CHAPTER 4

MODELLING

4.1 SIMULTANEOUS EQUATIONS MODELS

4.1.1 A simple three equation demand and supply model

We often wish to model a system of equations i.e. supply and demand models, macroeconomic models, etc. The following three-equation supply and demand for beef model (see Pindyck and Rubinfeld (1991)) is used to explain terminology and illustrate identification issues. The beef demand equation (in deviation from the mean form) is given by

$$q_t^d = \beta_2 p_t + \beta_3 y_t + \beta_4 w_t + u_t^d \tag{4.1}$$

where q^d is the quantity of beef demanded, p is the price beef, y is disposable income and w is wealth or some other variable that affects the quantity demanded. The supply of beef equation (in deviation from the mean form) is given by

$$q_t^s = \alpha_2 p_t + \alpha_3 p_{t-1} + \alpha_4 g_t + u_t^s \tag{4.2}$$

where q^s is the quantity of beef supplied and g is the price of feed. Finally equilibrium in the market is given by

$$q_t^d = q_t^s = q_t \tag{4.3}$$

In the market place p and q will be determined. These are often called *endogenous variables* as they are determined within the system of equations that are normalised with respect to q. On the other hand, the model contains three variables whose values are not determined directly within the system. Income, the price of feed and lagged price are *predetermined variables* and help cause movement of the endogenous variables within the system of equations. The disposable income and price of feed variables are determined completely outside the system and are called *exogenous variables*.

The lagged beef price variable is determined inside the system and is called a *lagged endogenous variable*. Note y_t, w_t and u_t^d shift the demand curve while p_{t-1}, g_t and u_t^s shift the supply curve. Since this model comes from underlying economic theory it is called a *structural model*. A structural model contains endogenous as well as predetermined variables. Applying *OLS* estimation to (4.1)-(4.3) will generate biased and inconsistent estimators (see Section 3.8).

We can solve the model for each of the endogenous variables as a function solely of the predetermined variables of the system. This is called the *reduced form solution* and can be consistently estimated by *OLS*. The *reduced form solution* is given as

$$p_t = \frac{\beta_3}{\alpha_2 - \beta_2} y_t + \frac{\beta_4}{\alpha_2 - \beta_2} w_t - \frac{\alpha_3}{\alpha_2 - \beta_2} p_{t-1} - \frac{\alpha_4}{\alpha_2 - \beta_2} g_t + \frac{u_t^d - u_t^s}{\alpha_2 - \beta_2} \tag{4.4}$$

$$q_t = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} y_t + \frac{\alpha_2 \beta_4}{\alpha_2 - \beta_2} w_t - \frac{\beta_2 \alpha_3}{\alpha_2 - \beta_2} p_{t-1} - \frac{\beta_2 \alpha_4}{\alpha_2 - \beta_2} g_t + \frac{\alpha_2 u_t^d - \beta_2 u_t^s}{\alpha_2 - \beta_2} \tag{4.5}$$

These can be simplified to

$$p_t = \pi_{11} y_t + \pi_{12} w_t + \pi_{13} p_{t-1} + \pi_{14} g_t + e_t^p \tag{4.6}$$

$$q_t = \pi_{21} y_t + \pi_{22} w_t + \pi_{23} p_{t-1} + \pi_{24} g_t + e_t^q \tag{4.7}$$

where the e^i are reduced-form error terms. The two equations can consistently estimated by *OLS*. Note that the reduced form parameters and error terms are functions of the true structural parameters and error terms. The problem is to figure out the interesting structural parameters (i.e. the α_i and β_i) from the reduced form parameters (i.e. the π_i). This is called the identification problem.

4.1.2 Reduced form estimation and the identification problem

In the following three cases we will consider the model described in the last section. An equation said to be

1. *Unidentified*: if there is no way of estimating all the structural parameters from the reduced form.
2. *Exactly identified*: if there is a unique way of estimating all the structural parameters from the reduced form.
3. *Over identified*: if there is more than one way of estimating all the structural parameters from the reduced form.

The latter two cases are sometimes called identified. Note it is important to realise that within a given structural model, some equations may be identified and others may not. Within a single equation some parameters may be identified and others may not.

Case 1: $\alpha_3 = \beta_3 = \alpha_4 = \beta_4 = 0$

At each time period there is one equilibrium value for *p* and *q* that the econometrician has to estimate the model. There is no way to ascertain the true supply and demand curves; both are unidentified. Since in Case 1 you are only using data on price and quantity you will not knwo whether you have estimated a demand or supply equation. There is no way to obtain structural parameters from the reduced form coefficients. In this case the reduced form is

$$p_t = \frac{u_t^d - u_t^s}{\alpha_2 - \beta_2} \tag{4.8}$$

$$q_t = \frac{\alpha_2 u_t^d - \beta_2 u_t^s}{\alpha_2 - \beta_2} \tag{4.9}$$

or

$$p_t = e_t^p \tag{4.10}$$

$$q_t = e_t^q \tag{4.11}$$

Any pair of supply and demand curves intersecting at the equilibrium point could be the "true" curves. There are an infinite number of structural models which are consistent with the reduced form. Note if prediction is the objective one can use the reduced form, as identification is not required. Identification of equations in a model system requires further information.

