model is given by

$$\ln q_{it} = a_0 + a_1 \ln P_{it} + a_2 \ln P_{jt} + a_3 \ln Y_t + a_4 \ln Z_t + a_5 t + u_{i1t} \tag{6.30}$$

$$\ln TR_{it} = b_0 + b_1 \ln q_{it} + \sum_{k=1}^{3} b_{k+1} \ln w_{ikt} + u_{i2t} \tag{6.31}$$

where $i = 1, 2$ are the two banks, q is bank output, P_t is the price of the output at time t, Y_t is a measure of aggregate income at time t, Z_t is the price of a substitute for bank activity at time t, the variable t represents a time trend, TR_{it} is the total revenue of bank i at time t, w_{ikt}

are the prices of input k ($k = 1, 2, 3$ for labour, bank deposits, and physical capital) for bank i at time t and the u are unobservable error terms. The coefficient estimates are not presented here, but suffice to say that a simultaneous framework, with the resulting model estimated separately using annual time series data for each bank, is necessary. Output is a function of price on the RHS of (6.30), while in (6.31), total revenue, which is a function of output on the RHS, is obviously related to price. Therefore, OLS is again an inappropriate estimation technique. Both of the equations in this system are overidentified, since there are only two equations, and the income, the substitute for banking activity, and the trend terms are missing from (6.31), whereas the three input prices are missing from (6.30).

6.6 A definition of exogeneity

Leamer (1985) defines a variable x as exogenous if the conditional distribution of y given x does not change with modifications of the process generating x. Although several slightly different definitions exist, it is possible to classify two forms of exogeneity – predeterminedness and strict exogeneity:

- A *predetermined* variable is one that is independent of the contemporaneous and future errors in that equation
- A *strictly exogenous* variable is one that is independent of all contemporaneous, future and past errors in that equation.

6.6.1 *Tests for exogeneity*

How can a researcher tell whether variables really need to be treated as endogenous or not? In other words, financial theory might suggest that there should be a two-way relationship between two or more variables, but how can it be tested whether a simultaneous equations model is necessary in practice?

Example 6.2

Consider again (6.27)–(6.29). Equation (6.27) contains Y_2 and Y_3 – but are separate equations required for them, or could the variables Y_2 and Y_3 be treated as exogenous variables (in which case, they would be called X_3 and X_4!)? This can be formally investigated using a Hausman test, which is calculated as shown in box 6.2.

Box 6.2 Conducting a Hausman test for exogeneity

(1) *Obtain the reduced form equations corresponding to (6.27)–(6.29).*
The reduced form equations are obtained as follows.

Substituting in (6.28) for Y_3 from (6.29):

$$Y_2 = \beta_0 + \beta_1(\gamma_0 + \gamma_1 Y_2 + u_3) + \beta_2 X_1 + u_2 \tag{6.32}$$

$$Y_2 = \beta_0 + \beta_1\gamma_0 + \beta_1\gamma_1 Y_2 + \beta_1 u_3 + \beta_2 X_1 + u_2 \tag{6.33}$$

$$Y_2(1 - \beta_1\gamma_1) = (\beta_0 + \beta_1\gamma_0) + \beta_2 X_1 + (u_2 + \beta_1 u_3) \tag{6.34}$$

$$Y_2 = \frac{(\beta_0 + \beta_1\gamma_0)}{(1 - \beta_1\gamma_1)} + \frac{\beta_2 X_1}{(1 - \beta_1\gamma_1)} + \frac{(u_2 + \beta_1 u_3)}{(1 - \beta_1\gamma_1)} \tag{6.35}$$

(6.35) is the reduced form equation for Y_2, since there are no endogenous variables on the RHS. Substituting in (6.27) for Y_3 from (6.29)

$$Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_3(\gamma_0 + \gamma_1 Y_2 + u_3) + \alpha_4 X_1 + \alpha_5 X_2 + u_1 \tag{6.36}$$

$$Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_3\gamma_0 + \alpha_3\gamma_1 Y_2 + \alpha_3 u_3 + \alpha_4 X_1 + \alpha_5 X_2 + u_1 \tag{6.37}$$

$$Y_1 = (\alpha_0 + \alpha_3\gamma_0) + (\alpha_1 + \alpha_3\gamma_1)Y_2 + \alpha_4 X_1 + \alpha_5 X_2 + (u_1 + \alpha_3 u_3) \tag{6.38}$$

Substituting in (6.38) for Y_2 from (6.35):

$$\begin{aligned} Y_1 = {} & (\alpha_0 + \alpha_3\gamma_0) + (\alpha_1 + \alpha_3\gamma_1)\left(\frac{(\beta_0 + \beta_1\gamma_0)}{(1 - \beta_1\gamma_1)} + \frac{\beta_2 X_1}{(1 - \beta_1\gamma_1)} + \frac{(u_2 + \beta_1 u_3)}{(1 - \beta_1\gamma_1)}\right) \\ & + \alpha_4 X_1 + \alpha_5 X_2 + (u_1 + \alpha_3 u_3) \end{aligned} \tag{6.39}$$

$$\begin{aligned} Y_1 = {} & \left(\alpha_0 + \alpha_3\gamma_0 + (\alpha_1 + \alpha_3\gamma_1)\frac{(\beta_0 + \beta_1\gamma_0)}{(1 - \beta_1\gamma_1)}\right) + \frac{(\alpha_1 + \alpha_3\gamma_1)\beta_2 X_1}{(1 - \beta_1\gamma_1)} \\ & + \frac{(\alpha_1 + \alpha_3\gamma_1)(u_2 + \beta_1 u_3)}{(1 - \beta_1\gamma_1)} + \alpha_4 X_1 + \alpha_5 X_2 + (u_1 + \alpha_3 u_3) \end{aligned} \tag{6.40}$$

$$\begin{aligned} Y_1 = {} & \left(\alpha_0 + \alpha_3\gamma_0 + (\alpha_1 + \alpha_3\gamma_1)\frac{(\beta_0 + \beta_1\gamma_0)}{(1 - \beta_1\gamma_1)}\right) + \left(\frac{(\alpha_1 + \alpha_3\gamma_1)\beta_2}{(1 - \beta_1\gamma_1)} + \alpha_4\right) X_1 \\ & + \alpha_5 X_2 + \left(\frac{(\alpha_1 + \alpha_3\gamma_1)(u_2 + \beta_1 u_3)}{(1 - \beta_1\gamma_1)} + (u_1 + \alpha_3 u_3)\right) \end{aligned} \tag{6.41}$$

(6.41) is the reduced form equation for Y_1. Finally, to obtain the reduced form equation for Y_3, substitute in (6.29) for Y_2 from (6.35)

$$Y_3 = \left(\gamma_0 + \frac{\gamma_1(\beta_0 + \beta_1\gamma_0)}{(1 - \beta_1\gamma_1)}\right) + \frac{\gamma_1\beta_2 X_1}{(1 - \beta_1\gamma_1)} + \left(\frac{\gamma_1(u_2 + \beta_1 u_3)}{(1 - \beta_1\gamma_1)} + u_3\right) \tag{6.42}$$

So, the reduced form equations corresponding to (6.27)–(6.29) are, respectively, given by (6.41), (6.35) and (6.42). These three equations can also be expressed using π_{ij} for the coefficients, as discussed above

$$Y_1 = \pi_{10} + \pi_{11} X_1 + \pi_{12} X_2 + v_1 \tag{6.43}$$

$$Y_2 = \pi_{20} + \pi_{21} X_1 + v_2 \tag{6.44}$$

$$Y_3 = \pi_{30} + \pi_{31} X_1 + v_3 \tag{6.45}$$

Estimate the reduced form equations (6.43)–(6.45) using OLS, and obtain the fitted values, $\hat{Y}_1^1,\ \hat{Y}_2^1,\ \hat{Y}_3^1$, where the superfluous superscript ^{1} denotes the fitted values from the reduced form estimation.

(2) Run the regression corresponding to (6.27) – i.e. the structural form equation, at this stage ignoring any possible simultaneity.

(3) Run the regression (6.27) again, but now also including the fitted values from the reduced form equations, $\hat{Y}_2^1, \hat{Y}_3^1$, as additional regressors

$$Y_1 = \alpha_0 + \alpha_1 Y_2 + \alpha_3 Y_3 + \alpha_4 X_1 + \alpha_5 X_2 + \lambda_2 \hat{Y}_2^1 + \lambda_3 \hat{Y}_3^1 + \varepsilon_1 \tag{6.46}$$

(4) Use an *F*-test to test the joint restriction that $\lambda_2 = 0$, and $\lambda_3 = 0$. If the null hypothesis is rejected, Y_2 and Y_3 should be treated as endogenous. If λ_2 and λ_3 are significantly different from zero, there is extra important information for modelling Y_1 from the reduced form equations. On the other hand, if the null is not rejected, Y_2 and Y_3 can be treated as exogenous for Y_1, and there is no useful additional information available for Y_1 from modelling Y_2 and Y_3 as endogenous variables.

Steps 2–4 would then be repeated for (6.28) and (6.29).

6.7 Triangular systems

Consider the following system of equations, with time subscripts omitted for simplicity

$$Y_1 = \beta_{10} + \gamma_{11} X_1 + \gamma_{12} X_2 + u_1 \tag{6.47}$$
$$Y_2 = \beta_{20} + \beta_{21} Y_1 + \gamma_{21} X_1 + \gamma_{22} X_2 + u_2 \tag{6.48}$$
$$Y_3 = \beta_{30} + \beta_{31} Y_1 + \beta_{32} Y_2 + \gamma_{31} X_1 + \gamma_{32} X_2 + u_3 \tag{6.49}$$

Assume that the error terms from each of the three equations are not correlated with each other. Can the equations be estimated individually using OLS? At first blush, an appropriate answer to this question might appear to be, 'No, because this is a simultaneous equations system.' But consider the following:

- Equation (6.47): contains no endogenous variables, so X_1 and X_2 are not correlated with u_1. So OLS can be used on (6.47).
- Equation (6.48): contains endogenous Y_1 together with exogenous X_1 and X_2. OLS can be used on (6.48) if all the RHS variables in (6.48) are uncorrelated with that equation's error term. In fact, Y_1 is not correlated with u_2 because there is no Y_2 term in (6.47). So OLS can be used on (6.48).
- Equation (6.49): contains both Y_1 and Y_2; these are required to be uncorrelated with u_3. By similar arguments to the above, (6.47) and (6.48) do not contain Y_3. So OLS can be used on (6.49).

This is known as a *recursive or triangular system*, which is really a special case – a set of equations that looks like a simultaneous equations system, but isn't. In fact, there is not a simultaneity problem here, since the dependence is not bi-directional, for each equation it all goes one way.

6.8 Estimation procedures for simultaneous equations systems

Each equation that is part of a recursive system can be estimated separately using OLS. But in practice, not many systems of equations will be recursive, so a direct way to address the estimation of equations that are from a true simultaneous system must be sought. In fact, there are potentially many methods that can be used, three of which – indirect least squares, two-stage least squares and instrumental variables – will be detailed here. Each of these will be discussed below.

6.8.1 *Indirect least squares (ILS)*

Although it is not possible to use OLS directly on the structural equations, it is possible to validly apply OLS to the reduced form equations. If the system is just identified, ILS involves estimating the reduced form equations using OLS, and then using them to substitute back to obtain the structural parameters. ILS is intuitive to understand in principle; however, it is not widely applied because:

(1) *Solving back to get the structural parameters can be tedious*. For a large system, the equations may be set up in a matrix form, and to solve them may therefore require the inversion of a large matrix.
(2) *Most simultaneous equations systems are overidentified*, and ILS can be used to obtain coefficients only for just identified equations. For overidentified systems, ILS would not yield unique structural form estimates.

ILS estimators are consistent and asymptotically efficient, but in general they are biased, so that in finite samples ILS will deliver biased structural form estimates. In a nutshell, the bias arises from the fact that the structural form coefficients under ILS estimation are transformations of the reduced form coefficients. When expectations are taken to test for unbiasedness, it is in general not the case that the expected value of a (non-linear) combination of reduced form coefficients will be equal to the combination of their expected values (see Gujarati, 1995, pp. 704–5 for a proof).

6.8.2 *Estimation of just identified and overidentified systems using 2SLS*

This technique is applicable for the estimation of overidentified systems, where ILS cannot be used. In fact, it can also be employed for estimating the coefficients of just identified systems, in which case the method would yield asymptotically equivalent estimates to those obtained from ILS.

Two-stage least squares (2SLS or TSLS) is done in two stages:

- *Stage 1* Obtain and estimate the reduced form equations using OLS. Save the fitted values for the dependent variables.
- *Stage 2* Estimate the structural equations using OLS, but replace any RHS endogenous variables with their stage 1 fitted values.

