FROM: THOMAS - MODERN ECONOMETRICS

COMPUTER EXERCISE I

In this exercise, we shall again make use of Data Set 1, previously used in Chapters 7, 9 and 11, involving annual observations on the US demand for food. On this occasion, however, we will make a start at employing the general-to-specific approach to estimating an acceptable demand for food equation, taking as our initial general model Equation (12.31). A first attempt at using this approach can be made now, but Chapter 13 will need to be studied before the exercise can be satisfactorily completed.

Estimation of (12.31) for the years 1965–1989 should yield, with lower-case letters denoting natural logarithms,

$$\begin{aligned}\hat{q}_t = {} & \underset{(1.87)}{1.80} + \underset{(4.10)}{0.577}q_{t-1} + \underset{(3.09)}{0.571}x_t - \underset{(-1.82)}{0.337}x_{t-1} \\ & -\underset{(-2.15)}{0.350}p_t + \underset{(4.63)}{0.486}p_{t-1} - \underset{(-1.36)}{0.459}g_t + \underset{(0.14)}{0.039}g_{t-1}\end{aligned} \qquad (12.44)$$

$$R^2 = 0.986, \quad \sum e^2 = 0.001\,66, \quad dw = 1.95$$

$$z_{11} = 0.02, \quad z_2 = 0.28, \quad z_3 = 2.04, \quad z_4 = 0.45, \quad z_5 = 1.14$$

Figures in parentheses are, as usual, t ratios. z_{11} above is the LM statistic for first-order autocorrelation and z_2 is the RESET statistic, described earlier in this chapter, using just the square of the fitted values. z_3 is the Jarque–Bera statistic for normality in the residuals, also described earlier in this chapter, and z_4 is the LM statistic for heteroskedasticity described in Chapter 10.3. z_5 is the Chow test statistic for predictive failure, described in Section 9.5. We have used the observations in Data Set 1 for the years 1990–92 in the computing of this last statistic. Recall that ability to predict well 'out of sample' is an important attribute of a satisfactory model. $z_{11}, \ldots, z_5$, or similar diagnostic statistics, are standard output from most modern regression packages. A MICROFIT printout showing the basic regression result and most of these statistics is presented in Table 12.1.

The LM statistics z_{11} and z_4 are distributed as χ^2 with 1 degree of freedom, and so have critical values of $\chi^2_{0.05} = 3.841$. z_2, the RESET statistic, has an F distribution with [1,16] d.f., with a critical value $F_{0.05} = 4.49$. z_3, the normality statistic, is distributed as χ^2 with 2 degrees of freedom, and therefore has a critical value of $\chi^2_{0.05} = 5.991$. z_5, the Chow statistic, has an F distribution with [3,17] d.f. and a critical value of $F_{0.05} = 3.20$. None of these statistics exceed their critical values, and the Durbin–Watson statistic takes a value close to 2. The reader should also confirm, using LM statistics, that there is no sign of higher-order autocorrelation in the residuals of (12.44).

All these test suggest that specification in 12.44 is OK

As explained earlier, diagnostic statistics can be regarded as tests of mis-specification in this context, so that their values can be regarded as confirmation that we have an appropriate functional form and are justified in omitting lags of more than one period from the equation. However, the reader should also verify that the addition of second-order lags (i.e. the variables q_{t-2}, x_{t-2}, p_{t-2} and g_{t-2}) to (12.44) leads to no improvement in the explanatory power of the equation. This may be verified by estimating an equation including such second-order lags and applying the F-test statistic (9.28) for additional explanatory variables, or its equivalent (9.19). Note also that the addition of second-order lags leads to a value for z_3, the normality statistic, well in excess of its critical value of 5.991.

Earlier in this chapter we used the Hausman test for contemporaneous correlation to assess the specification of the static demand equation (7.66). The reader should apply this test to the general model (12.44). You will need 8 instruments in all for the test. Use the intercept, one-period-lagged values of q_t, x_t, p_t and g_t, and two-period-lagged values of x_t, p_t and g_t. You should find that (12.44) passes the Hausman test, confirming that it is a reasonable specification.