Case 2: $\alpha_3 = \alpha_4 = \beta_4 = 0$

In this case the reduced form is

$$p_t = \frac{\beta_3}{\alpha_2 - \beta_2} y_t + \frac{u_t^d - u_t^s}{\alpha_2 - \beta_2} \tag{4.12}$$

$$q_t = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} y_t + \frac{\alpha_2 u_t^d - \beta_2 u_t^s}{\alpha_2 - \beta_2} \tag{4.13}$$

or

$$p_t = \pi_{11} y_t + e_t^p \tag{4.14}$$

$$q_t = \pi_{21} y_t + e_t^q \tag{4.15}$$

Suppose we estimate the reduced form of Case 2 by performing *OLS* on (4.14) and (4.15). Notice that $\hat{\alpha}_2 = \hat{\pi}_{21} / \hat{\pi}_{11}$. This procedure is called *indirect least squares* (*ILS*) and can be used to obtain consistent parameter estimates. The estimators may be biased because they are a ratio of two unbiased reduced form estimators. There is no way to estimate the demand equation parameters (i.e. the β_i) from the reduced form. If income varies over time then the demand curve shifts. The equilibrium values trace out the supply curve and the slope coefficient is identified. The slope of the demand curve is not identified.

Case 3 : $\alpha_3 = \beta_4 = 0$

The reduced form is given by

$$p_t = \frac{\beta_3}{\alpha_2 - \beta_2} y_t - \frac{\alpha_4}{\alpha_2 - \beta_2} g_t + \frac{u_t^d - u_t^s}{\alpha_2 - \beta_2} \tag{4.16}$$

$$q_t = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} y_t - \frac{\beta_2 \alpha_4}{\alpha_2 - \beta_2} g_t + \frac{\alpha_2 u_t^d - \beta_2 u_t^s}{\alpha_2 - \beta_2} \tag{4.17}$$

or

$$p_t = \pi_{11} y_t + \pi_{14} g_t + e_t^p \tag{4.18}$$

$$q_t = \pi_{21} y_t + \pi_{24} g_t + e_t^q \tag{4.19}$$

Notice that $\hat{\alpha}_2 = \hat{\pi}_{21} / \hat{\pi}_{11}$ and $\hat{\beta}_2 = \hat{\pi}_{24} / \hat{\pi}_{14}$. It is then easy to work out $\hat{\beta}_3$ from $\hat{\pi}_{21}$ and $\hat{\alpha}_4$ from $\hat{\pi}_{24}$. In this case both supply and demand are identified. If *y* and *g* vary over time (and they are not perfectly correlated), both curves will shift. The shifting of the supply curve via shifts in *g* will help identify the demand curve. The shifting of the demand curve via shifts in *y* will help identify the supply curve.

Case 4 : $\alpha_3 = \alpha_4 = 0$

The reduced form model is given as

$$p_t = \frac{\beta_3}{\alpha_2 - \beta_2} y_t + \frac{\beta_4}{\alpha_2 - \beta_2} w_t + \frac{u_t^d - u_t^s}{\alpha_2 - \beta_2} \tag{4.20}$$

$$q_t = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2} y_t + \frac{\alpha_2 \beta_4}{\alpha_2 - \beta_2} w_t + \frac{\alpha_2 u_t^d - \beta_2 u_t^s}{\alpha_2 - \beta_2} \tag{4.21}$$

or

$$p_t = \pi_{11} y_t + \pi_{12} w_t + e_t^p \tag{4.22}$$

$$q_t = \pi_{21} y_t + \pi_{22} w_t + e_t^q \tag{4.23}$$

In this case the supply equation is over identified as there are at least two ways of obtaining structural parameters from the reduced form. The demand equation is not identified. Now suppose we estimate (4.22) and (4.23) by *OLS* to get consistent estimates. There are two ways of calculating $\hat{\alpha}_2 = \hat{\pi}_{21} / \hat{\pi}_{11} = \hat{\pi}_{22} / \hat{\pi}_{12}$. In general both these estimators will yield consistent estimates but they may not be identical. Either one of these choices involve the loss of information about the model being estimated. In this case a researcher should estimate the equation by 2*SLS* (see Chapter 2) as this estimation procedure avoids information loss.

Identification conditions

Prior information about the excluded variable in the supply equation allows the supply curve to be identified. This leads us to a *necessary but not sufficient* condition for identification of an equation in a system of equations; a counting rule known as the *order condition*. Let

g = the number of endogenous variables in the model,
e.g. g=2 (*p* and *q*).

k = the number of excluded exogenous or endogenous variables in the equation of interest,
e.g. for supply k = 1 and demand k = 0 in Case 2.

Then if

1. k = g-1 the equation of interest is exactly identified
2. k > g-1 the equation of interest is over identified
3. k < g-1 the equation of interest is under identified

Applying this rule to Case 1 implies that no equation is identified. While in Case 2 supply is identified. An alternative way of looking at the identification problem is to see whether the equation under consideration can be obtained as a linear combination of the other equations. Consider, for example, Case 2 above. Take a weighted average (let the weights be equal to δ and (1-δ)) of the demand and supply equations

$$\begin{aligned} q_t &= \delta\left(\beta_2 p_t + \beta_3 y_t + u_t^d\right) + (1-\delta)\left(\alpha_2 p_t + u_t^s\right) \\ &= a_1 p_t + a_2 y_t + u_t \end{aligned} \tag{4.24}$$

where $a_1 = \delta\beta_2 + (1-\delta)\alpha_2$, $a_2 = \delta\beta_3$. Thus when we estimate the parameters of the demand function, we do not know whether we are getting estimates of the parameters of the demand function or in some weighted average of the demand and supply functions. Thus the parameters in the demand function are not identified. The supply function is identified because the only way (4.24) will look the supply function is if δ=0. Thus when we estimate the supply function we are sure that the estimates are those of the supply function.

In larger models we use a more systematic way of checking this condition. We use the *rank condition*. It is a *necessary and sufficient condition for identification* in a simultaneous equations model. Consider a model with 3 endogenous variables, y_{1t}, y_{2t}, and y_{3t}, and three exogenous variables x_{1t}, x_{2t}, and x_{3t}. In the following table we will mark with a • if a variable occurs in an equation, and 0 if not.

