CHAPTER 6
Distributed Lags and Dynamic Models

6.1 Introduction

Many economic models have lagged values of the regressors in the regression equation. For example, it takes time to build roads and highways. Therefore, the effect of this public investment on growth in GNP will show up with a lag, and this effect will probably linger on for several years. It takes time before investment in research and development pays off in new inventions which in turn take time to develop into commercial products. In studying consumption behavior, a change in income may affect consumption over several periods. This is true in the permanent income theory of consumption, where it may take the consumer several periods to determine whether the change in real disposable income was temporary or permanent. For example, is the extra consulting money earned this year going to continue next year? Also, lagged values of real disposable income appear in the regression equation because the consumer takes into account his life time earnings in trying to smooth out his consumption behavior. In turn, one's life time income may be guessed by looking at past as well as current earnings. In other words, the regression relationship would look like

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + .. + \beta_s X_{t-s} + u_t \quad t = 1, 2, \ldots, T \tag{6.1}$$

where Y_t denotes the t-th observation on the dependent variable Y and X_{t-s} denotes the $(t\text{-}s)$th observation on the independent variable X. α is the intercept and $\beta_0, \beta_1, \ldots, \beta_s$ are the current and lagged coefficients of X_t. Equation (6.1) is known as a *distributed lag* since it distributes the effect of an increase in income on consumption over s periods. Note that the *short-run* effect of a unit change in X on Y is given by β_o, while the *long-run* effect of a unit change in X on Y is $(\beta_0 + \beta_1 + .. + \beta_s)$.

Suppose that you observe X_t from 1959 to 2007. X_{t-1} is the same variable but for the previous period, i.e., 1958–2006. Since 1958 is not available in this data, the software you are using will start from 1959 for X_{t-1}, and end at 2006. This means that when we lag once, the current X_t series will have to start at 1960 and end at 2007. For practical purposes, this means that when we lag once we loose one observation from the sample. So if we lag s periods, we loose s observations. Furthermore, we are estimating one extra β with every lag. Therefore, there is double jeopardy with respect to loss of degrees of freedom. The number of observations fall (because we are lagging the same series), and the number of parameters to be estimated increase with every lagging variable introduced. Besides the loss of degrees of freedom, the regressors in (6.1) are likely to be highly correlated with each other. In fact most economic time series are usually trended and very highly correlated with their lagged values. This introduces the problem of among the regressors and as we saw in Chapter 4, the higher the multicollinearity among these regressors, the lower is the reliability of the regression estimates.

In this model, OLS is still BLUE because the classical assumptions are still satisfied. All we have done in (6.1) is introduce the additional regressors $(X_{t-1}, \ldots, X_{t-s})$. These regressors are uncorrelated with the disturbances since they are lagged values of X_t, which are by assumption not correlated with u_t for every t.

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_6,

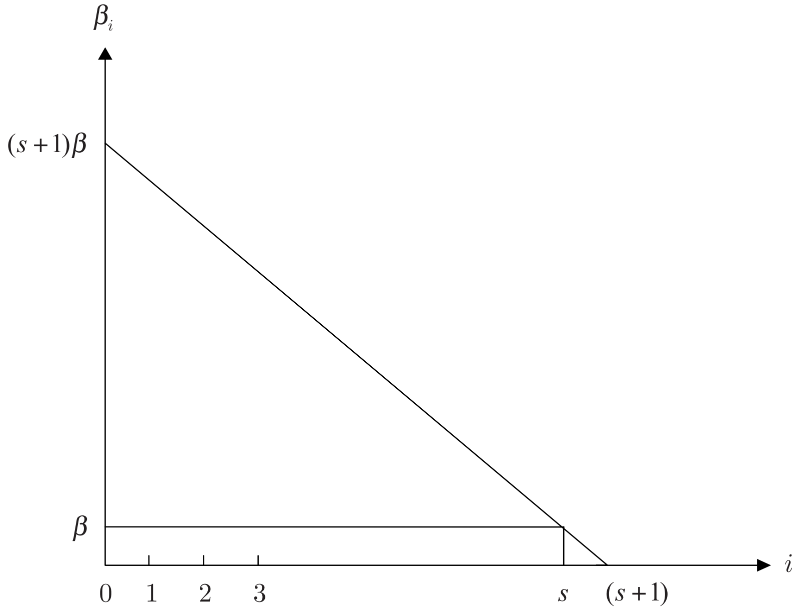

Figure 6.1 Linear Distributed Lag

In order to reduce the degrees of freedom problem, one could impose more structure on the β's. One of the simplest forms imposed on these coefficients is the *linear arithmetic lag*, (see Figure 6.1), which can be written as