Example 6.3

Suppose that (6.27)–(6.29) are required. 2SLS would involve the following two steps:

- *Stage 1* Estimate the reduced form equations (6.43)–(6.45) individually by OLS and obtain the fitted values, and denote them $\hat{Y}_1^1$, $\hat{Y}_2^1$, $\hat{Y}_3^1$, where the superfluous superscript ^{1} indicates that these are the fitted values from the first stage.
- *Stage 2* Replace the RHS endogenous variables with their stage 1 estimated values

$$Y_1 = \alpha_0 + \alpha_1 \hat{Y}_2^1 + \alpha_3 \hat{Y}_3^1 + \alpha_4 X_1 + \alpha_5 X_2 + u_1 \tag{6.50}$$

$$Y_2 = \beta_0 + \beta_1 \hat{Y}_3^1 + \beta_2 X_1 + u_2 \tag{6.51}$$

$$Y_3 = \gamma_0 + \gamma_1 \hat{Y}_2^1 + u_3 \tag{6.52}$$

where $\hat{Y}_2^1$ and $\hat{Y}_3^1$ are the fitted values from the reduced form estimation. Now $\hat{Y}_2^1$ and $\hat{Y}_3^1$ will not be correlated with u_1, $\hat{Y}_3^1$ will not be correlated with u_2, and $\hat{Y}_2^1$ will not be correlated with u_3. The simultaneity problem has therefore been removed. It is worth noting that the 2SLS estimator is consistent, but not unbiased.

In a simultaneous equations framework, it is still of concern whether the usual assumptions of the CLRM are valid or not, although some of the test statistics require modifications to be applicable in the systems context. Most econometrics packages will automatically make any required changes. To illustrate one potential consequence of the violation of the CLRM assumptions, if the disturbances in the structural equations are autocorrelated, the 2SLS estimator is not even consistent.

The standard error estimates also need to be modified compared with their OLS counterparts (again, econometrics software will usually do this automatically), but once this has been done, the usual t-tests can be used to test hypotheses about the structural form coefficients. This modification arises as a result of the use of the reduced form fitted values on the RHS rather than actual variables, which implies that a modification to the error variance is required.

6.8.3 Instrumental variables

Broadly, the method of instrumental variables (IV) is another technique for parameter estimation that can be validly used in the context of a simultaneous equations system. Recall that the reason that OLS cannot be used directly on the structural equations is that the endogenous variables are correlated with the errors.

One solution to this would be not to use Y_2 or Y_3, but rather to use some other variables instead. These other variables should be (highly) correlated with Y_2 and Y_3, but not correlated with the errors – such variables would be known as *instruments*. Suppose that suitable instruments for Y_2 and Y_3, were found and denoted z_2 and z_3, respectively. The instruments are not used in the structural equations directly, but rather, regressions of the following form are run

$$Y_2 = \lambda_1 + \lambda_2 z_2 + \varepsilon_1 \tag{6.53}$$

$$Y_3 = \lambda_3 + \lambda_4 z_3 + \varepsilon_2 \tag{6.54}$$

Obtain the fitted values from (6.53) and (6.54), $\hat{Y}_2^1$ and $\hat{Y}_3^1$, and replace Y_2 and Y_3 with these in the structural equation. It is typical to use more than one instrument per endogenous variable. If the instruments are the variables in the reduced form equations, then IV is equivalent to 2SLS, so that the latter can be viewed as a special case of the former.

6.8.4 What happens if IV or 2SLS are used unnecessarily?

In other words, suppose that one attempted to estimate a simultaneous system when the variables specified as endogenous were in fact independent of one another. The consequences are similar to those of including irrelevant variables in a single equation OLS model. That is, the coefficient estimates will still be consistent, but will be inefficient compared to those that just used OLS directly.

6.8.5 *Other estimation techniques*

There are, of course, many other estimation techniques available for systems of equations, including three-stage least squares (3SLS), full information maximum likelihood (FIML) and limited information maximum likelihood (LIML). Three-stage least squares provides a third step in the estimation process that allows for non-zero covariances between the error terms in the structural equations. It is asymptotically more efficient than 2SLS since the latter ignores any information that may be available concerning the error covariances (and also any additional information that may be contained in the endogenous variables of other equations). Full information maximum likelihood involves estimating all of the equations in the system simultaneously using maximum likelihood (see chapter 8 for a discussion of the principles of maximum likelihood estimation). Thus under FIML, all of the parameters in all equations are treated jointly, and an appropriate likelihood function is formed and maximised. Finally, limited information maximum likelihood involves estimating each equation separately by maximum likelihood. LIML and 2SLS are asymptotically equivalent. For further technical details on each of these procedures, see Greene (2002, chapter 15).

The following section presents an application of the simultaneous equations approach in finance to the joint modelling of bid–ask spreads and trading activity in the S&P100 index options market. Two related applications of this technique that are also worth examining are by Wang *et al*. (1997) and by Wang and Yau (2000). The former employs a bivariate system to model trading volume and bid–ask spreads and they show using a Hausman test that the two are indeed simultaneously related and so must both be treated as endogenous variables and are modelled using 2SLS. The latter paper employs a trivariate system to model trading volume, spreads and intra-day volatility.

6.9 An application of a simultaneous equations approach to modelling bid–ask spreads and trading activity

6.9.1 *Introduction*

One of the most rapidly growing areas of empirical research in finance is the study of market microstructure. This research is involved with issues such as price formation in financial markets, how the structure of the market may affect the way it operates, determinants of the bid–ask spread, and so on. One application of simultaneous equations methods in the market microstructure literature is a study by George and Longstaff (1993). Among other issues, this paper considers the questions:

- Is trading activity related to the size of the bid–ask spread?
- How do spreads vary across options, and how is this related to the volume of contracts traded? 'Across options' in this case means for different maturities and strike prices for an option on a given underlying asset.

This chapter will now examine the George and Longstaff models, results and conclusions.

6.9.2 *The data*

The data employed by George and Longstaff comprise options prices on the S&P100 index, observed on all trading days during 1989. The S&P100 index has been traded on the Chicago Board Options Exchange (CBOE) since 1983 on a continuous open-outcry auction basis. The option price as used in the paper is defined as the average of the bid and the ask. The average bid and ask prices are calculated for each option during the time 2.00p.m.–2.15p.m. (US Central Standard Time) to avoid time-of-day effects, such as differences in behaviour at the open and the close of the market. The following are then dropped from the sample for that day to avoid any effects resulting from stale prices:

- Any options that do not have bid and ask quotes reported during the 1/4 hour
- Any options with fewer than ten trades during the day.

This procedure results in a total of 2,456 observations. A 'pooled' regression is conducted since the data have both time series and cross-sectional dimensions. That is, the data are measured every trading day and across options with different strikes and maturities, and the data is stacked in a single column for analysis.

6.9.3 *How might the option price/trading volume and the bid–ask spread be related?*

George and Longstaff argue that the bid–ask spread will be determined by the interaction of market forces. Since there are many market makers trading the S&P100 contract on the CBOE, the bid–ask spread will be set to just cover marginal costs. There are three components of the costs associated with being a market maker. These are administrative costs, inventory holding costs, and 'risk costs'. George and Longstaff consider three possibilities for how the bid–ask spread might be determined:

- *Market makers equalise spreads across options* This is likely to be the case if order-processing (administrative) costs make up the majority of costs associated with being a market maker. This could be the case since the CBOE charges market makers the same fee for each option traded. In fact, for every contract (100 options) traded, a CBOE fee of 9 cents and an Options Clearing Corporation (OCC) fee of 10 cents is levied on the firm that clears the trade.
- *The spread might be a constant proportion of the option value* This would be the case if the majority of the market maker's cost is in inventory holding costs, since the more expensive options will cost more to hold and hence the spread would be set wider.
- *Market makers might equalise marginal costs across options irrespective of trading volume* This would occur if the riskiness of an unwanted position were the most important cost facing market makers. Market makers typically do not hold a particular view on the direction of the market – they simply try to make money by buying and selling. Hence, they would like to be able to offload any unwanted (long or short) positions quickly. But trading is not continuous, and in fact the average time between trades in 1989 was approximately five minutes. The longer market makers hold an option, the higher the risk they face since the higher the probability that there will be a large adverse price movement. Thus options with low trading volumes would command higher spreads since it is more likely that the market maker would be holding these options for longer.

In a non-quantitative exploratory analysis, George and Longstaff find that, comparing across contracts with different maturities, the bid–ask spread does indeed increase with maturity (as the option with longer maturity is worth more) and with 'moneyness' (that is, an option that is deeper in the money has a higher spread than one which is less in the money). This is seen to be true for both call and put options.

6.9.4 *The influence of tick-size rules on spreads*

The CBOE limits the *tick size* (the minimum granularity of price quotes), which will of course place a lower limit on the size of the spread. The tick sizes are:

- \$1/8 for options worth \$3 or more
- \$1/16 for options worth less than \$3.

6.9.5 The models and results

The intuition that the bid–ask spread and trading volume may be simultaneously related arises since a wider spread implies that trading is relatively more expensive so that marginal investors would withdraw from the market. On the other hand, market makers face additional risk if the level of trading activity falls, and hence they may be expected to respond by increasing their fee (the spread). The models developed seek to simultaneously determine the size of the bid–ask spread and the time between trades.

For the calls, the model is:

$$CBA_i = \alpha_0 + \alpha_1 CDUM_i + \alpha_2 C_i + \alpha_3 CL_i + \alpha_4 T_i + \alpha_5 CR_i + e_i \tag{6.55}$$
$$CL_i = \gamma_0 + \gamma_1 CBA_i + \gamma_2 T_i + \gamma_3 T_i^2 + \gamma_4 M_i^2 + v_i \tag{6.56}$$

And symmetrically for the puts:

$$PBA_i = \beta_0 + \beta_1 PDUM_i + \beta_2 P_i + \beta_3 PL_i + \beta_4 T_i + \beta_5 PR_i + u_i \tag{6.57}$$
$$PL_i = \delta_0 + \delta_1 PBA_i + \delta_2 T_i + \delta_3 T_i^2 + \delta_4 M_i^2 + w_i \tag{6.58}$$

where CBA_i and PBA_i are the call bid–ask spread and the put bid–ask spread for option i, respectively

C_i and P_i are the call price and put price for option i, respectively
CL_i and PL_i are the times between trades for the call and put option i, respectively
CR_i and PR_i are the squared deltas of the options
$CDUM_i$ and $PDUM_i$ are dummy variables to allow for the minimum tick size

$$\begin{aligned} &= 0 \quad \text{if } C_i \text{ or } P_i < \$3 \\ &= 1 \quad \text{if } C_i \text{ or } P_i \geq \$3 \end{aligned}$$

T is the time to maturity
T^2 allows for a non-linear relationship between time to maturity and the spread M^2 is the square of moneyness, which is employed in quadratic form since at-the-money options have a higher trading volume, while out-of-the-money and in-the-money options both have lower trading activity
CR_i and PR_i are measures of risk for the call and put, respectively, given by the square of their deltas.

Equations (6.55) and (6.56), and then separately (6.57) and (6.58), are estimated using 2SLS. The results are given here in tables 6.1 and 6.2.

Table 6.1 Call bid–ask spread and trading volume regression

$$CBA_i = \alpha_0 + \alpha_1 CDUM_i + \alpha_2 C_i + \alpha_3 CL_i + \alpha_4 T_i + \alpha_5 CR_i + e_i \quad (6.55)$$
$$CL_i = \gamma_0 + \gamma_1 CBA_i + \gamma_2 T_i + \gamma_3 T_i^2 + \gamma_4 M_i^2 + v_i \quad (6.56)$$

α_0	α_1	α_2	α_3	α_4	α_5	Adj. R^2
0.08362 (16.80)	0.06114 (8.63)	0.01679 (15.49)	0.00902 (14.01)	−0.00228 (−12.31)	−0.15378 (−12.52)	0.688
γ_0	γ_1	γ_2	γ_3	γ_4	Adj. R^2	
−3.8542 (−10.50)	46.592 (30.49)	−0.12412 (−6.01)	0.00406 (14.43)	0.00866 (4.76)	0.618	

Note: t-ratios in parentheses.
Source: George and Longstaff (1993). Reprinted with the permission of School of Business Administration, University of Washington.

Table 6.2 Put bid–ask spread and trading volume regression

$$PBA_i = \beta_0 + \beta_1 PDUM_i + \beta_2 P_i + \beta_3 PL_i + \beta_4 T_i + \beta_5 PR_i + u_i \quad (6.57)$$
$$PL_i = \delta_0 + \delta_1 PBA_i + \delta_2 T_i + \delta_3 T_i^2 + \delta_4 M_i^2 + w_i \quad (6.58)$$

β_0	β_1	β_2	β_3	β_4	β_5	Adj.R^2
0.05707 (15.19)	0.03258 (5.35)	0.01726 (15.90)	0.00839 (12.56)	−0.00120 (−7.13)	−0.08662 (−7.15)	0.675
δ_0	δ_1	δ_2	δ_3	δ_4	Adj. R^2	
−2.8932 (−8.42)	46.460 (34.06)	−0.15151 (−7.74)	0.00339 (12.90)	0.01347 (10.86)	0.517	

Note: t-ratios in parentheses.
Source: George and Longstaff (1993). Reprinted with the permission of School of Business Administration, University of Washington.