In Section 12.1 we also used the Zarembka version of the Box–Cox procedure to demonstrate that with this data set for the simple static demand model, a double-log specification provided a superior fit to a linear specification. The reader should verify that this superiority still holds when first-order lags are introduced into the equation as in (12.44). This will confirm the findings of the above diagnostic tests that the double-log specification of (12.44) is satisfactory.

Simplification search:

models that are nested within (12.44). The relationship between these models was illustrated in Figure 12.3. Our next step is therefore to test such models against the general model (12.44).

First, estimation of Equation (12.34), the simple partial adjustment model, yields

$$\hat{q}_t = \underset{(3.31)}{3.94} + \underset{(2.59)}{0.423}q_{t-1} + \underset{(1.72)}{0.216}x_t - \underset{(-2.00)}{0.186}p_t - \underset{(-0.27)}{0.055}g_t \tag{12.45}$$

$$R^2 = 0.957, \quad \sum e^2 = 0.005\,21, \quad dw = 1.09$$

$$z_{11} = 7.72, \quad z_2 = 4.13, \quad z_3 = 0.82, \quad z_4 = 0.36, \quad z_5 = 0.15$$

What is immediately striking about (12.45) is the sharp deterioration in the autocorrelation statistics z_{11} and dw compared with (12.44). There is now clear evidence of first-order autocorrelation in the residuals, and moreover the RESET statistic z_2 is now close to its critical value. This suggests we have made a serious specification error in omitting the lagged variables x_{t-1}, p_{t-1} and g_{t-1} from (12.44).

We can test the three restrictions placed on (12.44) to obtain (12.45), using the F-test statistic (9.19) and the residual sums of squares obtained for (12.44) and (12.45). In this case we have

$$\frac{(\text{SSR}_\text{R} - \text{SSR}_\text{U})/h}{\text{SSR}_\text{U}/(n-k)} = \frac{(0.005\,21 - 0.001\,66)/3}{0.001\,66/(25-8)} = 12.1$$

Since, with [3,17] d.f., the critical F value is $F_{0.05} = 3.20$, we see that the restrictions are strongly rejected by the data.

It is clear that the nested model (12.45) has to be rejected when compared with the general model (12.44). The simplification search in this particular direction therefore terminates. There is no point in carrying on and estimating equations of the forms (12.32) or (12.35), which are nested within (12.45). Note that we did, in fact, estimate Equation (12.32) in Chapter 7 (see Equation (7.66)) and Equation (12.35) in Chapter 11 (see Equation (11.24)). However, since (12.45) is rejected against the general model, we now see that these models must also be rejected. Thus, although when we estimated them in earlier chapters, equations such as (11.24) looked superficially attractive, with high R^2s etc., they, in fact, represent a highly inadequate description of the data set. We must revert to the general model and look for other models nested within it.

Reverting to the general model (12.44), suppose we estimate the special case (12.33), which is also nested in (12.44). This should yield, over the same sample period,

$$\hat{q}_t = \underset{(4.35)}{5.28} + \underset{(2.89)}{0.917}x_t - \underset{(-1.36)}{0.435}x_{t-1} - \underset{(-3.89)}{0.871}g_t + \underset{(1.52)}{0.316}g_{t-1} \tag{12.46}$$

$$R^2 = 0.944, \quad \sum e^2 = 0.006\,83, \quad dw = 0.89$$

$$z_{11} = 6.66, \quad z_2 = 6.97, \quad z_3 = 0.43, \quad z_4 = 3.67, \quad z_5 = 0.19$$

Again there is a deterioration in the autocorrelation and RESET statistics, suggesting specification problems with (12.46). Moreover, if we F-test the three restrictions necessary to obtain this equation, we obtain a value for the test statistic (9.19) as high as 17.7. The simplification search must end here and we must again revert to the general model 12.44.

Exercise:

1) Explain, why are models 12.45 and 12.46 nested in 12.44

2) Do you understand the tests after estimation of 12.44?