Table 4.1
The Rank Condition for Identification

Equation	y_{1t}	y_{2t}	y_{3t}	x_{1t}	x_{2t}	x_{3t}
1	•	0	•	•	0	•
2	•	0	0	•	0	•
3	0	•	•	•	•	0

The rule for identification is as follows

1. Delete a particular row. In order to check the rank of equation 1. Delete row 1.
2. Pick up the columns corresponding to the elements that have zeros in that row, i.e. the missing variables, y_{2t} and x_{2t} are missing from equation 1.
3. If from this array of columns we can find (g-1) rows and columns that are not all zeros, where g is the number of endogenous variables, then the equation is identified. Otherwise it is not. In this example g is three. The number of missing variables from the first row is 2 (y_2 and x_2). Thus the order condition gives equation 1 to be exactly identified. The rank condition does not!! The array of columns corresponding to the missing variables in equation 1 is

0	0
•	•

and we cannot find 2 rows and columns that are not all zeros. In matrix language the rank of this matrix is less than (g-1). In this case the estimates we obtain for the parameters in equation 1 are actually estimates of some linear combinations of the parameters in all the equations and thus have no special economic interpretation.

4.2 SIMULATION MODELS

In this section we discuss the construction, evaluation, and analysis of large scale agri-food simultaneous-equation simulation models. We will also discuss the dynamic properties of these models. Finally we will illustrate their use in policy analysis and forecasting. We will confine ourselves to models where the individual equations have been estimated using econometric methods (obviously in some cases one could substitute parameters, say price elasticities, estimated from another study).

Combining several estimated equations into a simultaneous-equation simulation model may produce poor results even though each equation fits well (the converse is also true). This problem can happen because of the dynamic structure of the system. A simulation is essentially a mathematical solution of a simultaneous set of difference equations. Simulations of a model might be performed for a variety of reasons, including model testing and evaluation, historical policy analysis, and forecasting. Usually the time horizon over which the simulation is formed will depend on the objective of the simulation. Consider Table 4.2.

Table 4.2
Simulation Time Horizons

Backcasting	Historical Simulation	*Forecasting*	
		ex post Forecast	*ex ante* Forecast
	T_1	T_2	T_3

T_1 and T_2 represent the time bounds over which the individual equations are estimated, the *estimation period.* T_3 represents time today. The first mode of simulation, called *historical simulation*, begins in year T_1 and runs forward until year T_2. Historical values in year T_1 are supplied as initial conditions for the endogenous variables and the values after this year are determined in the dynamic simulation. Historical values beginning in year T_1 and ending in year T_2 are used for the exogenous variables. At time T_2 the simulated endogenous variables can be compared with the actual values using percentage root mean squared error (*%rmse*) or Theil's *U* statistics given below. Note these are statistics used for model evaluation, they are not tests.

The *%rmse* is calculated as

$$\%rmse = 100 * \left(\frac{1}{n} \sum_{t=1}^{n} \left(\frac{Y_t^s - Y_t^a}{Y_t^a} \right)^2 \right)^{0.5} \tag{4.25}$$

where Y^s and Y^a are the simulated and actual values of Y_t respectively. One could also calculate mean error, mean percentage error, mean absolute error and mean absolute percentage error (see Pindyck and Rubinfeld (1991) for example). The Theil's *U* inequality coefficient is calculated as

$$U = \frac{\left(\frac{1}{n} \sum_{t=1}^{n} \left(Y_t^s - Y_t^a \right)^2 \right)^{0.5}}{\left(\frac{1}{n} \sum_{t=1}^{n} \left(Y_t^s \right)^2 \right)^{0.5} + \left(\frac{1}{n} \sum_{t=1}^{n} \left(Y_t^a \right)^2 \right)^{0.5}} \tag{4.26}$$

The Theil's *U* statistic falls between 0 and 1. When *U*=0 the numerator is zero and the historical or *ex post* fit is perfect. When *U*=1 the predictive performance of the model is as bad as possible. The mean squared error (numerator of the Theil's *U* statistic) can be decomposed into the three following proportions

$$U^m = \frac{\left(\bar{Y}^s - \bar{Y}^a\right)^2}{\left(\frac{1}{n}\sum_{t=1}^{n}\left(Y_t^s - Y_t^a\right)^2\right)^{0.5}} \tag{4.27}$$

$$U^s = \frac{\left(\sigma^s - \sigma^a\right)^2}{\left(\frac{1}{n}\sum_{t=1}^{n}\left(Y_t^s - Y_t^a\right)^2\right)^{0.5}} \tag{4.28}$$

and

$$U^c = \frac{2\left(1-\rho\right)\sigma^s\sigma^a}{\left(\frac{1}{n}\sum_{t=1}^{n}\left(Y_t^s - Y_t^a\right)^2\right)^{0.5}} \tag{4.29}$$

where the proportions U^m, U^s, U^c, are called the bias, the variance, and the covariance proportions, respectively, and they sum to unity. The $\bar{Y}^i$ and σ^i are the means and standard deviations of the simulated and actual series and ρ is their correlation coefficient.

The bias proportion is a scaled difference of the average values of the simulated and actual series. If it is greater than 0.1 it would mean that a bias was present in the model and the model needs to be revised. The variance proportion is a scaled difference of the standard deviations of the simulated and actual series. If this is large then the model is not picking up enough of the volatility in the series or producing to much. The covariance proportion measures unsystematic error which is the remaining error after deviations from average have been accounted for. A model performs well if this is close to unity (for a further discussion see Pindyck and Rubinfeld (1991)).

Another test of the models performance is to test for forecast accuracy. Since the model is estimated using data up to T_2 and data is available up to T_3 it is useful to forecast the endogenous variables in the model and compare the *ex post* forecast with the actual data percentage mean squared error or Theil's *U* statistics. At this stage one would re-estimate each equation in the model using data up to T_3.