The adjusted $R^2 \approx 0.6$ for all four equations, indicating that the variables selected do a good job of explaining the spread and the time between trades. George and Longstaff argue that strategic market maker behaviour, which cannot be easily modelled, is important in influencing the spread and that this precludes a higher adjusted R^2.

A next step in examining the empirical plausibility of the estimates is to consider the sizes, signs and significances of the coefficients. In the call and put spread regressions, respectively, α_1 and β_1 measure the tick size constraint on the spread – both are statistically significant and positive. α_2 and β_2 measure the effect of the option price on the spread. As expected, both of these coefficients are again significant and positive since these are inventory or holding costs. The coefficient value of approximately 0.017 implies that a 1 dollar increase in the price of the option will on average lead to a 1.7 cent increase in the spread. α_3 and β_3 measure the effect of trading activity on the spread. Recalling that an inverse trading activity variable is used in the regressions, again, the coefficients have their correct sign. That is, as the time between trades increases (that is, as trading activity falls), the bid–ask spread widens. Furthermore, although the coefficient values are small, they are statistically significant. In the put spread regression, for example, the coefficient of approximately 0.009 implies that, even if the time between trades widened from one minute to one hour, the spread would increase by only 54 cents. α_4 and β_4 measure the effect of time to maturity on the spread; both are negative and statistically significant. The authors argue that this may arise as market making is a more risky activity for near-maturity options. A possible alternative explanation, which they dismiss after further investigation, is that the early exercise possibility becomes more likely for very short-dated options since the loss of time value would be negligible. Finally, α_5 and β_5 measure the effect of risk on the spread; in both the call and put spread regressions, these coefficients are negative and highly statistically significant. This seems an odd result, which the authors struggle to justify, for it seems to suggest that more risky options will command lower spreads.

Turning attention now to the trading activity regressions, γ_1 and δ_1 measure the effect of the spread size on call and put trading activity, respectively. Both are positive and statistically significant, indicating that a rise in the spread will increase the time between trades. The coefficients are such that a 1 *cent* increase in the spread would lead to an increase in the average time between call and put trades of nearly half a minute. γ_2 and δ_2 give the effect of an increase in time to maturity, while γ_3 and δ_3 are coefficients attached to the square of time to maturity. For both the call and put regressions, the coefficient on the level of time to maturity is negative and significant, while that on the square is positive and significant. As time to maturity increases, the squared term would dominate, and one could therefore conclude that the time between trades will show a U-shaped relationship with time to maturity. Finally, γ_4 and δ_4 give the effect of an increase in the square of moneyness (i.e. the effect of an option going deeper into the money or deeper out of the money) on the time between trades. For both the call and put regressions, the coefficients are statistically significant and positive, showing that as the option moves further from the money in either direction, the time between trades rises. This is consistent with the authors' supposition that trade is most active in at-the-money options, and less active in both out-of-the-money and in-the-money options.

6.9.6 *Conclusions*

The value of the bid–ask spread on S&P100 index options and the time between trades (a measure of market liquidity) can be usefully modelled in a simultaneous system with exogenous variables such as the options' deltas, time to maturity, moneyness, etc.

This study represents a nice example of the use of a simultaneous equations system, but, in this author's view, it can be criticised on several grounds. First, there are no diagnostic tests performed. Second, clearly the equations are all overidentified, but it is not obvious how the overidentifying restrictions have been generated. Did they arise from consideration of financial theory? For example, why do the *CL* and *PL* equations not contain the *CR* and *PR* variables? Why do the *CBA* and *PBA* equations not contain moneyness or squared maturity variables? The authors could also have tested for endogeneity of *CBA* and *CL*. Finally, the wrong sign on the highly statistically significant squared deltas is puzzling.

6.10 Simultaneous equations modelling using EViews

What is the relationship between inflation and stock returns? Holding stocks is often thought to provide a good hedge against inflation, since the payments to equity holders are not fixed in nominal terms and represent a claim on real assets (unlike the coupons on bonds, for example). However, the majority of empirical studies that have investigated the sign of this relationship have found it to be negative. Various explanations of this puzzling empirical phenomenon have been proposed, including a link through real activity, so that real activity is negatively related to inflation but positively related to stock returns and therefore stock returns and inflation vary positively. Clearly, inflation and stock returns ought to be simultaneously related given that the rate of inflation will affect the discount rate applied to cashflows and therefore the value of equities, but the performance of the stock market may also affect consumer demand and therefore inflation through its impact on householder wealth (perceived or actual).[1]

[1] Crucially, good econometric models are based on solid financial theory. This model is clearly not, but represents a simple way to illustrate the estimation and interpretation of simultaneous equations models using EViews with freely available data!

This simple example uses the same macroeconomic data as used previously to estimate this relationship simultaneously. Suppose (without justification) that we wish to estimate the following model, which does not allow for dynamic effects or partial adjustments and does not distinguish between expected and unexpected inflation

$$inflation_t = \alpha_0 + \alpha_1\ returns_t + \alpha_2\ dcredit_t + \alpha_3\ dprod_t + \alpha_4\ dmoney + u_{1t} \tag{6.59}$$

$$returns_t = \beta_0 + \beta_1\ dprod_t + \beta_2\ dspread_t + \beta_3\ inflation_t + \beta_4\ rterm_t + u_{2t} \tag{6.60}$$

where 'returns' are stock returns and all of the other variables are defined as in the previous example in chapter 4.

It is evident that there is feedback between the two equations since the *inflation* variable appears in the *stock returns* equation and vice versa. Are the equations identified? Since there are two equations, each will be identified if one variable is missing from that equation. Equation (6.59), the inflation equation, omits two variables. It does not contain the default spread or the term spread, and so is over-identified. Equation (6.60), the stock returns equation, omits two variables as well – the consumer credit and money supply variables – and so is over-identified too. Two-stage least squares (2SLS) is therefore the appropriate technique to use.

In EViews, to do this we need to specify a list of instruments, which would be all of the variables from the reduced form equation. In this case, the reduced form equations would be

$$inflation = f(constant, dprod, dspread, rterm, dcredit, qrev, dmoney) \tag{6.61}$$

$$returns = g(constant, dprod, dspread, rterm, dcredit, qrev, dmoney) \tag{6.62}$$

We can perform both stages of 2SLS in one go, but by default, EViews estimates each of the two equations in the system separately. To do this, click **Quick**, **Estimate Equation** and then select **TSLS – Two Stage Least Squares (TSNLS and ARMA)** from the list of estimation methods. Then fill in the dialog box as in screenshot 6.1 to estimate the inflation equation.

Thus the format of writing out the variables in the first window is as usual, and the full structural equation for inflation as a dependent variable should be specified here. In the instrument list, include every variable from the reduced form equation, including the constant, and click **OK**.

The results would then appear as in the following table.

Dependent Variable: INFLATION
Method: Two-Stage Least Squares
Date: 09/02/07 Time: 20:55
Sample (adjusted): 1986M04 2007M04
Included observations: 253 after adjustments
Instrument list: C DCREDIT DPROD RTERM DSPREAD DMONEY

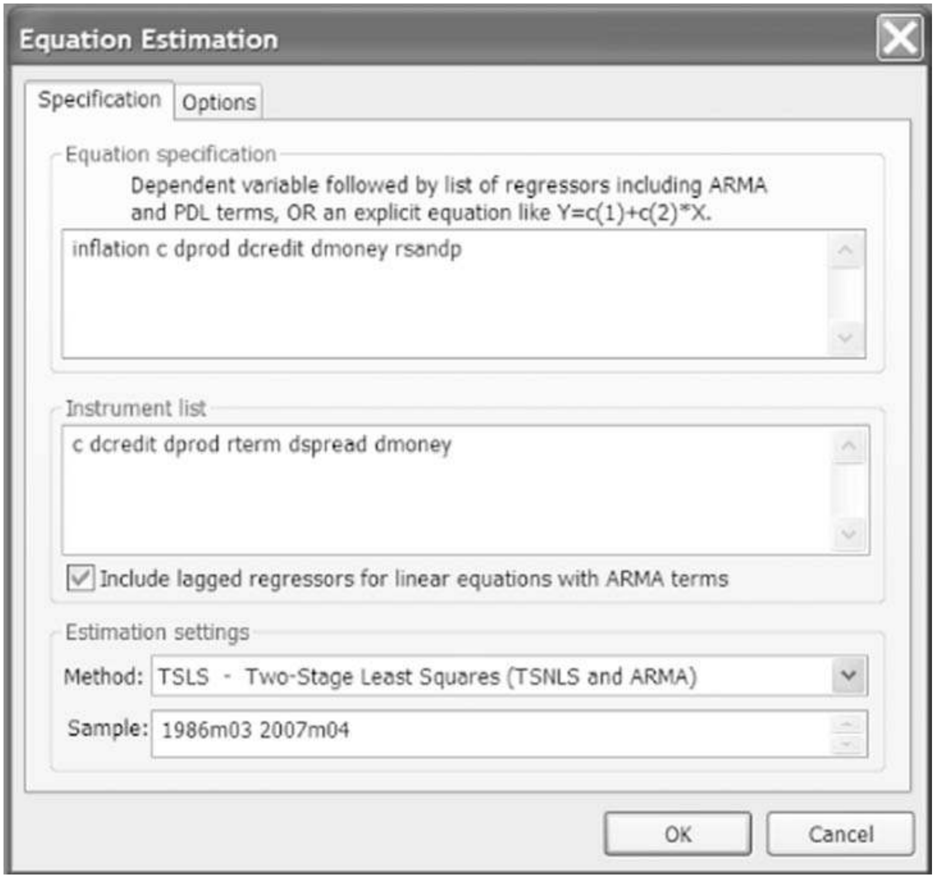
	Coefficient	Std. Error	t-Statistic	Prob.
C	0.066248	0.337932	0.196038	0.8447
DPROD	0.068352	0.090839	0.752453	0.4525
DCREDIT	4.77E-07	1.38E-05	0.034545	0.9725
DMONEY	0.027426	0.05882	0.466266	0.6414
RSANDP	0.238047	0.363113	0.655573	0.5127
R-squared	−15.398762	Mean dependent var		0.253632
Adjusted R-squared	−15.663258	S.D. dependent var		0.269221
S.E. of regression	1.098980	Sum squared resid		299.5236
F-statistic	0.179469	Durbin-Watson stat		1.923274
Prob(F-statistic)	0.948875	Second-Stage SSR		17.39799

Similarly, the dialog box for the rsandp equation would be specified as in screenshot 6.2. The output for the returns equation is shown in the following table.

Dependent Variable: RSANDP
Method: Two-Stage Least Squares
Date: 09/02/07 Time: 20:30
Sample (adjusted): 1986M04 2007M04
Included observations: 253 after adjustments
Instrument list: C DCREDIT DPROD RTERM DSPREAD DMONEY

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.682709	3.531687	0.193310	0.8469
DPROD	−0.242299	0.251263	−0.964322	0.3358
DSPREAD	−2.517793	10.57406	−0.238110	0.8120
RTERM	0.138109	1.263541	0.109303	0.9131
INFLATION	0.322398	14.10926	0.02285	0.9818
R-squared	0.006553	Mean dependent var		0.721483
Adjusted R-squared	−0.009471	S.D. dependent var		4.355220
S.E. of regression	4.375794	Sum squared resid		4748.599
F-statistic	0.688494	Durbin-Watson stat		2.017386
Prob(F-statistic)	0.600527	Second-Stage SSR		4727.189

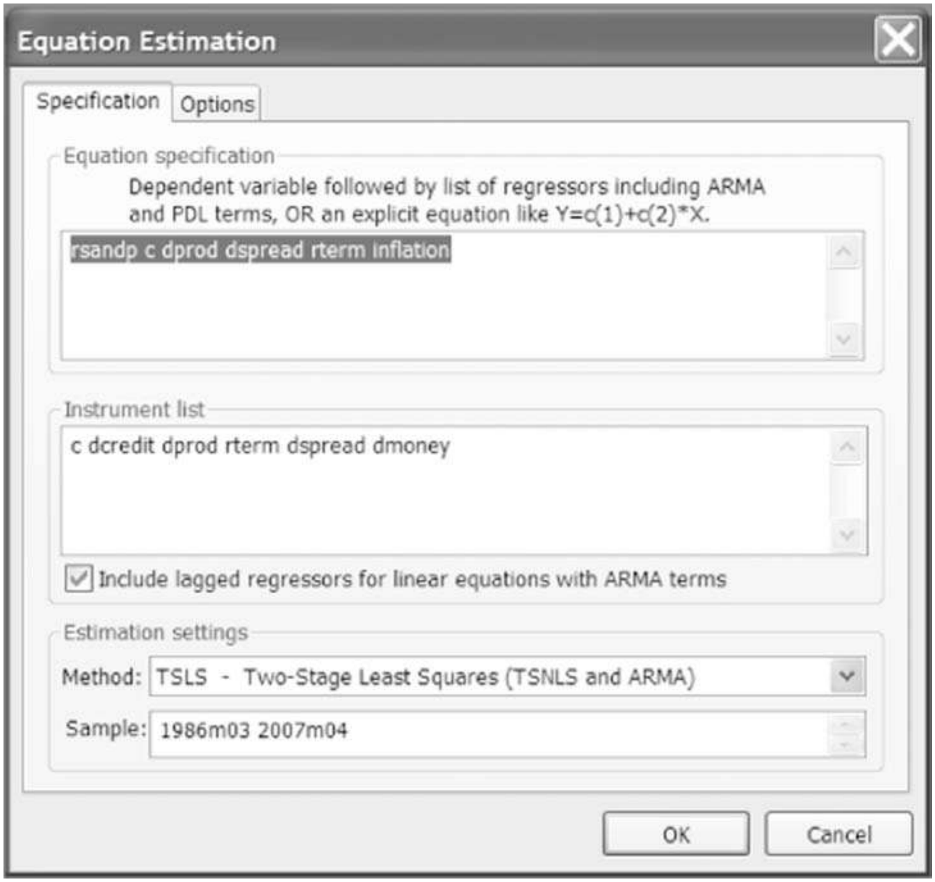
Screenshot 6.1
Estimating the inflation equation



The results overall are not very enlightening. None of the parameters is even close to statistical significance in either equation, although interestingly, the fitted relationship between the stock returns and inflation series is positive (albeit not significantly so). The $\bar{R}^2$ values from both equations are also negative, so should be interpreted with caution. As the EViews *User's Guide* warns, this can sometimes happen even when there is an intercept in the regression.