FD2-3

COMPUTER EXERCISE II

We have estimated demand equations using the annual US series in Data Set 1 on the floppy disk on several previous occasions, notably in Section 12.5. So far, however, we have only used variables in 'level' form. But we have seen that the levels of the variables Q, X, P and G exhibit definite upward trends, and are almost certainly non-stationary. This casts considerable suspicion on most of the results we have obtained previously.

In this exercise we shall again apply the general-to-specific approach to Data Set 1, but, unlike in Section 12.5, this time making use of the error correction type models just described. We begin by estimating a reparameterized version of the general model (12.44), which is merely an extension to three explanatory variables of the ECM (13.36). The implied long-run equilibrium relationship is therefore, again using lower-case letters to denote logarithms,

$$q_t = \beta_0^* + \beta_1 x_t + \beta_2 p_t + \beta_3 g_t \qquad (13.38)$$

$\Delta q_t = b_1 \Delta x_t + b_2 \Delta p_t + b_3 \Delta g_t - \lambda (q_{t-1} - \beta_0 x_{t-1} - \beta_2 p_{t-1} - \beta_3 g_{t-1})$

Estimating the ECM with no restrictions on its coefficients, for the same period 1965 through 1989 as for Equation (12.44), should yield

$$\widehat{\Delta q_t} = \underset{(1.87)}{1.80} + \underset{(3.09)}{0.571}\,\Delta x_t - \underset{(-2.15)}{0.350}\,\Delta p_t - \underset{(-1.36)}{0.459}\,\Delta g_t$$

$$-\underset{(-3.01)}{0.423}q_{t-1} + \underset{(2.43)}{0.235}x_{t-1} + \underset{(0.90)}{0.136}p_{t-1} - \underset{(-2.44)}{0.420}g_{t-1} \qquad (13.39)$$

$R^2 = 0.855, \quad \sum e^2 = 0.001\,66, \quad dw = 1.95$

$z_{11} = 0.02, \quad z_2 = 1.23, \quad z_3 = 2.04, \quad z_4 = 1.49, \quad z_5 = 1.14$

The z statistics in (13.39) are as defined in Section 12.5, where their critical values are also given. Notice first that the residual sum of squares and many of the values of the diagnostic statistics in (13.39) are identical to those obtained for the general model (12.44) of the last chapter.[5] This is because, just as Equation (13.36) above is simply a reparameterization of Equation (13.35), so Equation (13.39) is no more than a reparameterization of the general model (12.44). They are effectively the same equation. As a simple arithmetic exercise, the reader is invited to obtain (13.39) by rearranging (12.44).

Although (13.39) and (12.44) are really the same equation, notice that the R^2 for (13.39) of 0.855 is considerably less than the 0.986 obtained for (12.44). This is because, as we noted in Chapter 12, it is far easier to explain variations in a trending variable such as q_t than it is to explain variations in a variable such as Δq_t, from which the trend has, apparently, been eliminated. However, the R^2 in (13.39) probably means rather more than that for (12.44), since, now that we are working largely in first differences, problems of spurious correlation will, we hope, be absent.

We have already tested the general model (12.44) for possible mis-specification, so that, since (13.39) is no more than a rearrangement of (12.44) with similar diagnostic statistics, we can take it as our general model for this exercise.

Of the individual variables in (13.39), $\Delta x_t, \Delta p_t, q_{t-1}, x_{t-1}$ and g_{t-1} have coefficients that are significantly different from zero (with $n - k = 17$ d.f, the critical t value is $t_{0.05} = 1.74$). Δg_t, the change in the general price index, and p_{t-1}, the lagged food price, have insignificant coefficients, although the t ratio on Δg_t is -1.36.