A good simulation model should duplicate *turning points* in the data. A simulated series may have a higher *%rmse* or *U* statistic for an endogenous series and still be a preferred model if it is better at tracking turning points of that endogenous series. Other criterion of model performance is overall sensitivity to changing estimation period, small percentage changes in exogenous variables and small changes in estimated coefficients. A model is preferred if the simulation performance does not change very much.

Ex ante forecasts are made after T_3. FAPRI produce 10-year forecasts. In order to generate a forecast of an endogenous variable one must have forecasts of all the exogenous variables. Some of these may come from another model. Forecasts of future disposable income and other macroeconomic variables may be obtained from the ESRI's macro model. Different sets of assumptions may be used for government or EU policy variables. Using the model for policy analysis not only involves using different assumptions about future time series but also involves changing the values of various policy parameters such as tax rates. Any other remaining forecasts of future exogenous variables may have to be made by using time-series methods such as Box-Jenkins or vector-autoregression analysis.

4.2.1 Properties of dynamic models

In models with lagged endogenous variables, the entire previous time path of the exogenous variables and disturbances, not just their current values, determines the current value of the dependent variables. The intrinsic dynamic properties of the autoregressive model, such as stability and existence of an equilibrium, are embodied in their autoregressive parameters. Simulation models are often used to study and compare short-run and long-run responses of one variable to another variable. Two common and obviously related ways of analysing these responses is by calculating *dynamic multipliers* and plotting *impulse response functions*. We illustrate the dynamic properties of models with three examples.

Example 1

In the first example consider the following macroeconomic model

$$Y_t = C_t + I_t \tag{4.30}$$

and

$$C_t = \alpha_0 + \alpha_1 Y_t \tag{4.31}$$

Assume I_t is fixed at level I. Then substitute C_t from (4.31) into (4.30) to get

$$Y_t = \beta_0 + \beta_1 Y_t \tag{4.32}$$

where $\beta_0 = \alpha_0 + I$. This is a nonhomogenous first order difference equation. There is also a first order difference equation for C_t. The general solution to (4.32) is a function rather than a number.

NOTE: The general solution = particular solution + complementary solution

The particular solution is a long run, steady state or equilibrium solution. This can be solved algebraically or by simulating the model far into the future, using a given set of exogenous data, until the values of endogenous variables do not change (to so many decimal places). The former approach gives you the exact long-run solution. The latter approach gives the approximate long run solution and the number of periods it takes. The complementary (or homogenous or transient) solution is time dependent and will usually determine whether the equation (and therefore the time series) is stable or not.

The particular solution

If $\beta_1 \neq 1$ the long run solution for Y is a constant so try $Y_t^p = \overline{Y}$. This gives

$$Y_t^p = \overline{Y} = \frac{\beta_0}{1-\beta_1} \tag{4.33}$$

An alternative method is to iterate backwards on (4.32). If $\beta_1 = 1$ the particular solution is still the long run solution but is now a function of time and not constant so try $Y_t^p = t\overline{Y}$. This gives

$$\begin{aligned} t\overline{Y} &= \beta_0 + \beta_1(t-1)\overline{Y} \\ \Rightarrow \overline{Y} &= \beta_0 \end{aligned} \tag{4.34}$$

This implies that the long run solution is

$$Y_t^p = t\overline{Y} = t\beta_0 = \beta_0 t \tag{4.35}$$

The complementary solution

This is a solution to the homogenous part of (4.32) which can be written as

$$Y_t - \beta_1 Y_t = 0 \tag{4.36}$$

Since the solution is dependent on time try $Y_t^c = Ab^t$ in (4.36). This implies

$$\begin{aligned} Ab^t - \beta_1 Ab^{t-1} &= 0 \\ b^2 - \beta_1 &= 0 \end{aligned} \tag{4.37}$$

The second line in (4.37) is a *characteristic equation* and possesses a *characteristic root* which is b. The complementary solution when $\beta_1 \neq 1$ is of the following form

$$Y_t^c = Ab^t = A\beta_1^t \tag{4.38}$$

The complementary solution when $\beta_1 = 1$ is any arbitrary constant as it satisfies $Y_t = Y_{t-1}$.

The complete solution

The complete solution when $\beta_1 \neq 1$ is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \frac{\beta_0}{1-\beta_1} + A\beta_1^t \end{aligned} \tag{4.39}$$

The complete solution when $\beta_1 = 1$ is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \beta_0 t + A \end{aligned} \tag{4.40}$$

In both cases the arbitrary constant A will depend on an initial condition e.g. when t=0 let Y_t=Y_0.

Stability of first order linear difference equations

The stability of the first difference equation for Y depends on the absolute size of the characteristic root.

If $|b| > 1$ we have a series that is divergent (explosive).
If $|b| < 1$ we have a series that is convergent (damped).
If $|b| = 1$ we have a series that is a horizontal line (a constant trend if $\beta_0 \neq 0$).
If b is positive the time path is smooth (nonoscillatory).
If b is negative the time path is oscillatory.

Example 2

In the second example we will consider an extension to the macroeconomic model

$$Y_t = C_t + I_t \tag{4.41}$$

and

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_{t-1} \tag{4.42}$$

Assume I_t is fixed at level I. Then substitute C_t from (4.42) into (4.41) to get

$$Y_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} \tag{4.43}$$

where $\beta_0 = \alpha_0 + I$. This is a nonhomogenous second order difference equation. There is also a second order difference equation for C_t. The general solution to (4.43) is a function rather than a number. The particular solution is a long run or steady state solution. The complementary (or homogenous) solution is time dependent and will usually determine whether the equation (and therefore the time series) is stable or not.