It may also be of relevance to conduct a Hausman test for the endogeneity of the inflation and stock return variables. To do this, **estimate the reduced form equations and save the residuals**. Then **create series of fitted values** by constructing new variables which are equal to the actual values minus the residuals. Call the fitted value series **inflation_fit** and **rsandp_fit**. Then estimate the structural equations (separately), adding the fitted values from the relevant reduced form equations. The two sets of

Screenshot 6.2
Estimating the rsandp equation



variables (in EViews format, with the dependent variables first followed by the lists of independent variables) are as follows.

For the stock returns equation:

rsandp c dprod dspread rterm inflation inflation_fit

and for the inflation equation:

inflation c dprod dcredit dmoney rsandp rsandp_fit

The conclusion is that the inflation fitted value term is not significant in the stock return equation and so inflation can be considered exogenous for stock returns. Thus it would be valid to simply estimate this equation (minus the fitted value term) on its own using OLS. But the fitted stock return term is significant in the inflation equation, suggesting that stock returns are endogenous.

6.11 Vector autoregressive models

Vector autoregressive models (VARs) were popularised in econometrics by Sims (1980) as a natural generalisation of univariate autoregressive models discussed in chapter 5. A VAR is a systems regression model (i.e. there is more than one dependent variable) that can be considered a kind of hybrid between the univariate time series models considered in chapter 5 and the simultaneous equations models developed previously in this chapter. VARs have often been advocated as an alternative to large-scale simultaneous equations structural models.

The simplest case that can be entertained is a bivariate VAR, where there are only two variables, y_{1t} and y_{2t}, each of whose current values depend on different combinations of the previous k values of both variables, and error terms

$$y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \cdots + \beta_{1k} y_{1t-k} + \alpha_{11} y_{2t-1} + \cdots + \alpha_{1k} y_{2t-k} + u_{1t} \tag{6.63}$$

$$y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \cdots + \beta_{2k} y_{2t-k} + \alpha_{21} y_{1t-1} + \cdots + \alpha_{2k} y_{1t-k} + u_{2t} \tag{6.64}$$

where u_{it} is a white noise disturbance term with $\mathrm{E}(u_{it}) = 0$, $(i = 1, 2)$, $\mathrm{E}(u_{1t}u_{2t}) = 0$.

As should already be evident, an important feature of the VAR model is its flexibility and the ease of generalisation. For example, the model could be extended to encompass moving average errors, which would be a multivariate version of an ARMA model, known as a VARMA. Instead of having only two variables, y_{1t} and y_{2t}, the system could also be expanded to include g variables, $y_{1t}, y_{2t}, y_{3t}, \ldots, y_{gt}$, each of which has an equation.

Another useful facet of VAR models is the compactness with which the notation can be expressed. For example, consider the case from above where $k = 1$, so that each variable depends only upon the immediately previous values of y_{1t} and y_{2t}, plus an error term. This could be written as

$$y_{1t} = \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + u_{1t} \tag{6.65}$$

$$y_{2t} = \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + u_{2t} \tag{6.66}$$

or

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{6.67}$$

or even more compactly as

$$\underset{g \times 1}{y_t} = \underset{g \times 1}{\beta_0} + \underset{g \times g g \times 1}{\beta_1 y_{t-1}} + \underset{g \times 1}{u_t} \tag{6.68}$$

In (6.68), there are $g = 2$ variables in the system. Extending the model to the case where there are k lags of each variable in each equation is also easily accomplished using this notation

$$\underset{g\times 1}{y_t} = \underset{g\times 1}{\beta_0} + \underset{g\times g}{\beta_1}\,\underset{g\times 1}{y_{t-1}} + \underset{g\times g}{\beta_2}\,\underset{g\times 1}{y_{t-2}} + \cdots + \underset{g\times g}{\beta_k}\,\underset{g\times 1}{y_{t-k}} + \underset{g\times 1}{u_t} \tag{6.69}$$

The model could be further extended to the case where the model includes first difference terms and cointegrating relationships (a vector error correction model (VECM) – see chapter 7).

6.11.1 *Advantages of VAR modelling*

VAR models have several advantages compared with univariate time series models or simultaneous equations structural models:

- The researcher does not need to specify which variables are endogenous or exogenous – *all are endogenous*. This is a very important point, since a requirement for simultaneous equations structural models to be estimable is that all equations in the system are identified. Essentially, this requirement boils down to a condition that some variables are treated as exogenous and that the equations contain different RHS variables. Ideally, this restriction should arise naturally from financial or economic theory. However, in practice theory will be at best vague in its suggestions of which variables should be treated as exogenous. This leaves the researcher with a great deal of discretion concerning how to classify the variables. Since Hausman-type tests are often not employed in practice when they should be, the specification of certain variables as exogenous, required to form identifying restrictions, is likely in many cases to be invalid. Sims termed these identifying restrictions 'incredible'. VAR estimation, on the other hand, requires no such restrictions to be imposed.
- VARs allow the value of a variable to depend on more than just its own lags or combinations of white noise terms, so VARs are more flexible than univariate AR models; the latter can be viewed as a restricted case of VAR models. VAR models can therefore offer a very *rich structure*, implying that they may be able to capture more features of the data.
- Provided that there are no contemporaneous terms on the RHS of the equations, it is possible to *simply use OLS separately on each equation.* This arises from the fact that all variables on the RHS are pre-determined – that is, at time t, they are known. This implies that there is no possibility

for feedback from any of the LHS variables to any of the RHS variables. Pre-determined variables include all exogenous variables and lagged values of the endogenous variables.

- The forecasts generated by VARs are often *better than 'traditional structural' models*. It has been argued in a number of articles (see, for example, Sims, 1980) that large-scale structural models performed badly in terms of their out-of-sample forecast accuracy. This could perhaps arise as a result of the ad hoc nature of the restrictions placed on the structural models to ensure identification discussed above. McNees (1986) shows that forecasts for some variables (e.g. the US unemployment rate and real GNP, etc.) are produced more accurately using VARs than from several different structural specifications.

6.11.2 Problems with VARs

VAR models of course also have drawbacks and limitations relative to other model classes:

- VARs are *a-theoretical* (as are ARMA models), since they use little theoretical information about the relationships between the variables to guide the specification of the model. On the other hand, valid exclusion restrictions that ensure identification of equations from a simultaneous structural system will inform on the structure of the model. An upshot of this is that VARs are less amenable to theoretical analysis and therefore to policy prescriptions. There also exists an increased possibility under the VAR approach that a hapless researcher could obtain an essentially spurious relationship by mining the data. It is also often not clear how the VAR coefficient estimates should be interpreted.
- How should the appropriate *lag lengths* for the VAR be determined? There are several approaches available for dealing with this issue, which will be discussed below.
- *So many parameters*! If there are g equations, one for each of g variables and with k lags of each of the variables in each equation, $(g + kg^2)$ parameters will have to be estimated. For example, if $g = 3$ and $k = 3$ there will be 30 parameters to estimate. For relatively small sample sizes, degrees of freedom will rapidly be used up, implying large standard errors and therefore wide confidence intervals for model coefficients.
- Should *all of the components of the VAR be stationary*? Obviously, if one wishes to use hypothesis tests, either singly or jointly, to examine the statistical significance of the coefficients, then it is essential that all of the components in the VAR are stationary. However, many proponents of the VAR approach recommend that differencing to induce

stationarity should not be done. They would argue that the purpose of VAR estimation is purely to examine the relationships between the variables, and that differencing will throw information on any long-run relationships between the series away. It is also possible to combine levels and first differenced terms in a VECM – see chapter 7.

6.11.3 Choosing the optimal lag length for a VAR

Often, financial theory will have little to say on what is an appropriate lag length for a VAR and how long changes in the variables should take to work through the system. In such instances, there are broadly two methods that could be used to arrive at the optimal lag length: cross-equation restrictions and information criteria.

6.11.4 Cross-equation restrictions for VAR lag length selection

A first (but incorrect) response to the question of how to determine the appropriate lag length would be to use the block *F*-tests highlighted in section 6.13 below. These, however, are not appropriate in this case as the *F*-test would be used separately for the set of lags in each equation, and what is required here is a procedure to test the coefficients on a set of lags on all variables for all equations in the VAR at the same time.

It is worth noting here that in the spirit of VAR estimation (as Sims, for example, thought that model specification should be conducted), the models should be as unrestricted as possible. A VAR with different lag lengths for each equation could be viewed as a restricted VAR. For example, consider a VAR with 3 lags of both variables in one equation and 4 lags of each variable in the other equation. This could be viewed as a restricted model where the coefficient on the fourth lags of each variable in the first equation have been set to zero.

An alternative approach would be to specify the same number of lags in each equation and to determine the model order as follows. Suppose that a VAR estimated using quarterly data has 8 lags of the two variables in each equation, and it is desired to examine a restriction that the coefficients on lags 5–8 are jointly zero. This can be done using a likelihood ratio test (see chapter 8 for more general details concerning such tests). Denote the variance–covariance matrix of residuals (given by $\hat{u}\hat{u}'$), as $\hat{\Sigma}$. The likelihood ratio test for this joint hypothesis is given by

$$LR = T[\log|\hat{\Sigma}_r| - \log|\hat{\Sigma}_u|] \tag{6.70}$$

where $|\hat{\Sigma}_r|$ is the determinant of the variance–covariance matrix of the residuals for the restricted model (with 4 lags), $|\hat{\Sigma}_u|$ is the determinant of the variance–covariance matrix of residuals for the unrestricted VAR (with 8 lags) and T is the sample size. The test statistic is asymptotically distributed as a χ^2 variate with degrees of freedom equal to the total number of restrictions. In the VAR case above, 4 lags of two variables are being restricted in each of the 2 equations = a total of $4 \times 2 \times 2 = 16$ restrictions. In the general case of a VAR with g equations, to impose the restriction that the last q lags have zero coefficients, there would be $g^2 q$ restrictions altogether. Intuitively, the test is a multivariate equivalent to examining the extent to which the *RSS* rises when a restriction is imposed. If $\hat{\Sigma}_r$ and $\hat{\Sigma}_u$ are 'close together', the restriction is supported by the data.

6.11.5 Information criteria for VAR lag length selection

The likelihood ratio (LR) test explained above is intuitive and fairly easy to estimate, but has its limitations. Principally, one of the two VARs must be a special case of the other and, more seriously, only pairwise comparisons can be made. In the above example, if the most appropriate lag length had been 7 or even 10, there is no way that this information could be gleaned from the LR test conducted. One could achieve this only by starting with a VAR(10), and successively testing one set of lags at a time.

A further disadvantage of the LR test approach is that the χ^2 test will strictly be valid asymptotically only under the assumption that the errors from each equation are normally distributed. This assumption is unlikely to be upheld for financial data. An alternative approach to selecting the appropriate VAR lag length would be to use an information criterion, as defined in chapter 5 in the context of ARMA model selection. Information criteria require no such normality assumptions concerning the distributions of the errors. Instead, the criteria trade off a fall in the *RSS* of each equation as more lags are added, with an increase in the value of the penalty term. The univariate criteria could be applied separately to each equation but, again, it is usually deemed preferable to require the number of lags to be the same for each equation. This requires the use of multivariate versions of the information criteria, which can be defined as

$$MAIC = \log\left|\hat{\Sigma}\right| + 2k'/T \tag{6.71}$$

$$MSBIC = \log\left|\hat{\Sigma}\right| + \frac{k'}{T}\log(T) \tag{6.72}$$

$$MHQIC = \log\left|\hat{\Sigma}\right| + \frac{2k'}{T}\log(\log(T)) \tag{6.73}$$

where again $\hat{\Sigma}$ is the variance–covariance matrix of residuals, T is the number of observations and k' is the total number of regressors in all equations, which will be equal to $p^2k + p$ for p equations in the VAR system, each with k lags of the p variables, plus a constant term in each equation. As previously, the values of the information criteria are constructed for 0, 1, ..., $\bar{k}$ lags (up to some pre-specified maximum $\bar{k}$), and the chosen number of lags is that number minimising the value of the given information criterion.