$$\beta_i = [(s+1) - i]\beta \quad \text{for } i = 0, 1, \ldots, s \tag{6.2}$$

The lagged coefficients of X follow a linear distributed lag declining arithmetically from $(s+1)\beta$ for X_t to β for X_{t-s}. Substituting (6.2) in (6.1) one gets

$$Y_t = \alpha + \sum_{i=0}^{s} \beta_i X_{t-i} + u_t = \alpha + \beta \sum_{i=0}^{s} [(s+1) - i] X_{t-i} + u_t \tag{6.3}$$

where the latter equation can be estimated by the regression of Y_t on a constant and Z_t, where

$$Z_t = \sum_{i=0}^{s} [(s+1) - i] X_{t-i}$$

This Z_t can be calculated given s and X_t. Hence, we have reduced the estimation of $\beta_0, \beta_1, \ldots, \beta_s$ into the estimation of just one β. Once $\widehat{\beta}$ is obtained, $\widehat{\beta}_i$ can be deduced from (6.2), for $i = 0, 1, \ldots, s$. Despite its simplicity, this lag is too restrictive to impose on the regression and is not usually used in practice.

Alternatively, one can think of $\beta_i = f(i)$ for $i = 0, 1, \ldots, s$. If $f(i)$ is a continuous function, over a closed interval, then it can be approximated by an r-th degree polynomial,

$$f(i) = a_0 + a_1 i + \ldots + a_r i^r$$

For example, if $r = 2$, then

$$\beta_i = a_0 + a_1 i + a_2 i^2 \quad \text{for } i = 0, 1, 2, \ldots, s$$

so that

$$\begin{aligned}
\beta_0 &= a_0 \\
\beta_1 &= a_0 + a_1 + a_2 \\
\beta_2 &= a_0 + 2a_1 + 4a_2 \\
&\vdots \;\; \vdots \;\; \vdots \\
\beta_s &= a_0 + sa_1 + s^2 a_2
\end{aligned}$$

Once a_0, a_1, and a_2 are estimated, $\beta_0, \beta_1, \ldots, \beta_s$ can be deduced. In fact, substituting $\beta_i = a_0 + a_1 i + a_2 i^2$ in (6.1) we get

$$\begin{aligned}
Y_t &= \alpha + \textstyle\sum_{i=0}^{s}(a_0 + a_1 i + a_2 i^2)X_{t-i} + u_t \qquad (6.4)\\
&= \alpha + a_0 \textstyle\sum_{i=0}^{s} X_{t-i} + a_1 \sum_{i=0}^{s} iX_{t-i} + a_2 \sum_{i=0}^{s} i^2 X_{t-i} + u_t
\end{aligned}$$

This last equation, shows that α, a_0, a_1 and a_2 can be estimated from the regression of Y_t on a constant, $Z_0 = \sum_{i=0}^{s} X_{t-i}$, $Z_1 = \sum_{i=0}^{s} iX_{t-i}$ and $Z_2 = \sum_{i=0}^{s} i^2 X_{t-i}$. This procedure was proposed by Almon (1965) and is known as the *Almon lag*. One of the problems with this procedure is the choice of s and r, the number of lags on X_t, and the degree of the polynomial, respectively. In practice, neither is known. Davidson and MacKinnon (1993) suggest starting with a maximum reasonable lag s^* that is consistent with the theory and then based on the unrestricted regression, given in (6.1), checking whether the fit of the model deteriorates as s^* is reduced. Some criteria suggested for this choice include: (i) maximizing $\bar{R}^2$; (ii) minimizing *Akaike's* (1973) *Information Criterion* (AIC) with respect to s. This is given by $AIC(s) = (RSS/T)e^{2s/T}$; or (iii) minimizing Schwarz (1978) *Bayesian Information Criterion* (BIC) with respect to s. This is given by $BIC(s) = (RSS/T)T^{s/T}$ where RSS denotes the residual sum of squares. Note that the AIC and BIC criteria, like $\bar{R}^2$, reward good fit but penalize loss of degrees of freedom associated with a high value of s. These criteria are printed by most regression software including SHAZAM, EViews and SAS. Once the lag length s is chosen it is straight forward to determine r, the degree of the polynomial. Start with a high value of r and construct the Z variables as described in (6.4). If $r = 4$ is the highest degree polynomial chosen and a_4, the coefficient of $Z_4 = \sum_{i=0}^{s} i^4 X_{t-4}$ is insignificant, drop Z_4 and run the regression for $r = 3$. Stop, if the coefficient of Z_3 is significant, otherwise drop Z_3 and run the regression for $r = 2$.