The particular solution

If $\beta_1 + \beta_2 \neq 1$ the long run solution is constant so try $Y_t^p = \overline{Y}$. This gives

$$Y_t^p = \overline{Y} = \frac{\beta_0}{1-\beta_1-\beta_2} \tag{4.44}$$

If $\beta_1 + \beta_2 = 1$ the long run solution depends on time so try $Y_t^p = t\overline{Y}$. This gives

$$t\overline{Y} = \beta_0 + \beta_1(t-1)\overline{Y} + \beta_2(t-2)\overline{Y} \tag{4.45}$$

Now solve for $\overline{Y}$

$$\overline{Y} = \frac{\beta_0}{t(1-\beta_1-\beta_2)+\beta_1+2\beta_2} = \frac{\beta_0}{2-\beta_1} \tag{4.46}$$

which implies

$$Y_t^p = t\overline{Y} = t\frac{\beta_0}{2-\beta_1} \tag{4.47}$$

because $\beta_2 = 1 - \beta_1$. If $\beta_1 = 2$ try $Y_t^p = t^2\overline{Y}$ (see Chiang (1984) for example).

The complementary solution

This is a solution to the homogenous part of (4.43) which can be written as

$$Y_t - \beta_1 Y_t - \beta_2 Y_{t-1} = 0 \tag{4.48}$$

Since the solution is dependent on time try $Y_t^c = Ab^t$ in (4.48). This implies

$$\begin{aligned} Ab^t - \beta_1 Ab^{t-1} - \beta_2 Ab^{t-2} &= 0 \\ b^2 - \beta_1 b - \beta_2 &= 0 \end{aligned} \tag{4.49}$$

This is a characteristic equation and possesses the following characteristic roots

$$b_1, b_2 = \frac{\beta_1 \pm \sqrt{\beta_1^2 + 4\beta_2}}{2} \tag{4.50}$$

The solution gives are three possible cases, all of which depend on the square root expression.

Case 1 : Distinct real roots

$$\beta_1^2 + 4\beta_2 > 0 \tag{4.51}$$

In this case the complementary solution is of the following form

$$Y_t^c = A_1 b_1^t + A_2 b_2^t \tag{4.52}$$

This solution will solve (4.48) for any two arbitrary constants. When $\beta_1 + \beta_2 \neq 1$ the complete solution is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \frac{\beta_0}{1 - \beta_1 - \beta_2} + A_1 b_1^t + A_2 b_2^t \end{aligned} \tag{4.53}$$

When $\beta_1 + \beta_2 = 1$ the complete solution is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \frac{\beta_0}{2 - \beta_1} t + A_1 b_1^t + A_2 b_2^t \end{aligned} \tag{4.54}$$

Stability of second order linear difference equations : Case 1

The stability of the first difference equation for Y depends on the absolute size of the characteristic roots.

If $|b_1| > 1$ and $|b_2| > 1$ we have a series that is explosive.

If $|b_1| < 1$ and $|b_2| < 1$ we have a series that is convergent.

If $|b_1| > 1$ and $|b_2| < 1$ we have a series that is explosive.

If $|b_1| < 1$ and $|b_2| > 1$ we have a series that is explosive.

i.e. The *dominant root* sets the tone of the time path of the series. If both roots are positive and less than one then there is smooth convergence. If they have different signs or both are negative then there may be oscillatory convergence. Stability imposes restrictions on the coefficients (see Chiang (1984)).

Case 2 : Repeated roots

$$\beta_1^2 + 4\beta_2 = 0 \tag{4.55}$$

Thus $b_1 = b_2 = b$. In this case the complementary solution is of the following form

$$Y_t^c = A_1 b_1^t + A_2 b_2^t = (A_1 + A_2) b^t = A_3 b^t \tag{4.56}$$

Therefore we need to supply a missing component. Thus

$$Y_t^c = A_3 b^t + A_4 b^t t \tag{4.57}$$

This solution will solve (4.48) for any two arbitrary constants. When $\beta_1 + \beta_2 \neq 1$ the complete solution is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \frac{\beta_0}{1 - \beta_1 - \beta_2} + A_3 b^t + A_4 b^t t \end{aligned} \tag{4.58}$$

When $\beta_1 + \beta_2 = 1$ the complete solution is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \frac{\beta_0}{2 - \beta_1} t + A_3 b^t + A_4 b^t t \end{aligned} \tag{4.59}$$

Stability of second order linear difference equations : Case 2

If $|b| > 1$ we have a series that is explosive and the multiplicative t terms will intensify the explosiveness as t increases.

If $|b| < 1$ we have a series that diminishes as t increases and but the other terms explode.

If the root is positive then the series will display smooth convergence (divergence). If the root is negative then the series will display oscillatory convergence (divergence).

Case 3 : Complex roots

$$\beta_1^2 + 4\beta_2 < 0 \tag{4.60}$$

The characteristic roots are a *complex conjugate*. Specifically they will have the form

$$\begin{aligned} &b_1, b_2 = h \pm vi \\ &h = \frac{\beta_1}{2} \quad v = \frac{\sqrt{-\beta_1^2 - 4\beta_2}}{2} \end{aligned} \tag{4.61}$$

In this case the complementary solution is of the following form

$$Y_t^c = A_5 b_1^t + A_6 b_2^t = A_5 (h + vi)^t + A_6 (h - vi)^t \tag{4.62}$$

This is not easily interpreted. However, due to De Moivre's theorem (4.61) can be transformed into trigonometric terms, which we can interpret. We can write

$$\begin{aligned} (h \pm vi)^t &= R^t \left(\cos(\theta t) \pm i \sin(\theta t)\right) \\ R &= \sqrt{h^2 + v^2} > 0 \\ \cos(\theta) = \frac{h}{r} &\quad \sin(\theta) = \frac{v}{r} \end{aligned} \tag{4.63}$$

θ is the radian measure of the angle in the interval [0,2π).