6.12 Does the VAR include contemporaneous terms?

So far, it has been assumed that the VAR specified is of the form

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + u_{1t} \tag{6.74}$$
$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + u_{2t} \tag{6.75}$$

so that there are no contemporaneous terms on the RHS of (6.74) or (6.75) – i.e. there is no term in y_{2t} on the RHS of the equation for y_{1t} and no term in y_{1t} on the RHS of the equation for y_{2t}. But what if the equations had a contemporaneous feedback term, as in the following case?

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + \alpha_{12}y_{2t} + u_{1t} \tag{6.76}$$
$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + \alpha_{22}y_{1t} + u_{2t} \tag{6.77}$$

Equations (6.76) and (6.77) could also be written by stacking up the terms into matrices and vectors:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{12} & 0 \\ 0 & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{2t} \\ y_{1t} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{6.78}$$

This would be known as a *VAR in primitive form*, similar to the structural form for a simultaneous equations model. Some researchers have argued that the a-theoretical nature of reduced form VARs leaves them unstructured and their results difficult to interpret theoretically. They argue that the forms of VAR given previously are merely reduced forms of a more general structural VAR (such as (6.78)), with the latter being of more interest.

The contemporaneous terms from (6.78) can be taken over to the LHS and written as

$$\begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{22} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{6.79}$$

or

$$\mathrm{A}y_t = \beta_0 + \beta_1 y_{t-1} + u_t \tag{6.80}$$

If both sides of (6.80) are pre-multiplied by A^{-1}

$$y_t = \mathrm{A}^{-1}\beta_0 + \mathrm{A}^{-1}\beta_1 y_{t-1} + \mathrm{A}^{-1}u_t \tag{6.81}$$

or

$$y_t = \mathrm{A}_0 + \mathrm{A}_1 y_{t-1} + e_t \tag{6.82}$$

This is known as a *standard form VAR*, which is akin to the reduced form from a set of simultaneous equations. This VAR contains only pre-determined values on the RHS (i.e. variables whose values are known at time t), and so there is no contemporaneous feedback term. This VAR can therefore be estimated equation by equation using OLS.

Equation (6.78), the structural or primitive form VAR, is not identified, since identical pre-determined (lagged) variables appear on the RHS of both equations. In order to circumvent this problem, a restriction that one of the coefficients on the contemporaneous terms is zero must be imposed. In (6.78), either α_{12} or α_{22} must be set to zero to obtain a triangular set of VAR equations that can be validly estimated. The choice of which of these two restrictions to impose is ideally made on theoretical grounds. For example, if financial theory suggests that the current value of y_{1t} should affect the current value of y_{2t} but not the other way around, set $\alpha_{12} = 0$, and so on. Another possibility would be to run separate estimations, first imposing $\alpha_{12} = 0$ and then $\alpha_{22} = 0$, to determine whether the general features of the results are much changed. It is also very common to estimate only a reduced form VAR, which is of course perfectly valid provided that such a formulation is not at odds with the relationships between variables that financial theory says should hold.

One fundamental weakness of the VAR approach to modelling is that its a-theoretical nature and the large number of parameters involved make the estimated models difficult to interpret. In particular, some lagged variables may have coefficients which change sign across the lags, and this, together with the interconnectivity of the equations, could render it difficult to see what effect a given change in a variable would have upon the future values of the variables in the system. In order to partially alleviate this problem, three sets of statistics are usually constructed for an estimated VAR model: block significance tests, impulse responses and variance decompositions. How important an intuitively interpretable model is will of course depend on the purpose of constructing the model. Interpretability may not be an issue at all if the purpose of producing the VAR is to make forecasts.

Table 6.3 Granger causality tests and implied restrictions on VAR models

	Hypothesis	Implied restriction
1	Lags of y_{1t} do not explain current y_{2t}	$\beta_{21} = 0$ and $\gamma_{21} = 0$ and $\delta_{21} = 0$
2	Lags of y_{1t} do not explain current y_{1t}	$\beta_{11} = 0$ and $\gamma_{11} = 0$ and $\delta_{11} = 0$
3	Lags of y_{2t} do not explain current y_{1t}	$\beta_{12} = 0$ and $\gamma_{12} = 0$ and $\delta_{12} = 0$
4	Lags of y_{2t} do not explain current y_{2t}	$\beta_{22} = 0$ and $\gamma_{22} = 0$ and $\delta_{22} = 0$

6.13 Block significance and causality tests

It is likely that, when a VAR includes many lags of variables, it will be difficult to see which sets of variables have significant effects on each dependent variable and which do not. In order to address this issue, tests are usually conducted that restrict all of the lags of a particular variable to zero. For illustration, consider the following bivariate VAR(3)

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{10} \\ \alpha_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix}$$
$$+ \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-3} \\ y_{2t-3} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{6.83}$$

This VAR could be written out to express the individual equations as

$$\begin{aligned} y_{1t} &= \alpha_{10} + \beta_{11} y_{1t-1} + \beta_{12} y_{2t-1} + \gamma_{11} y_{1t-2} + \gamma_{12} y_{2t-2} \\ &\quad + \delta_{11} y_{1t-3} + \delta_{12} y_{2t-3} + u_{1t} \\ y_{2t} &= \alpha_{20} + \beta_{21} y_{1t-1} + \beta_{22} y_{2t-1} + \gamma_{21} y_{1t-2} + \gamma_{22} y_{2t-2} \\ &\quad + \delta_{21} y_{1t-3} + \delta_{22} y_{2t-3} + u_{2t} \end{aligned} \tag{6.84}$$

One might be interested in testing the hypotheses and their implied restrictions on the parameter matrices given in table 6.3.

Assuming that all of the variables in the VAR are stationary, the joint hypotheses can easily be tested within the *F*-test framework, since each individual set of restrictions involves parameters drawn from only one equation. The equations would be estimated separately using OLS to obtain the unrestricted *RSS*, then the restrictions imposed and the models re-estimated to obtain the restricted *RSS*. The *F*-statistic would then take the usual form described in chapter 3. Thus, evaluation of the significance of variables in the context of a VAR almost invariably occurs on the basis of joint tests on all of the lags of a particular variable in an equation, rather than by examination of individual coefficient estimates.

In fact, the tests described above could also be referred to as causality tests. Tests of this form were described by Granger (1969) and a slight variant due to Sims (1972). Causality tests seek to answer simple questions of the type, 'Do changes in y_1 cause changes in y_2?' The argument follows that if y_1 causes y_2, lags of y_1 should be significant in the equation for y_2. If this is the case and not vice versa, it would be said that y_1 'Granger-causes' y_2 or that there exists unidirectional causality from y_1 to y_2. On the other hand, if y_2 causes y_1, lags of y_2 should be significant in the equation for y_1. If both sets of lags were significant, it would be said that there was 'bi-directional causality' or 'bi-directional feedback'. If y_1 is found to Granger-cause y_2, but not vice versa, it would be said that variable y_1 is strongly exogenous (in the equation for y_2). If neither set of lags are statistically significant in the equation for the other variable, it would be said that y_1 and y_2 are independent. Finally, the word 'causality' is somewhat of a misnomer, for Granger-causality really means only a correlation between the *current* value of one variable and the *past* values of others; it does not mean that movements of one variable cause movements of another.

6.14 VARs with exogenous variables

Consider the following specification for a VAR(1) where X_t is a vector of exogenous variables and B is a matrix of coefficients

$$y_t = A_0 + A_1 y_{t-1} + BX_t + e_t \tag{6.85}$$

The components of the vector X_t are known as exogenous variables since their values are determined outside of the VAR system – in other words, there are no equations in the VAR with any of the components of X_t as dependent variables. Such a model is sometimes termed a VARX, although it could be viewed as simply a restricted VAR where there are equations for each of the exogenous variables, but with the coefficients on the RHS in those equations restricted to zero. Such a restriction may be considered desirable if theoretical considerations suggest it, although it is clearly not in the true spirit of VAR modelling, which is not to impose any restrictions on the model but rather to 'let the data decide'.

6.15 Impulse responses and variance decompositions

Block *F*-tests and an examination of causality in a VAR will suggest which of the variables in the model have statistically significant impacts on the

Box 6.3 Forecasting with VARs

One of the main advantages of the VAR approach to modelling and forecasting is that since only lagged variables are used on the right hand side, forecasts of the future values of the dependent variables can be calculated using only information from within the system. We could term these *unconditional forecasts* since they are not constructed conditional on a particular set of assumed values. However, conversely it may be useful to produce forecasts of the future values of some variables *conditional upon* known values of other variables in the system. For example, it may be the case that the values of some variables become known before the values of the others. If the known values of the former are employed, we would anticipate that the forecasts should be more accurate than if estimated values were used unnecessarily, thus throwing known information away. Alternatively, conditional forecasts can be employed for counterfactual analysis based on examining the impact of certain scenarios. For example, in a trivariate VAR system incorporating monthly stock returns, inflation and GDP, we could answer the question: 'What is the likely impact on the stock market over the next 1–6 months of a 2-percentage point increase in inflation and a 1% rise in GDP?'

future values of each of the variables in the system. But *F*-test results will not, by construction, be able to explain the sign of the relationship or how long these effects require to take place. That is, *F*-test results will not reveal whether changes in the value of a given variable have a positive or negative effect on other variables in the system, or how long it would take for the effect of that variable to work through the system. Such information will, however, be given by an examination of the VAR's impulse responses and variance decompositions.

Impulse responses trace out the responsiveness of the dependent variables in the VAR to shocks to each of the variables. So, for each variable from each equation separately, a unit shock is applied to the error, and the effects upon the VAR system over time are noted. Thus, if there are g variables in a system, a total of g^2 impulse responses could be generated. The way that this is achieved in practice is by expressing the VAR model as a VMA – that is, the vector autoregressive model is written as a vector moving average (in the same way as was done for univariate autoregressive models in chapter 5). Provided that the system is stable, the shock should gradually die away.

To illustrate how impulse responses operate, consider the following bivariate VAR(1)

$$y_t = A_1 y_{t-1} + u_t \tag{6.86}$$

$$\text{where } A_1 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix}$$

The VAR can also be written out using the elements of the matrices and vectors as

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{6.87}$$

Consider the effect at time $t = 0, 1, \ldots,$ of a unit shock to y_{1t} at time $t = 0$

$$y_0 = \begin{bmatrix} u_{10} \\ u_{20} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{6.88}$$

$$y_1 = A_1 y_0 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} \tag{6.89}$$

$$y_2 = A_1 y_1 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0 \end{bmatrix} \tag{6.90}$$

and so on. It would thus be possible to plot the impulse response functions of y_{1t} and y_{2t} to a unit shock in y_{1t}. Notice that the effect on y_{2t} is always zero, since the variable y_{1t-1} has a zero coefficient attached to it in the equation for y_{2t}.

Now consider the effect of a unit shock to y_{2t} at time $t = 0$

$$y_0 = \begin{bmatrix} u_{10} \\ u_{20} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{6.91}$$

$$y_1 = A_1 y_0 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix} \tag{6.92}$$

$$y_2 = A_1 y_1 = \begin{bmatrix} 0.5 & 0.3 \\ 0.0 & 0.2 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.21 \\ 0.04 \end{bmatrix} \tag{6.93}$$

and so on. Although it is probably fairly easy to see what the effects of shocks to the variables will be in such a simple VAR, the same principles can be applied in the context of VARs containing more equations or more lags, where it is much more difficult to see by eye what are the interactions between the equations.

Variance decompositions offer a slightly different method for examining VAR system dynamics. They give the proportion of the movements in the dependent variables that are due to their 'own' shocks, versus shocks to the other variables. A shock to the ith variable will directly affect that variable of course, but it will also be transmitted to all of the other variables in the system through the dynamic structure of the VAR. Variance decompositions determine how much of the s-step-ahead forecast error variance of a given variable is explained by innovations to each explanatory variable for $s = 1, 2, \ldots$ In practice, it is usually observed that own series shocks explain most of the (forecast) error variance of the series in a VAR. To some extent, impulse responses and variance decompositions offer very similar information.

For calculating impulse responses and variance decompositions, the ordering of the variables is important. To see why this is the case, recall that the impulse responses refer to a unit shock to the errors of one VAR equation alone. This implies that the error terms of all other equations in the VAR system are held constant. However, this is not realistic since the error terms are likely to be correlated across equations to some extent. Thus, assuming that they are completely independent would lead to a misrepresentation of the system dynamics. In practice, the errors will have a common component that cannot be associated with a single variable alone.