It is the insignificance of the coefficient on p_{t-1} that has the more serious implication, since this suggests that in the long run the relative price of food has no influence on the demand for it. However, food is obviously a very basic necessity, so that, over the relative price ranges in our sample, the demand for it may well be eventually unaffected by its price. Dropping p_{t-1} from the equation, as the first stage of a simplification search, yields

$$\begin{aligned}\widehat{\Delta q_t} &= \underset{(3.05)}{2.33} + \underset{(3.58)}{0.624}\,\Delta x_t - \underset{(-4.57)}{0.464}\,\Delta p_t - \underset{(-1.17)}{0.194}\,\Delta g_t \\ &\quad -\underset{(-4.26)}{0.494} q_{t-1} + \underset{(2.77)}{0.257} x_{t-1} - \underset{(-2.57)}{0.311} g_{t-1}\end{aligned} \tag{13.40}$$

$R^2 = 0.848,\quad \sum e^2 = 0.001\,736,\quad dw = 1.91$

$z_{11} = 0.06,\quad z_2 = 1.08,\quad z_3 = 2.53,\quad z_4 = 1.48,\quad z_5 = 1.07$

There is no deterioration in the diagnostic statistics in (13.40) compared with (13.39). We do not need to F-test the simple restriction imposed to omit p_{t-1}, since this variable had an insignificant coefficient in (13.39).

Notice that the rate of change in general prices, Δg_t, still has an insignificant coefficient in (13.40). However, the three short-term elasticities in (13.40), that is, the coefficients on the rate-of-change variables, sum to 0.034, which is close to zero. This suggests that, as the next step in our simplification search, instead of omitting Δg_t, we impose the restriction that these elasticities sum to zero. This would imply that the demand for food is homogeneous of degree zero in total expenditure and prices in the short run. Imposing this restriction involves replacing the three rate-of-change variables by the rate of change in real expenditure, $\Delta x_t - \Delta g_t$, and the rate of change in relative prices,[6] $\Delta p_t - \Delta g_t$. You should find that this results in

$$\begin{aligned}\widehat{\Delta q_t} &= \underset{(3.29)}{2.38} + \underset{(5.46)}{0.657}(\Delta x_t - \Delta g_t) - \underset{(-4.68)}{0.463}(\Delta p_t - \Delta g_t) \\ &\quad -\underset{(-4.39)}{0.496} q_{t-1} + \underset{(2.83)}{0.255} x_{t-1} - \underset{(-2.62)}{0.307} g_{t-1}\end{aligned} \tag{13.41}$$

$R^2 = 0.847,\quad \sum e^2 = 0.001\,743,\quad dw = 1.93$

$z_{11} = 0.03,\quad z_2 = 1.01,\quad z_3 = 3.30,\quad z_4 = 1.42,\quad z_5 = 1.14$

Again there are no problems with the diagnostic statistics. The imposition of the restriction has left the residual sum of squares virtually unchanged. We can F-test the restriction imposed, using the test statistic (9.19), treating (13.40) as the unrestricted equation and (13.41) as the restricted equation. This yields a value of just 0.07, compared with a critical value, using [1,18] d.f. of $F_{0.05} = 4.41$. Imposing short-run homogeneity on the demand for food is clearly data-acceptable.

Examination of Equation (13.41) suggests that one further restriction, and hence one further simplification, might be made. The coefficients of x_{t-1} and g_{t-1} are of opposite sign and of roughly similar absolute value. This suggests that they be replaced by a single variable, the lagged change in real expenditure, $x_{t-1} - g_{t-1}$. Such a restriction implies that the demand for food is homogeneous not merely in the short run but also in the long run. Imposing the restriction results in

$$\widehat{\Delta q_t} = \underset{(4.42)}{2.99} + \underset{(4.87)}{0.581}(\Delta x_t - \Delta g_t) - \underset{(-4.10)}{0.374}(\Delta p_t - \Delta g_t)$$

$$-\underset{(-3.96)}{0.364}q_{t-1} + \underset{(3.00)}{0.101}(x_{t-1} - g_{t-1}) \qquad (13.42)$$

$R^2 = 0.820, \quad \sum e^2 = 0.002\,05, \quad dw = 1.81$

$z_{11} = 0.22, \quad z_2 = 0.50, \quad z_3 = 1.51, \quad z_4 = 1.25, \quad z_5 = 1.31$

LM test Reset Jarque Bera LM Chow test

There is again no obvious deterioration in the diagnostic statistics for (13.42) compared with (13.41). If we F-test the single restriction imposed on (13.41) to obtain (13.42), the value of the test statistic (9.19) is 3.36. This compares with a critical value, using [1,19] d.f., of $F_{0.05} = 4.38$, so the restriction is not rejected by the data. This restriction, however, comes a little closer to being rejected than the previous restrictions imposed.