Therefore the complementary solution is

$$\begin{aligned} Y_t^c &= A_5 R^t \left(\cos(\theta t) + i \sin(\theta t)\right) + A_6 R^t \left(\cos(\theta t) - i \sin(\theta t)\right) \\ &= R^t \left[\left(A_5 + A_6\right) \cos(\theta t) + \left(A_5 - A_6\right) i \sin(\theta t) \right] \\ &= R^t \left[A_7 \cos(\theta t) + A_8 \sin(\theta t) \right] \end{aligned} \tag{4.64}$$

In the complex roots case when $\beta_1 + \beta_2 \neq 1$ the complete solution is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \frac{\beta_0}{1 - \beta_1 - \beta_2} + R^t \left[A_7 \cos(\theta t) + A_8 \sin(\theta t) \right] \end{aligned} \tag{4.65}$$

In the complex roots case when $\beta_1 + \beta_2 = 1$ the complete solution is of the following form

$$\begin{aligned} Y_t &= Y_t^p + Y_t^c \\ &= \frac{\beta_0}{2 - \beta_1} t + R^t \left[A_7 \cos(\theta t) + A_8 \sin(\theta t) \right] \end{aligned} \tag{4.66}$$

Stability of 2nd order linear difference equations : Case 3

If $|R| > 1$ the series diverges.

If $|R| < 1$ the series converges in (4.65).

The time path will display a stepped fluctuation around the long run solution due to the cos and sin terms.

It is evident that the more lags in endogenous variables appearing in the various equations leads to a more autoregressive reduced form models. Although there is greater complexity the solution methods are similar. The main criterion for stability is that the characteristic roots of the characteristic equation be less than unity in absolute terms.

Example 3

In the third example we assume a general structural model with many equations. This model can be written in compactly in matrix form as

$$\boldsymbol{Y}_t'\boldsymbol{\Gamma} + \boldsymbol{X}_t'\boldsymbol{B} + \boldsymbol{Y}_{t-1}'\boldsymbol{\Phi} = \boldsymbol{u}_t' \tag{4.67}$$

where $\boldsymbol{Y}$ is a vector of all the endogenous variables in the model and $\boldsymbol{X}$ is a vector of all the exogenous variables (including predetermined variables), $\boldsymbol{\Gamma}$, $\mathbf{B}$, and $\boldsymbol{\Phi}$ are matrices of coefficients. In practice there will be many zero elements in these matrices. If there is more than one lag in the some of the endogenous variables the variables can be redefined so as to fit into (4.67). The reduced form is

$$\begin{aligned} \boldsymbol{Y}_t' &= -\boldsymbol{X}_t'\boldsymbol{B}\boldsymbol{\Gamma}^{-1} - \boldsymbol{Y}_{t-1}'\boldsymbol{\Phi}\boldsymbol{\Gamma}^{-1} + \boldsymbol{u}_t'\boldsymbol{\Gamma}^{-1} \\ &= \boldsymbol{X}_t'\boldsymbol{\Omega} + \boldsymbol{Y}_{t-1}'\boldsymbol{\Delta} + \boldsymbol{v}_t' \end{aligned} \tag{4.68}$$

From the reduced form the short-run (current) effects on the endogenous variables of a change in exogenous variables is given by $\boldsymbol{\Omega}$ which is a matrix of *impact multipliers*. This is just

$$\frac{\partial y_{t,i}}{\partial x_{t,j}} = \boldsymbol{\Omega}_{ij} \tag{4.69}$$

Iterating backwards on (4.68) gives

$$\boldsymbol{Y}_t' = \sum_{s=0}^{n-1} \boldsymbol{X}_{t-s}'\boldsymbol{\Omega}\boldsymbol{\Delta}^s + \boldsymbol{v}_t'\boldsymbol{\Delta}^s + \boldsymbol{Y}_0'\boldsymbol{\Delta}^s \tag{4.70}$$

This shows how the initial conditions $\boldsymbol{Y}_0$ and the subsequent time path of the exogenous variables and disturbances completely determine the current values of the endogenous variables. In the terminology used above this can be thought of as the complementary solution. The coefficient matrices in the bracketed sum are the *dynamic multipliers*,

$$\frac{\partial y_{t,i}}{\partial x_{t-s,j}} = \left(\boldsymbol{\Omega}\boldsymbol{\Delta}^s\right)_{ij} \tag{4.71}$$

A plot of the dynamic multipliers against lag length is called the *impulse response function*. The *cumulated multipliers* are obtained by adding the matrices of the dynamic multipliers. In order to obtain the long run, equilibrium or particular solution let *s* go to infinity in (4.68). This gives the *final form* of the model

$$Y_t' = \sum_{s=0}^{\infty} X_{t-s}' \Omega \Delta^s + v_t' \Delta^s \tag{4.72}$$

Assuming that $\lim_{t\to\infty} \Delta^t=0$, then the *long-run or equilibrium multipliers* are given by

$$\Omega (I - \Delta)^{-1} \tag{4.73}$$

and the cumulated multipliers are given as

$$\Omega (I - \Delta)^{-1} (I - \Delta^s) \tag{4.74}$$

Stability of a system of equations

In examples one and two the complementary solution gave the dynamic properties to the time series. Stability depended on the size of the dominant characteristic root. There is an analogous property in a system of simultaneous equations. For stability in a system of equations $\lim_{t\to\infty} \Delta^t=0$ is required. Thus the dominant characteristic root of **Δ** must be less than unity in absolute value. Most econometric software packages can find the characteristic roots of a matrix. See Greene (1997), Johnston (1984) and Johnston and DiNardo (1997).