The usual approach to this difficulty is to generate *orthogonalised impulse responses*. In the context of a bivariate VAR, the whole of the common component of the errors is attributed somewhat arbitrarily to the first variable in the VAR. In the general case where there are more than two variables in the VAR, the calculations are more complex but the interpretation is the same. Such a restriction in effect implies an 'ordering' of variables, so that the equation for y_{1t} would be estimated first and then that of y_{2t}, a bit like a recursive or triangular system.

Assuming a particular ordering is necessary to compute the impulse responses and variance decompositions, although the restriction underlying the ordering used may not be supported by the data. Again, ideally, financial theory should suggest an ordering (in other words, that movements in some variables are likely to follow, rather than precede, others). Failing this, the sensitivity of the results to changes in the ordering can be observed by assuming one ordering, and then exactly reversing it and re-computing the impulse responses and variance decompositions. It is also worth noting that the more highly correlated are the residuals from an estimated equation, the more the variable ordering will be important. But when the residuals are almost uncorrelated, the ordering of the variables will make little difference (see Lütkepohl, 1991, chapter 2 for further details).

Runkle (1987) argues that both impulse responses and variance decompositions are notoriously difficult to interpret accurately. He argues that confidence bands around the impulse responses and variance decompositions should always be constructed. However, he further states that, even then, the confidence intervals are typically so wide that sharp inferences are impossible.

6.16 VAR model example: the interaction between property returns and the macroeconomy

6.16.1 Background, data and variables

Brooks and Tsolacos (1999) employ a VAR methodology for investigating the interaction between the UK property market and various macroeconomic variables. Monthly data, in logarithmic form, are used for the period from December 1985 to January 1998. The selection of the variables for inclusion in the VAR model is governed by the time series that are commonly included in studies of stock return predictability. It is assumed that stock returns are related to macroeconomic and business conditions, and hence time series which may be able to capture both current and future directions in the broad economy and the business environment are used in the investigation.

Broadly, there are two ways to measure the value of property-based assets – *direct measures of property value* and *equity-based measures*. Direct property measures are based on periodic appraisals or valuations of the actual properties in a portfolio by surveyors, while equity-based measures evaluate the worth of properties indirectly by considering the values of stock market traded property companies. Both sources of data have their drawbacks. Appraisal-based value measures suffer from valuation biases and inaccuracies. Surveyors are typically prone to 'smooth' valuations over time, such that the measured returns are too low during property market booms and too high during periods of property price falls. Additionally, not every property in the portfolio that comprises the value measure is appraised during every period, resulting in some stale valuations entering the aggregate valuation, further increasing the degree of excess smoothness of the recorded property price series. Indirect property vehicles – property-related companies traded on stock exchanges – do not suffer from the above problems, but are excessively influenced by general stock market movements. It has been argued, for example, that over three-quarters of the variation over time in the value of stock exchange traded property companies can be attributed to general stock market-wide price movements. Therefore, the value of equity-based property series reflects much more the sentiment in the general stock market than the sentiment in the property market specifically.

Brooks and Tsolacos (1999) elect to use the equity-based FTSE Property Total Return Index to construct property returns. In order to purge the real estate return series of its general stock market influences, it is common to regress property returns on a general stock market index (in this case the FTA All-Share Index is used), saving the residuals. These residuals are expected to reflect only the variation in property returns, and thus become the property market return measure used in subsequent analysis, and are denoted PROPRES.

Hence, the variables included in the VAR are the property returns (with general stock market effects removed), the rate of unemployment, nominal interest rates, the spread between the long- and short-term interest rates, unanticipated inflation and the dividend yield. The motivations for including these particular variables in the VAR together with the property series, are as follows:

- *The rate of unemployment* (denoted UNEM) is included to indicate general economic conditions. In US research, authors tend to use aggregate consumption, a variable that has been built into asset pricing models and examined as a determinant of stock returns. Data for this variable and for alternative variables such as GDP are not available on a monthly basis in the UK. Monthly data are available for industrial production series but other studies have not shown any evidence that industrial production affects real estate returns. As a result, this series was not considered as a potential causal variable.
- *Short-term nominal interest rates* (denoted SIR) are assumed to contain information about future economic conditions and to capture the state of investment opportunities. It was found in previous studies that short-term interest rates have a very significant negative influence on property stock returns.
- *Interest rate spreads* (denoted SPREAD), i.e. the yield curve, are usually measured as the difference in the returns between long-term Treasury Bonds (of maturity, say, 10 or 20 years), and the one-month or three-month Treasury Bill rate. It has been argued that the yield curve has extra predictive power, beyond that contained in the short-term interest rate, and can help predict GDP up to four years ahead. It has also been suggested that the term structure also affects real estate market returns.
- *Inflation rate* influences are also considered important in the pricing of stocks. For example, it has been argued that unanticipated inflation could be a source of economic risk and as a result, a risk premium will also be added if the stock of firms has exposure to unanticipated inflation. The unanticipated inflation variable (denoted UNINFL) is defined as the difference between the realised inflation rate, computed as the percentage change in the Retail Price Index (RPI), and an estimated series of expected inflation. The latter series was produced by fitting an ARMA

model to the actual series and making a one-period(month)-ahead forecast, then rolling the sample forward one period, and re-estimating the parameters and making another one-step-ahead forecast, and so on.

- *Dividend yields* (denoted DIVY) have been widely used to model stock market returns, and also real estate property returns, based on the assumption that movements in the dividend yield series are related to long-term business conditions and that they capture some predictable components of returns.

All variables to be included in the VAR are required to be stationary in order to carry out joint significance tests on the lags of the variables. Hence, all variables are subjected to augmented Dickey–Fuller (ADF) tests (see chapter 7). Evidence that the log of the RPI and the log of the unemployment rate both contain a unit root is observed. Therefore, the first differences of these variables are used in subsequent analysis. The remaining four variables led to rejection of the null hypothesis of a unit root in the log-levels, and hence these variables were not first differenced.

6.16.2 *Methodology*

A reduced form VAR is employed and therefore each equation can effectively be estimated using OLS. For a VAR to be unrestricted, it is required that the same number of lags of all of the variables is used in all equations. Therefore, in order to determine the appropriate lag lengths, the multivariate generalisation of Akaike's information criterion (*AIC*) is used.

Within the framework of the VAR system of equations, the significance of all the lags of each of the individual variables is examined jointly with an F-test. Since several lags of the variables are included in each of the equations of the system, the coefficients on individual lags may not appear significant for all lags, and may have signs and degrees of significance that vary with the lag length. However, F-tests will be able to establish whether all of the lags of a particular variable are jointly significant. In order to consider further the effect of the macroeconomy on the real estate returns index, the impact multipliers (orthogonalised impulse responses) are also calculated for the estimated VAR model. Two standard error bands are calculated using the Monte Carlo integration approach employed by McCue and Kling (1994), and based on Doan (1994). The forecast error variance is also decomposed to determine the proportion of the movements in the real estate series that are a consequence of its own shocks rather than shocks to other variables.

Table 6.4 Marginal significance levels associated with joint *F*-tests

Dependent variable	Lags of variable					
	SIR	DIVY	SPREAD	UNEM	UNINFL	PROPRES
SIR	0.0000	0.0091	0.0242	0.0327	0.2126	0.0000
DIVY	0.5025	0.0000	0.6212	0.4217	0.5654	0.4033
SPREAD	0.2779	0.1328	0.0000	0.4372	0.6563	0.0007
UNEM	0.3410	0.3026	0.1151	0.0000	0.0758	0.2765
UNINFL	0.3057	0.5146	0.3420	0.4793	0.0004	0.3885
PROPRES	0.5537	0.1614	0.5537	0.8922	0.7222	0.0000

The test is that all 14 lags have no explanatory power for that particular equation in the VAR.
Source: Brooks and Tsolacos (1999).

6.16.3 Results

The number of lags that minimises the value of Akaike's information criterion is 14, consistent with the 15 lags used by McCue and Kling (1994). There are thus $(1 + 14 \times 6) = 85$ variables in each equation, implying 59 degrees of freedom. *F*-tests for the null hypothesis that all of the lags of a given variable are jointly insignificant in a given equation are presented in table 6.4.

In contrast to a number of US studies which have used similar variables, it is found to be difficult to explain the variation in the UK real estate returns index using macroeconomic factors, as the last row of table 6.4 shows. Of all the lagged variables in the real estate equation, only the lags of the real estate returns themselves are highly significant, and the dividend yield variable is significant only at the 20% level. No other variables have any significant explanatory power for the real estate returns. Therefore, based on the *F*-tests, an initial conclusion is that the variation in property returns, net of stock market influences, cannot be explained by any of the main macroeconomic or financial variables used in existing research. One possible explanation for this might be that, in the UK, these variables do not convey the information about the macroeconomy and business conditions assumed to determine the intertemporal behaviour of property returns. It is possible that property returns may reflect property market influences, such as rents, yields or capitalisation rates, rather than macroeconomic or financial variables. However, again the use of monthly data limits the set of both macroeconomic and property market variables that can be used in the quantitative analysis of real estate returns in the UK.

Table 6.5 Variance decompositions for the property sector index residuals

	Explained by innovations in											
	SIR		*DIVY*		*SPREAD*		*UNEM*		*UNINFL*		*PROPRES*	
Months ahead	I	II	I	II	I	II	I	II	I	II	I	II
1	0.0	0.8	0.0	38.2	0.0	9.1	0.0	0.7	0.0	0.2	100.0	51.0
2	0.2	0.8	0.2	35.1	0.2	12.3	0.4	1.4	1.6	2.9	97.5	47.5
3	3.8	2.5	0.4	29.4	0.2	17.8	1.0	1.5	2.3	3.0	92.3	45.8
4	3.7	2.1	5.3	22.3	1.4	18.5	1.6	1.1	4.8	4.4	83.3	51.5
12	2.8	3.1	15.5	8.7	15.3	19.5	3.3	5.1	17.0	13.5	46.1	50.0
24	8.2	6.3	6.8	3.9	38.0	36.2	5.5	14.7	18.1	16.9	23.4	22.0

Source: Brooks and Tsolacos (1999).

It appears, however, that lagged values of the real estate variable have explanatory power for some other variables in the system. These results are shown in the last column of table 6.4. The property sector appears to help in explaining variations in the term structure and short-term interest rates, and moreover since these variables are not significant in the property index equation, it is possible to state further that the property residual series Granger-causes the short-term interest rate and the term spread. This is a bizarre result. The fact that property returns are explained by own lagged values – i.e. that is there is interdependency between neighbouring data points (observations) – may reflect the way that property market information is produced and reflected in the property return indices.

Table 6.5 gives variance decompositions for the property returns index equation of the VAR for 1, 2, 3, 4, 12 and 24 steps ahead for the two variable orderings:

Order I: PROPRES, DIVY, UNINFL, UNEM, SPREAD, SIR
Order II: SIR, SPREAD, UNEM, UNINFL, DIVY, PROPRES.

Unfortunately, the ordering of the variables is important in the decomposition. Thus two orderings are applied, which are the exact opposite of one another, and the sensitivity of the result is considered. It is clear that by the two-year forecasting horizon, the variable ordering has become almost irrelevant in most cases. An interesting feature of the results is that shocks to the term spread and unexpected inflation together account for over 50% of the variation in the real estate series. The short-term interest rate and dividend yield shocks account for only 10–15% of the variance of

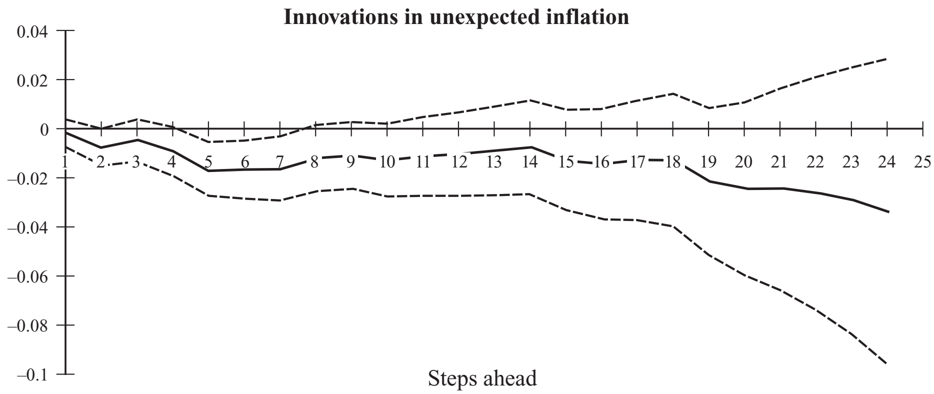


Figure 6.1 Impulse responses and standard error bands for innovations in unexpected inflation equation errors

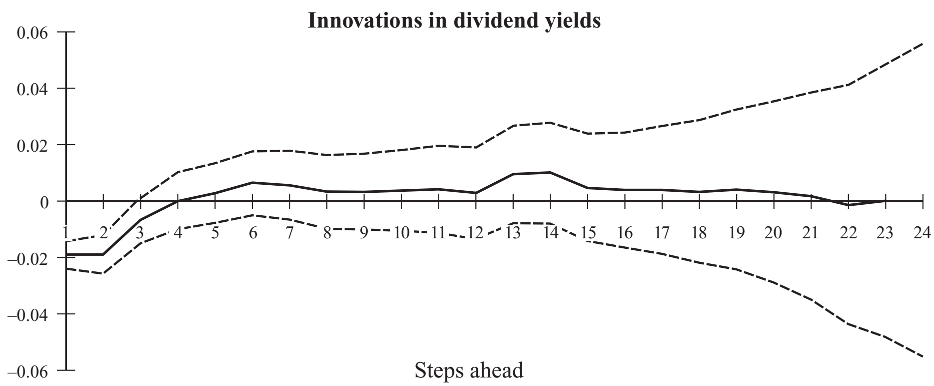


Figure 6.2 Impulse responses and standard error bands for innovations in the dividend yields

the property index. One possible explanation for the difference in results between the *F*-tests and the variance decomposition is that the former is a causality test and the latter is effectively an exogeneity test. Hence the latter implies the stronger restriction that both current and lagged shocks to the explanatory variables do not influence the current value of the dependent variable of the property equation. Another way of stating this is that the term structure and unexpected inflation have a contemporaneous rather than a lagged effect on the property index, which implies insignificant *F*-test statistics but explanatory power in the variance decomposition. Therefore, although the *F*-tests did not establish any significant effects, the error variance decompositions show evidence of a contemporaneous relationship between PROPRES and both SPREAD and UNINFL. The lack of lagged effects could be taken to imply speedy adjustment of the market to changes in these variables.