We have now reached the end of our simplification search, since there are no obvious further restrictions that can be imposed on equation (13.42). We have therefore tested down from the general error correction model (13.39), which contained eight variables including the intercept, to the more parsimonious error correction model (13.42) which contains just five variables. We must now test (13.42) against the general model. Three restrictions have been imposed in all to obtain (13.42) (note that each linear restriction imposed reduces the number of variables by one). Testing the combined effect of imposing these restrictions gives a value for the F-test statistic of 1.33 compared with a critical value, with [3,17] d.f., of $F_{0.05} = 3.20$. Hence Equation (13.42) cannot be rejected against the general model.

Although the z_5 statistic for Equation (13.42) indicates that the model passes the second Chow test for predictive failure, we can investigate parameter stability more thoroughly by employing the recursive least squares technique described in Section 9.5. In Figure 13.6 we illustrate the time paths of the recursive least squares estimated coefficients on the explanatory variables in (13.42). Notice that after initial instability (resulting from small subsample sizes), all the estimates become stable over time. This is particularly the case for the coefficients on the lagged level variables q_{t-1} and $x_{t-1} - g_{t-1}$, which relate to the underlying long-run relationship. We appear to have uncovered a relatively stable relationship.

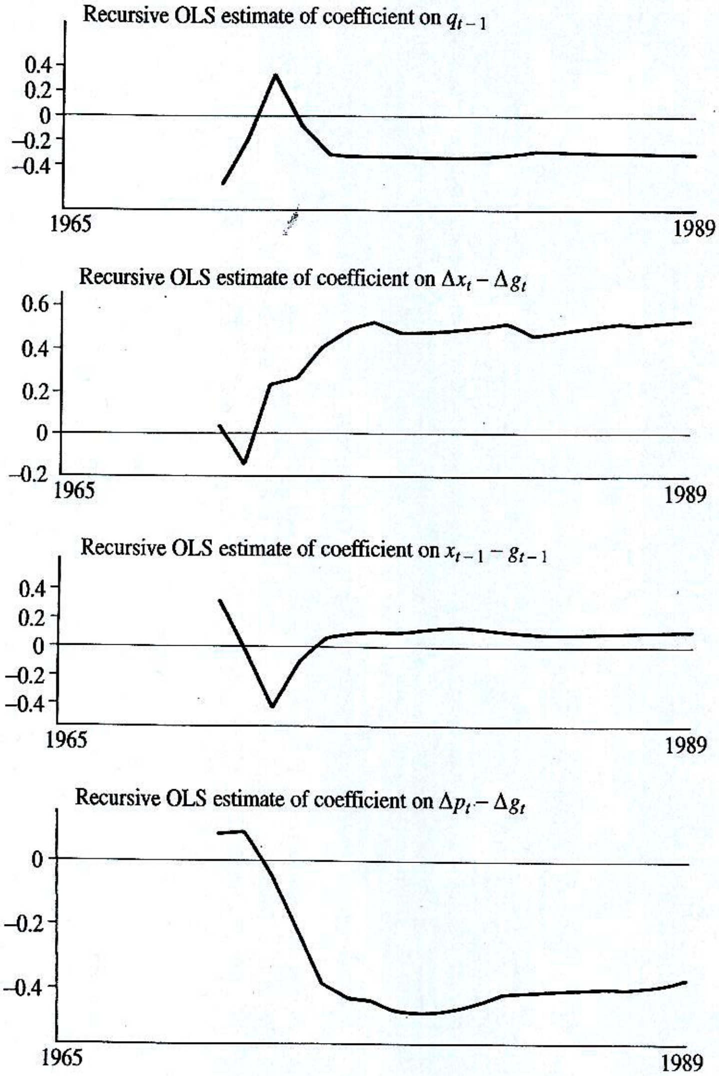

Figure 13.6 Recursive least squares estimates.