4.3 AN EXAMPLE

The example we take is a small nine equation dynamic model of the U.S. turkey market (see Brown (1994)). This model is small enough to illustrate some of the techniques described in this chapter. The following six equations were estimated using 2*SLS* over the 1970-91 period. Our results for 5 out of 6 equations are identical to those reported in Brown (1994). We used the following set of instrumental variables based on a SAS programme received from Scott Brown; *CONSTANT*, *TKEXPT*, *TREND*, *CRPFRM*, *SMP44D*, *ZCEW*, *POPTOTW*, *PCIUW*, *PPIW*, *CKRETP*, *D8234*, and *ZWRHP20W* (see below for a description of the variables).

The first estimated equation is for the number of turkeys slaughtered

$$TKHATCH_t = \underset{(-2.29)}{-54790.6} + \underset{(2.85)}{167902.3} \frac{(TKYWHP_t + TKYWHP_{t-1})/2}{TKFEED_t} + \underset{(7.87)}{0.88} TKHATCH_{t-1} + \underset{(2.12)}{1980.8} TREND + \underset{(1.76)}{30811.0} DUM89 \tag{4.75}$$
$$\bar{R}^2 = 0.979,$$

where *TKHATCH* is the number of turkey poults placed for slaughter, *TKYWHP* is the wholesale price of 14-22lb. young Tom turkeys, and *TKFEED* is turkey grower feed prices. The *t-test* statistics are reported in parenthesis. The coefficient on current and lagged wholesale turkey prices is restricted to be the same. This is testable. Some of the coefficients are large because the number of turkeys slaughtered is measured in thousands. For presentation purposes it would be better to scale the dependent variable to units measured in, say, millions. Note the Word 97 equation editor is not the best for presenting estimated equations in the style discussed in Chapter 1.

The second estimated equation is for turkey production

$$TKPROD_t = -\underset{(-2.50)}{500.5} + \underset{(0.41)}{189.6}\frac{TKYWHP_t}{TKFEED_t} + \underset{(42.15)}{0.016}\,TKHATCH_t - \underset{(-3.71)}{312.4}D7980 \qquad (4.76)$$
$$\bar{R}^2 = 0.994,$$

where *TKPROD* is the number of turkey produced. Note the ratio of wholesale turkey prices to grower feed prices is insignificantly different form zero yet retained in the model presumably because it improves the simulation results. The third estimated equation is for turkey grower feed prices

$$TKHATCH_t = -\underset{(-2.29)}{54790.6} + \underset{(2.85)}{167902.3}\frac{(TKYWHP_t + TKYWHP_{t-1})/2}{TKFEED_t} + \underset{(7.87)}{0.88}TKHATCH_{t-1} + \underset{(2.12)}{1980.8}TREND + \underset{(1.76)}{30811.0}DUM89 \qquad (4.77)$$
$$\bar{R}^2 = 0.979,$$

where *SMP44D* is soyabean meal price and *CRPFRM* is the corn meal price. There is probably no need to estimate this equation by 2*SLS* as all the right hand side variables are exogenous. A formal Durbin-Wu-Hausman test would indicate the most appropriate method of estimation to use.

The fourth estimated equation is for the wholesale price of turkeys

$$\ln\left(\frac{TKYWHP_t}{PPIW_t}\right) = \underset{(7.78)}{0.92}\ln\left(\frac{TKRETP_t}{PPIW_t}\right) - \underset{(-1.96)}{0.11}\ln(TKPROD_t) - \underset{(-1.04)}{0.19}\ln\left(\frac{ZWRHP20W_t}{PPIW_t}\right) + \underset{(2.15)}{0.29}DUM78 - \underset{(1.96)}{0.22}DUM73 \qquad (4.78)$$
$$\bar{R}^2 = 0.914,$$

where *PPIW* is the producer price index, *TKRETP* is the turkey retail price and *ZWRHP20W* is the wage rate in food and kindred products industry. This is the only equation that has slightly different results than Brown (1994) on page 113. We dropped the constant from the instrument set but it does not seem to matter. This is a log-linear model with a few dummy variables. This functional form is testable. Note the log real wage rate is also insignificantly different from zero yet retained in the model presumably because it improves the simulation results.

The fifth estimated equation is for turkey consumption per capita

$$\ln(TKPCCR_t) = \underset{(7.33)}{6.27} - \underset{(-5.61)}{1.37}\ln\left(\frac{TKRETP_t}{PCIUW_t}\right) + \underset{(2.88)}{0.59}\ln\left(\frac{CKRETP_t}{PCIUW_t}\right) + \underset{(1.64)}{0.39}\ln\left(\frac{ZCEW_t / POPTOTW_t}{PCIUW_t}\right) - \underset{(5.27)}{0.11}D7584 \qquad (4.79)$$
$$\bar{R}^2 = 0.975,$$

where *TKPCCR* is turkey consumption per capita, *CKRETP* is the chicken retail price, *PCIUW* is the consumer price index, *ZCEW* is the total consumption expenditures and *POPTOTW* is the total population in the U.S. This is a log-linear model with a dummy variable. Note the log real wage rate is also insignificantly different from zero yet retained in the model presumably because it improves the simulation results. As with the fourth equation this is a log-linear model with a few dummy variables which is testable. Note the log consumption expenditure per capita is also insignificantly different from zero yet retained in the model presumably because it improves the simulation results.