Figures 6.1 and 6.2 give the impulse responses for PROPRES associated with separate unit shocks to unexpected inflation and the dividend yield, as examples (as stated above, a total of 36 impulse responses could be calculated since there are 6 variables in the system).

Considering the signs of the responses, innovations to unexpected inflation (figure 6.1) always have a negative impact on the real estate index, since the impulse response is negative, and the effect of the shock does not die down, even after 24 months. Increasing stock dividend yields (figure 6.2) have a negative impact for the first three periods, but beyond that, the shock appears to have worked its way out of the system.

6.16.4 *Conclusions*

The conclusion from the VAR methodology adopted in the Brooks and Tsolacos paper is that overall, UK real estate returns are difficult to explain on the basis of the information contained in the set of the variables used in existing studies based on non-UK data. The results are not strongly suggestive of any significant influences of these variables on the variation of the filtered property returns series. There is, however, some evidence that the interest rate term structure and unexpected inflation have a contemporaneous effect on property returns, in agreement with the results of a number of previous studies.

6.17 VAR estimation in EViews

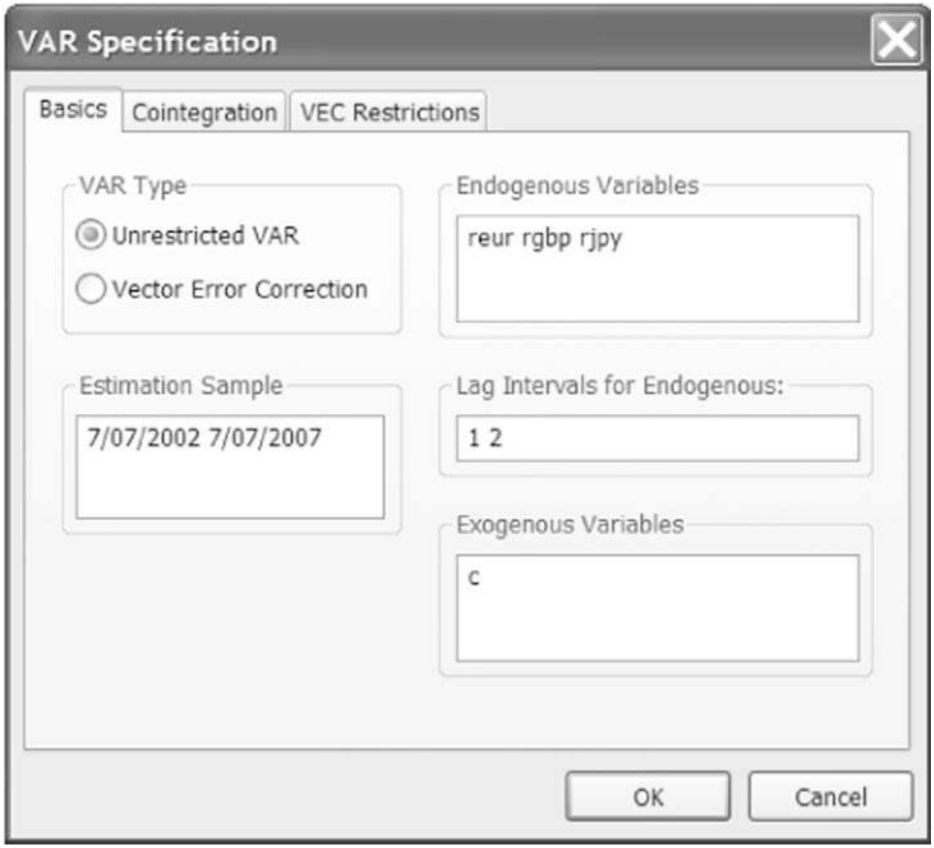
By way of illustration, a VAR is estimated in order to examine whether there are lead–lag relationships for the returns to three exchange rates against the US dollar – the euro, the British pound and the Japanese yen. The data are daily and run from 7 July 2002 to 7 July 2007, giving a total of 1,827 observations. The data are contained in the Excel file 'currencies.xls'. First **Create a new workfile**, called 'currencies.wf1', and import the three currency series. Construct a set of continuously compounded percentage returns called **'reur', 'rgbp' and 'rjpy'.** VAR estimation in EViews can be accomplished by clicking on the **Quick** menu and then **Estimate VAR**. The VAR inputs screen appears as in screenshot 6.3.

In the Endogenous variables box, type the three variable names, **reur rgbp rjpy**. In the Exogenous box, leave the default 'C' and in the Lag Interval box, enter **1 2** to estimate a VAR(2), just as an example. The output appears in a neatly organised table as shown on the following page, with one column for each equation in the first and second panels, and a single column of statistics that describes the system as a whole in the third. So values of the information criteria are given separately for each equation in the second panel and jointly for the model as a whole in the third.

Vector Autoregression Estimates
Date: 09/03/07 Time: 21:54
Sample (adjusted): 7/10/2002 7/07/2007
Included observations: 1824 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	REUR	RGBP	RJPY
REUR(−1)	0.031460	0.016776	0.040970
	(0.03681)	(0.03234)	(0.03444)
	[0.85471]	[0.51875]	[1.18944]
REUR(−2)	0.011377	0.045542	0.030551
	(0.03661)	(0.03217)	(0.03426)
	[0.31073]	[1.41574]	[0.89167]
RGBP(−1)	−0.070259	0.040547	−0.060907
	(0.04051)	(0.03559)	(0.03791)
	[−1.73453]	[1.13933]	[−1.60683]
RGBP(-2)	0.026719	−0.015074	−0.019407
	(0.04043)	(0.03552)	(0.03784)
	[0.66083]	[−0.42433]	[−0.51293]
RJPY(-1)	−0.020698	−0.029766	0.011809
	(0.03000)	(0.02636)	(0.02807)
	[−0.68994]	[−1.12932]	[0.42063]
RJPY(-2)	−0.014817	−0.000392	0.035524
	(0.03000)	(0.02635)	(0.02807)
	[−0.49396]	[−0.01489]	[1.26557]
C	−0.017229	−0.012878	0.002187
	(0.01100)	(0.00967)	(0.01030)
	[−1.56609]	[−1.33229]	[0.21239]
R-squared	0.003403	0.004040	0.003797
Adj. R-squared	0.000112	0.000751	0.000507
Sum sq. resids	399.0767	308.0701	349.4794
S.E. equation	0.468652	0.411763	0.438564
F-statistic	1.034126	1.228431	1.154191
Log likelihood	−1202.238	−966.1886	−1081.208
Akaike AIC	1.325919	1.067093	1.193210
Schwarz SC	1.347060	1.088234	1.214351
Mean dependent	−0.017389	−0.014450	0.002161
S.D. dependent	0.468679	0.411918	0.438676
Determinant resid covariance (dof adj.)		0.002214	
Determinant resid covariance		0.002189	
Log likelihood		−2179.054	
Akaike information criterion		2.412339	
Schwarz criterion		2.475763	

Screenshot 6.3
VAR inputs screen



We will shortly discuss the interpretation of the output, but the example so far has assumed that we know *the appropriate lag length* for the VAR. However, in practice, the first step in the construction of any VAR model, once the variables that will enter the VAR have been decided, will be to determine the appropriate lag length. This can be achieved in a variety of ways, but one of the easiest is to employ a multivariate information criterion. In EViews, this can be done easily from the EViews VAR output we have by clicking **View/Lag Structure/Lag Length Criteria**. . . . You will be invited to specify the maximum number of lags to entertain including in the model, and for this example, arbitrarily select **10**. The output in the following table would be observed.

EViews presents the values of various information criteria and other methods for determining the lag order. In this case, the Schwartz and Hannan–Quinn criteria both select a zero order as optimal, while Akaike's criterion chooses a VAR(1). **Estimate a VAR(1)** and examine the results. Does the model look as if it fits the data well? Why or why not?

VAR Lag Order Selection Criteria
Endogenous variables: REUR RGBP RJPY
Exogenous variables: C
Date: 09/03/07 Time: 21:58
Sample: 7/07/2002 7/07/2007
Included observations: 1816

Lag	LogL	LR	FPE	AIC	SC	HQ
0	−2192.395	NA	0.002252	2.417836	2.426929*	2.421191*
1	−2175.917	32.88475	0.002234*	2.409600*	2.445973	2.423020
2	−2170.888	10.01901	0.002244	2.413973	2.477625	2.437459
3	−2167.760	6.221021	0.002258	2.420441	2.511372	2.453992
4	−2158.361	18.66447	0.002257	2.420001	2.538212	2.463617
5	−2151.563	13.47494	0.002263	2.422426	2.567917	2.476109
6	−2145.132	12.72714	0.002269	2.425256	2.598026	2.489004
7	−2141.412	7.349932	0.002282	2.431071	2.631120	2.504884
8	−2131.693	19.17197	0.002281	2.430278	2.657607	2.514157
9	−2121.823	19.43540*	0.002278	2.429320	2.683929	2.523264
10	−2119.745	4.084453	0.002296	2.436944	2.718832	2.540953

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Next, run a Granger causality test by clicking **View/Lag Structure/Granger Causality/Block Exogeneity Tests**. The table of statistics will appear immediately as on the following page.

The results, unsurprisingly, show very little evidence of lead–lag interactions between the series. Since we have estimated a tri-variate VAR, three panels are displayed, with one for each dependent variable in the system. None of the results shows any causality that is significant at the 5% level, although there is causality from the pound to the euro and from the pound to the yen that is almost significant at the 10% level, but no causality in the opposite direction and no causality between the euro–dollar and the yen–dollar in either direction. These results might be interpreted as suggesting that information is incorporated slightly more quickly in the pound–dollar rate than in the euro–dollar or yen–dollar rates.