Before interpreting (13.42) from an economic viewpoint, we can rewrite it in the error correction form

$$\widehat{\Delta q_t} = 0.581(\Delta x_t - \Delta g_t) - 0.374(\Delta p_t - \Delta g_t) - 0.364[q_{t-1} - 8.21 - 0.277(x_{t-1} - g_{t-1})] \qquad (13.43)$$

The term in square brackets in (13.43) is clearly the disequilibrium error, reflecting the extent of departure from the long-run relationship. We can see that 36.4% of any disequilibrium present in period $t-1$ is corrected for in the current period. The long-run demand-for-food relationship implied is

$$q_t = \beta_0^* + 0.277(x_t - g_t) \tag{13.44}$$

where we shall derive β_0^* shortly. As noted earlier, this data set suggests that in the long run the demand for food is independent of its price and depends only on the level of real income. The long-run real income elasticity is 0.277, well below unity, as is to be expected for a basic necessity.

The demand for food is also homogeneous of degree zero in the short run. However, in the short run, demand depends not only on real total expenditure but also on relative prices, with elasticities of 0.581 and 0.374 respectively. Notice that the short-run real expenditure elasticity is actually greater than the long-run elasticity. It appears that the initial response of demand to change in prices and total expenditure is greater than that in the long run when basic forces of habit and necessity take over and demand to some extent reverts to earlier patterns.

To obtain β_0^* in (13.44), suppose that, in the long run, real income grows at a rate θ, but relative prices remain unchanged. That is,

$$\Delta x_t - \Delta g_t = (x_t - g_t) - (x_{t-1} - g_{t-1}) = \theta$$

and

$$\Delta p_t - \Delta g_t = 0$$

Since the long-run total expenditure elasticity of demand is 0.277, the demand for food therefore grows at a rate

$$\widehat{\Delta q_t} = q_t - q_{t-1} = 0.277\theta$$

Substituting these values into (13.43) gives

$$0.277\theta = 0.581\theta - 0.364[q_t - 8.21 - 0.277(x_t - g_t)]$$

Thus

$$q_t = 8.21 - 0.835\theta + 0.277(x_t - g_t)$$

The long-run demand equation for food is therefore

$$Q = K(X/G)^{0.277}$$

where $K = e^{8.21-0.835\theta}$ depends on the long-run growth rate in real total expenditure, X/G. For example, with a long-run growth rate of 4% per annum, $\theta = 0.04$ and $K = e^{8.177} = 3557$. The long-run relationship is then

$$Q = 3557(X/G)^{0.277}$$

and β_0^* in (13.44) is 8.18. If, however, $\theta = 0$ then $K = e^{8.21} = 3678$, and the long-run relationship is

$$Q = 3678(X/G)^{0.277}$$

and $\beta_0^* = 8.21$, as indicated by the disequilibrium error in (13.43). Thus, as usual in error correction models, the implied equilibrium relationship depends on long-run growth rates.

This exercise illustrates the usefulness of combining the concept of an error correction model with the general-to-specific approach. We have tested down to a final model (13.42) that was not only statistically satisfactory (i.e., in the terminology of Section 12.4, it was data-coherent), but also has sensible economic properties.

This data set also demonstrates the superiority of the general-to-specific methodology over the simple-to-general approach described earlier. We adopted a simple-to-general approach to analysing this data set in Chapters 7, 9 and 11. It led to a number of promising-looking equations, which we now see have to be rejected against the general model used in this and the last chapter. There is, in fact, almost certainly no way in which the simple-to-general approach could have uncovered the model (13.42) that we have finally selected.

Exercise: Replicate everything, try to understand both the methodology and interpretations.