The sixth estimated equation is for the turkey ending stocks, *TKSTK*,

$$TKHATCH_t = \underset{(-2.29)}{-54790.6} + \underset{(2.85)}{167902.3}\frac{(TKYWHP_t + TKYWHP_{t-1})/2}{TKFEED_t} + \underset{(7.87)}{0.88}TKHATCH_{t-1} + \underset{(2.12)}{1980.8}TREND + \underset{(1.76)}{30811.0}DUM89 \quad (4.80)$$
$$\bar{R}^2 = 0.979,$$

As with other equations an insignificant variable is included in the regression equation. The 1979 dummy is not significant. There are three identities in the model. An identity for turkey supply, *TKSUPP*,

$$TKSUPP_t = TKPROD_t + TKSTK_{t-1} - TKCONDM_t \quad (4.81)$$

where *TKCONDM* are turkey condemnations. An identity for turkey civilian disappearance, *TKCDIS*,

$$TKCDIS_t = TKSUPP_t - TKEXPT_t - TKSTK_t \quad (4.82)$$

where *TKEXPT* are turkey exports. An identity for turkey consumption per capita

$$TKPCCR_t = \frac{TKCDIS_t}{POPTOTW_t} \quad (4.83)$$

It is evident from the table that with the exception of the turkey stock series the model performs quite well within sample. The poor performance of the model in replicating the turkey stock series reflects the fact the turkey stock equation has the poorest fit. In addition to producing these statistics, one would plot each simulated series with the actual historical series and examine if all the turning points were captured by the model (see examples in Brown (1994)). We first dynamically simulated the model over the 1970-91 and present the *%rmse* and *Theil's U* and related statistics in Table 4.3.

The next stage in evaluating the model would be to dynamically forecast the model over the 1992-93 and compare forecasts of the endogenous variables with the actual data. We present the *%rmse* and *Theil's U* and related statistics in Table 4.4. In terms of *%rmse* and *Theil's U* the model performed well out of sample. It would be preferable to have a few more data points to use in the *ex post* evaluation.

We must also check and determine whether the model is dynamically stable. We do so by forecasting the model from 1994 forward to the year 2100 (or greater) and noting when the endogenous variables reach long run equilibrium. In Table 4.5 we present the year in which long run equilibrium is achieved, the number of years required and the long run equilibrium value of the endogenous variables. The are a few points to note. First the model is stable as all variables reach long run equilibrium. The *TKHATCH* variable takes the longest to reach long run equilibrium as the value of the coefficient on the lagged dependent variable in that equation is 0.88 which implies long adjustment period. This effect feeds into the other variables except *TKFEED* which only depends on exogenous variables.

Econometric code

The following is an example of part of the code of a SAS programme that will produce an example of the tables and analysis discussed above (much more programming is required in TSP and RATS – see users manuals for examples).

```
PROC MODEL DATA=MYDATA;
EXOGENOUS X1 X2; PARAMS BETA01 BETA11 BETA21 BETA02 BETA12 BETA22;
Y1 = BETA01 + BETA11*Y2 + BETA12*X1 ;
Y2 = BETA02 + BETA12*X2 + BETA22*X2*X2 ;
FIT Y1 Y2 / 2SLS; INSTRUMENTS _EXOG_ ;
/* Within sample evaluation */
SOLVE Y1 Y2 / DYNAMIC THEIL OUTPREDICT ;
/* Ex post 5 year sample evaluation */
SOLVE Y1 Y2 / DYNAMIC THEIL OUTPREDICT NAHEAD=5;
/* 100 year sample stability evaluation */
SOLVE Y1 Y2 / DYNAMIC OUTPREDICT NAHEAD=100;
RUN;
```

This is an example of a two-equation model with two exogenous variables. The FIT command estimates the models by 2*SLS* using all the exogenous variables as instruments (INSTRUMENTS includes a constant by default).

Table 4.3
Within Sample Performance Statistics of the U.S. Turkey Model

	%rmse	U^m	U^s	U^c	*U*
TKHATCH	5.74	0.000	0.002	0.998	0.025
TKPROD	6.44	0.000	0.012	0.988	0.027
TKFEED	6.34	0.000	0.002	0.998	0.025
TKYWHP	7.53	0.002	0.000	0.998	0.039
TKRETP	4.55	0.007	0.003	0.990	0.023
TKSTK	14.37	0.000	0.041	0.959	0.062
TKSUPP	6.39	0.006	0.011	0.988	0.028
TKCDIS	6.37	0.000	0.016	0.984	0.027
TKPCCR	6.38	0.000	0.019	0.981	0.028

Table 4.4
Ex Post Sample Performance Statistics of the U.S. Turkey Model

	%rmse	U^m	U^s	U^c	*U*
TKHATCH	5.21	0.880	0.100	0.020	0.025
TKPROD	1.68	0.001	0.999	0.000	0.009
TKFEED	5.82	0.273	0.727	0.000	0.008
TKYWHP	4.23	0.928	0.072	0.000	0.021
TKRETP	3.68	0.986	0.014	0.000	0.018
TKSTK	2.29	0.131	0.869	0.000	0.011
TKSUPP	1.61	0.300	0.700	0.000	0.008
TKCDIS	1.57	0.359	0.423	0.208	0.009
TKPCCR	1.69	0.361	0.001	0.639	0.009

Table 4.5
Dynamic Properties of the U.S. Turkey Model

	YEAR IN WHICH LONG RUN EQUILIBRIUM REACHED	*YEARS TO REACH LONG RUN EQUILIBRIUM*	*VALUE*
TKHATCH	2091	97	376.961m
TKPROD	2059	65	5704m lbs.
TKFEED	1995	1	$226 per ton
TKYWHP	2047	53	$0.531 per lb.
TKRETP	2073	79	$0.847 per lb.
TKSTK	2036	42	358m lbs.
TKSUPP	2061	67	6007m lbs.
TKCDIS	2055	61	5539m lbs.
TKPCCR	2037	43	21.9lbs.