It is worth also noting that the term ‘Granger causality’ is something of a misnomer since a finding of ‘causality’ does not mean that movements

VAR Granger Causality/Block Exogeneity Wald Tests
Date: 09/04/07 Time: 13:50
Sample: 7/07/2002 7/07/2007
Included observations: 1825

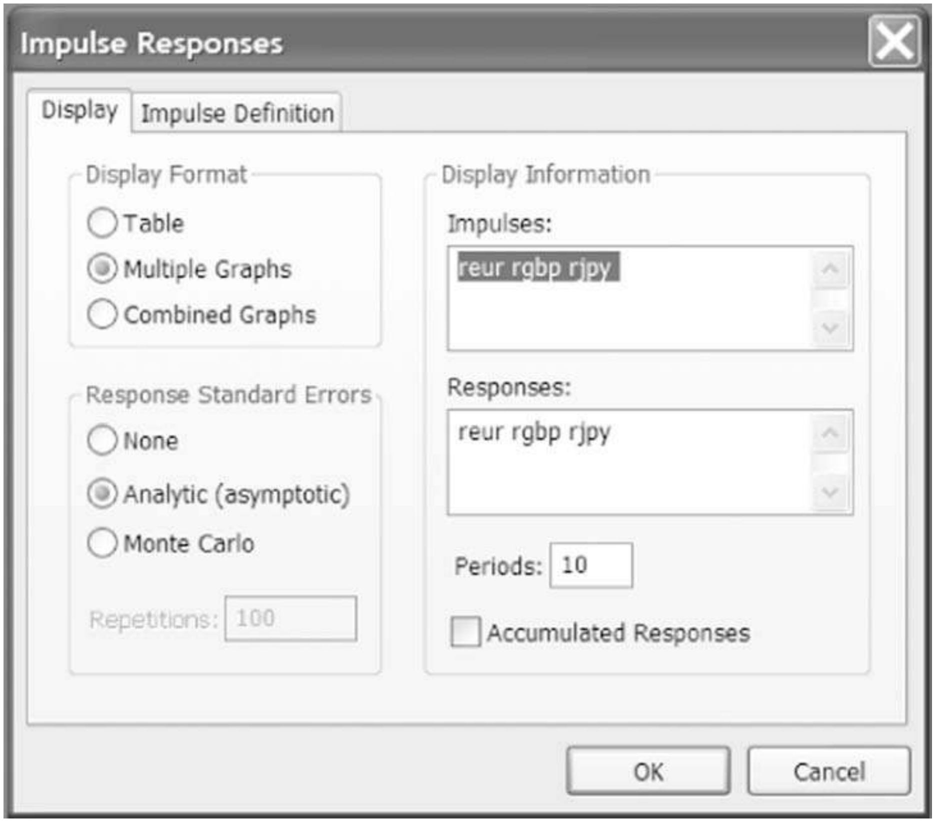
Dependent variable: REUR			
Excluded	Chi-sq	df	Prob.
RGBP	2.617817	1	0.1057
RJPY	0.473950	1	0.4912
All	3.529180	2	0.1713
Dependent variable: RGBP			
Excluded	Chi-sq	df	Prob.
REUR	0.188122	1	0.6645
RJPY	1.150696	1	0.2834
All	1.164752	2	0.5586
Dependent variable: RJPY			
Excluded	Chi-sq	df	Prob.
REUR	1.206092	1	0.2721
RGBP	2.424066	1	0.1195
All	2.435252	2	0.2959

in one variable physically cause movements in another. For example, in the above analysis, if movements in the euro–dollar market were found to Granger-cause movements in the pound–dollar market, this would not have meant that the pound–dollar rate changed as a direct result of, or because of, movements in the euro–dollar market. Rather, causality simply implies a *chronological ordering of movements in the series*. It could validly be stated that movements in the pound–dollar rate appear to lead those of the euro–dollar rate, and so on.

The EViews manual suggests that block *F*-test restrictions can be performed by estimating the VAR equations individually using OLS and then by using the **View** then **Lag Structure** then **Lag Exclusion Tests**. EViews tests for whether the parameters for a given lag of all the variables in a particular equation can be restricted to zero.

To obtain the impulse responses for the estimated model, simply click the **Impulse** on the button bar above the VAR object and a new dialog box will appear as in screenshot 6.4.

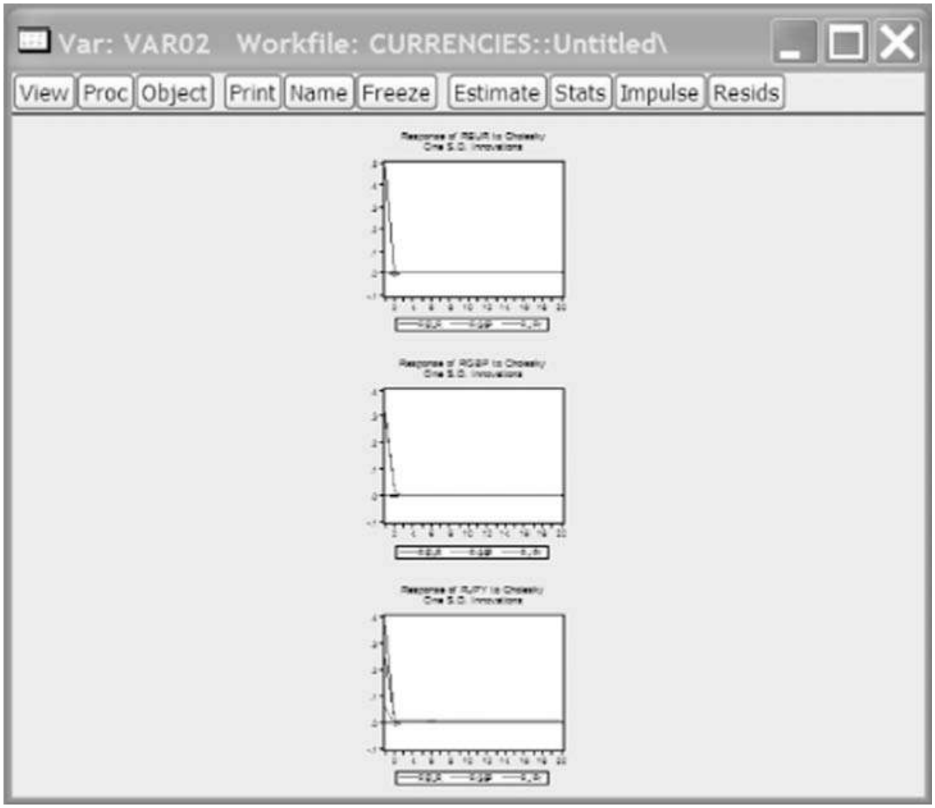
Screenshot 6.4
Constructing the VAR impulse responses



By default, EViews will offer to estimate and plot all of the responses to separate shocks of all of the variables in the order that the variables were listed in the estimation window, using ten steps and confidence intervals generated using analytic formulae. If 20 steps ahead had been selected, with 'combined response graphs', you would see the graphs in the format in screenshot 6.5 (obviously they appear small on the page and the colour has been lost, but the originals are much clearer). As one would expect given the parameter estimates and the Granger causality test results, again few linkages between the series are established here. The responses to the shocks are very small, except for the response of a variable to its own shock, and they die down to almost nothing after the first lag.

Plots of the variance decompositions can also be generated by clicking on **View** and then **Variance Decomposition**. A similar plot for the variance decompositions would appear as in screenshot 6.6.

There is little again that can be seen from these variance decomposition graphs that appear small on a printed page apart from the fact that the

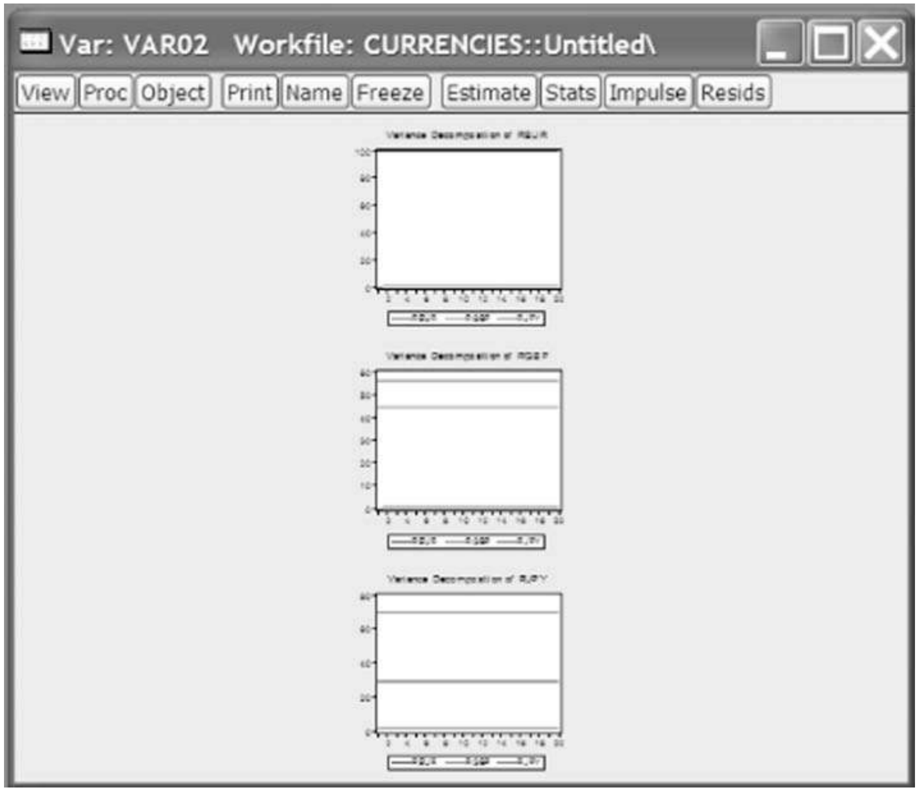


Screenshot 6.5
Combined impulse response graphs

behaviour is observed to settle down to a steady state very quickly. Interestingly, while the percentage of the errors that is attributable to own shocks is 100% in the case of the euro rate, for the pound, the euro series explains around 55% of the variation in returns, and for the yen, the euro series explains around 30% of the variation.

We should remember that the ordering of the variables has an effect on the impulse responses and variance decompositions, and when, as in this case, theory does not suggest an obvious ordering of the series, some sensitivity analysis should be undertaken. This can be achieved by clicking on the 'Impulse Definition' tab when the window that creates the impulses is open. A window entitled 'Ordering for Cholesky' should be apparent, and it would be possible to reverse the order of variables or to select any other order desired. For the variance decompositions, the 'Ordering for Cholesky' box is observed in the window for creating the decompositions without having to select another tab.

Screenshot 6.6
Variance decomposition graphs



Key concepts

The key terms to be able to define and explain from this chapter are

- endogenous variable
- exogenous variable
- simultaneous equations bias
- identified
- order condition
- rank condition
- Hausman test
- reduced form
- structural form
- instrumental variables
- indirect least squares
- two-stage least squares
- vector autoregression
- Granger causality
- impulse response
- variance decomposition

Review questions

1. Consider the following simultaneous equations system

$$y_{1t} = \alpha_0 + \alpha_1 y_{2t} + \alpha_2 y_{3t} + \alpha_3 X_{1t} + \alpha_4 X_{2t} + u_{1t} \tag{6.94}$$
$$y_{2t} = \beta_0 + \beta_1 y_{3t} + \beta_2 X_{1t} + \beta_3 X_{3t} + u_{2t} \tag{6.95}$$
$$y_{3t} = \gamma_0 + \gamma_1 y_{1t} + \gamma_2 X_{2t} + \gamma_3 X_{3t} + u_{3t} \tag{6.96}$$

(a) Derive the reduced form equations corresponding to (6.94)–(6.96).
(b) What do you understand by the term 'identification'? Describe a rule for determining whether a system of equations is identified. Apply this rule to (6.94–6.96). Does this rule guarantee that estimates of the structural parameters can be obtained?
(c) Which would you consider the more serious misspecification: treating exogenous variables as endogenous, or treating endogenous variables as exogenous? Explain your answer.
(d) Describe a method of obtaining the structural form coefficients corresponding to an overidentified system.
(e) Using EViews, estimate a VAR model for the interest rate series used in the principal components example of chapter 3. Use a method for selecting the lag length in the VAR optimally. Determine whether certain maturities lead or lag others, by conducting Granger causality tests and plotting impulse responses and variance decompositions. Is there any evidence that new information is reflected more quickly in some maturities than others?

2. Consider the following system of two equations

$$y_{1t} = \alpha_0 + \alpha_1 y_{2t} + \alpha_2 X_{1t} + \alpha_3 X_{2t} + u_{1t} \tag{6.97}$$
$$y_{2t} = \beta_0 + \beta_1 y_{1t} + \beta_2 X_{1t} + u_{2t} \tag{6.98}$$

(a) Explain, with reference to these equations, the undesirable consequences that would arise if (6.97) and (6.98) were estimated separately using OLS.
(b) What would be the effect upon your answer to (a) if the variable y_{1t} had not appeared in (6.98)?
(c) State the order condition for determining whether an equation which is part of a system is identified. Use this condition to determine whether (6.97) or (6.98) or both or neither are identified.
(d) Explain whether indirect least squares (ILS) or two-stage least squares (2SLS) could be used to obtain the parameters of (6.97) and (6.98). Describe how each of these two procedures (ILS and 2SLS) are used to calculate the parameters of an equation. Compare and evaluate the usefulness of ILS, 2SLS and IV.
(e) Explain briefly the Hausman procedure for testing for exogeneity.

3. Explain, using an example if you consider it appropriate, what you understand by the equivalent terms 'recursive equations' and 'triangular system'. Can a triangular system be validly estimated using OLS? Explain your answer.

4. Consider the following vector autoregressive model

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i y_{t-i} + u_t \tag{6.99}$$

where y_t is a $p \times 1$ vector of variables determined by k lags of all p variables in the system, u_t is a $p \times 1$ vector of error terms, β_0 is a $p \times 1$ vector of constant term coefficients and β_i are $p \times p$ matrices of coefficients on the ith lag of y.
   (a) If $p = 2$, and $k = 3$, write out all the equations of the VAR in full, carefully defining any new notation you use that is not given in the question.
   (b) Why have VARs become popular for application in economics and finance, relative to structural models derived from some underlying theory?
   (c) Discuss any weaknesses you perceive in the VAR approach to econometric modelling.
   (d) Two researchers, using the same set of data but working independently, arrive at different lag lengths for the VAR equation (6.99). Describe and evaluate two methods for determining which of the lag lengths is more appropriate.
5. Define carefully the following terms
   - Simultaneous equations system
   - Exogenous variables
   - Endogenous variables
   - Structural form model
   - Reduced form model