FROM THOMAS: MODERN ECONOMETRICS

COMPUTER EXERCISE

Now that we are aware of the calculations involved in multiple regression, we are well equipped to start estimating equations using a computer package. Suitable packages were mentioned in Chapter 1, and it is assumed the reader has ready access to one. We shall attempt to estimate a demand equation for food using Data Set 1 on the floppy disk which contains 30 annual observations for the US economy.

The economic theory of the consumer suggests that a demand equation for food might have the form

$$Q = f(X, P, G) \tag{7.61}$$

where Q is the quantity of food demanded by US consumers, X is the total money expenditure of consumers, P is a price index for food and G is a general price index. G is included because, although there are no obvious substitutes for food in the everyday sense, all other goods are substitutes in the general sense that they compete with food for the dollars in the consumer's total budget.[4]

Consumer theory also suggests that the above function should be *homogeneous of degree zero in its three explanatory variables*. What this means is that equiproportionate changes in total money expenditure and all prices should leave the demand for food unchanged. For example, a doubling in X, P and G should leave Q constant. If our demand equation is to have this property then it must be possible to rewrite (7.61) as

$$Q = g(X/G, P/G) \tag{7.62}$$

Equation (7.62) implies that the demand for food depends on the *real* total expenditure of consumers, X/G, and on the *relative* price of food, P/G. It is homogeneous of degree zero because, for example, a doubling of X, P and G leaves the ratios X/G and P/G unchanged and hence has no effect on Q.

We shall estimate equations of both forms (7.61) and (7.62) but first we must decide on what functional form to use. Purely for the sake of convenience we shall work in terms of the natural logarithms of variables. This means that we will be able to interpret regression coefficients as elasticities which are the quantities of greatest interest in demand studies. Thus we will give (7.61), for example, the form

$$Q = AX^{\beta_2}P^{\beta_3}G^{\beta_4} \tag{7.63}$$

Taking natural logarithms and adding a disturbance then gives

$$\ln(Q) = \beta_1 + \beta_2 \ln(X) + \beta_3 \ln(P) + \beta_4 \ln(G) + \epsilon \tag{7.64}$$

where $\beta_1 = \ln(A)$. β_2, β_3 and β_4 are demand elasticities. Thus we obtain a linear regression equation to estimate. Similarly, our empirical version of (7.62) is

$$\ln(Q) = \beta_1 + \beta_2 \ln(X/G) + \beta_3 \ln(P/G) + \epsilon \tag{7.65}$$

Another useful advantage of working in logarithmic terms is that the homogeneity property of demand equations can then be very easily expressed algebraically. For (7.64) to be homogeneous of degree zero, it is necessary for the three elasticities to sum to zero. That is, we require $\beta_2 + \beta_3 + \beta_4 = 0$. Enforcing this restriction on the regression parameters in (7.64) will yield the homogeneous equation (7.65). To show this, we simply substitute $\beta_4 = -\beta_2 - \beta_3$ into (7.64). Simple rearrangement will then yield (7.65). Thus if we estimate (7.65) we are implicitly assuming that the elasticities sum to zero.

Before we can actually estimate (7.64) or (7.65), we have to define empirical counterparts to the theoretical variables Q, X, P and G. Since we cannot add together such diverse items as, for example, bananas and sausages, we have to measure the total demand for food in monetary terms. However, since demand is a quantity variable, we shall work in constant prices and define demand as

$$Q = \text{consumers' expenditure on food in 1980 prices}$$

The other variables are easier to define. Since total expenditure X is in nominal money terms, we define it as

$$X = \text{total consumers' expenditure in current prices}$$

We obtain a price index for food by noting that

$$\text{food expenditure in constant prices} = \frac{\text{food expenditure in current prices}}{P}$$

where the price index P is known as the implicit deflator of food expenditure. P can therefore be calculated as

$$P = \frac{\text{food expenditure in current prices}}{\text{food expenditure in 1980 prices}}$$

Similarly, we define the general price index G as the implicit deflator of total consumer expenditure. That is,

$$G = \frac{\text{total consumer expenditure in current prices}}{\text{total consumer expenditure in 1980 prices}}$$

Annual observations for 1963–92 on all the necessary variables are contained in Data Set 1 on the floppy disk and also in Appendix III. The data is taken from the annual publication, *OECD National Accounts*. Transformation routines in your program will enable you to compute values for the ratio variables X/G and P/G very quickly. The routines will also enable you to form the natural logarithms of all variables.

For the moment, we shall make use only of the observations for 1965 through 1989 in Data Set 1.

Excercises:

1) Look at all time series. Do they have some trend?
2) Estimate 7.64 and do all standard tests
3) Estimate 7.64 with parameter restriction $\beta_2 + \beta_3 + \beta_4 = 0$
4) Estimate 7.65
5) Compare results from 3) and 4)
6) Realize that with trending time series, previous results are suspicious

COMPUTER EXERCISE I

In the exercise at the end of Chapter 7 we used Data Set 1 to estimate the demand for food equation (7.67). Although we obtained a high coefficient of determination R^2, we discovered in Section 10.5 that both the Durbin–Watson and the LM test statistics strongly suggested autocorrelation in the residuals of this equation.

Suppose we reinterpret Equation (7.67) as giving the equilibrium long-run demand for food, Q^*, that exists when consumers are fully adjusted to any changes in real expenditure X/G and relative price P/G. That is

β_2 et β_3 = long-run elasticities!

$$\ln (Q^*)_t = \beta_1 + \beta_2 \ln (X/G)_t + \beta_3 \ln (P/G)_t + \epsilon_t \qquad (11.21)$$

Actual demand for food, Q, may differ from equilibrium demand because of habit or the inertia of consumers. Let us specify a logarithmic version of the partial adjustment model (11.12) to relate Q to Q^*:

$$\ln (Q)_t - \ln (Q)_{t-1} = \theta[\ln (Q^*)_t - \ln (Q)_{t-1}] \qquad (11.22)$$

Substituting for $\ln (Q^*)_t$ in (11.22) now gives

$$\ln (Q)_t = \beta_1\theta + \beta_2\theta \ln (X/G)_t + \beta_3\theta \ln (P/G)_t + (1-\theta) \ln (Q)_{t-1} + u_t \qquad (11.23)$$

where $u_t = \theta\epsilon_t$. Equation (11.23) suggests that the problem with our estimated equation (7.67) may have been the fact that we failed to include the lagged dependent variable $\ln (Q)_{t-1}$ among its regressors. Recall from our discussion in Section 10.5 that a frequent cause of autocorrelated residuals is not so much genuine autocorrelation as the omission of relevant regressors. Our poor *dw* and LM statistics may be the result of a mis-specified regression equation. Such an omission of lagged dependent variable(s) is an example of what is sometimes referred to as a **dynamic mis-specification**.

Excercizes:

1) Estimate 11.23 and do all standard tests

2) Interprete parameters for this partial adjustment model

3) Assume for the moment, that 11.23 comes from geometric distributed lags. What would be the interpretation of parameters?

4) Assume for the moment, that 11.23 comes from adaptive expectations. What would be the interpretation of parameters?