Applied researchers usually impose end point constraints on this Almon lag. A near end point constraint means that $\beta_{-1} = 0$ in equation (6.1). This means that for equation (6.4), this constraint yields the following restriction on the second degree polynomial in a's: $\beta_{-1} = f(-1) = a_0 - a_1 + a_2 = 0$. This restriction allows us to solve for a_0 given a_1 and a_2. In fact, substituting $a_0 = a_1 - a_2$ into (6.4), the regression becomes

$$Y_t = \alpha + a_1(Z_1 + Z_0) + a_2(Z_2 - Z_0) + u_t \qquad (6.5)$$

and once a_1 and a_2 are estimated, a_0 is deduced, and hence the β_i's. This restriction essentially states that X_{t+1} has no effect on Y_t. This may not be a plausible assumption, especially in our consumption example, where income next year enters the calculation of permanent income or life time earnings. A more plausible assumption is the far end point constraint, where $\beta_{s+1} = 0$. This means that $X_{t-(s+1)}$ does not affect Y_t. The further you go back in time, the less is the effect on the current period. All we have to be sure of is that we have gone far back enough

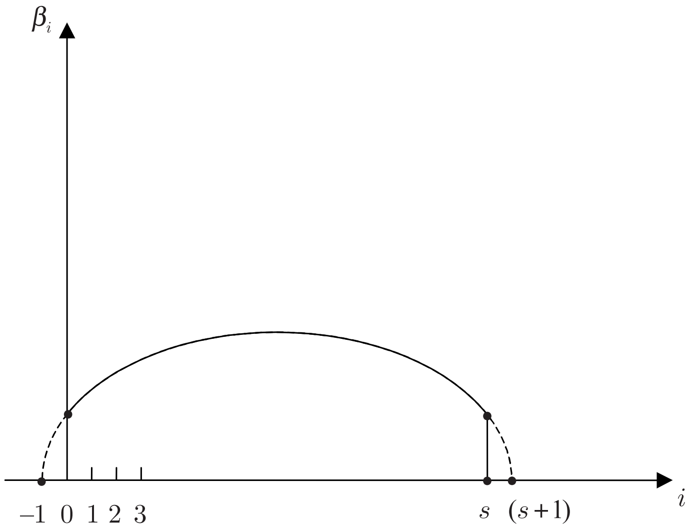

Figure 6.2 A Polynomial Lag with End Point Constraints

to reach an insignificant effect. This far end point constraint is imposed by removing $X_{t-(s+1)}$ from the equation as we have done above. But, some researchers impose this restriction on $\beta_i = f(i)$, i.e., by restricting $\beta_{s+1} = f(s+1) = 0$. This yields for $r = 2$ the following constraint: $a_0+(s+1)a_1+(s+1)^2a_2 = 0$. Solving for a_0 and substituting in (6.4), the constrained regression becomes

$$Y_t = \alpha + a_1[Z_1 - (s+1)Z_0] + a_2[Z_2 - (s+1)^2 Z_0] + u_t \tag{6.6}$$

One can also impose both end point constraints and reduce the regression into the estimation of one a rather than three a's. Note that $\beta_{-1} = \beta_{s+1} = 0$ can be imposed by not including X_{t+1} and $X_{t-(s+1)}$ in the regression relationship. However, these end point restrictions impose the *additional* restrictions that the polynomial on which the a's lie should pass through zero at $i = -1$ and $i = (s+1)$, see Figure 6.2.

These additional restrictions on the polynomial may not necessarily be true. In other words, the polynomial could intersect the X-axis at points other than -1 or $(s+1)$. Imposing a restriction, whether true or not, reduces the variance of the estimates, and introduces bias if the restriction is untrue. This is intuitive, because this restriction gives *additional* information which should increase the reliability of the estimates. The reduction in variance and the introduction of bias naturally lead to Mean Square Error criteria that help determine whether these restrictions should be imposed, see Wallace (1972). These criteria are beyond the scope of this chapter. In general, one should be careful in the use of restrictions that may not be plausible or even valid. In fact, one should always test these restrictions before using them. See Schmidt and Waud (1975).

Empirical Example: Using the Consumption-Income data from the Economic Report of the President over the period 1959–2007, given in Table 5.3, we estimate a consumption-income regression imposing a five year lag on income. In this case, all variables are in *log* and $s = 5$ in equation (6.1). Table 6.1 gives the Stata output imposing the *linear arithmetic* lag given in equation (6.2).

The regression output reports $\widehat{\beta} = 0.0498$ which is statistically significant with a t-value of 64.4. One can test the arithmetic lag restrictions jointly using an F-test. The Unrestricted

Table 6.1 Regression with Arithmetic Lag Restriction

```
. tsset year
        time variable:    year, 1959 to 2007
                delta:    1 unit
. gen ly=ln(y)
. gen lc=ln(c)
. gen z=6*ly+5*l.ly+4*l2.ly+3*l3.ly+2*l4.ly+l5.ly
(5 missing values generated)
. reg lc z
```

Source	SS	df	MS			
				Number of obs	=	44
				F(1, 42)	=	4149.96
Model	3.5705689	1	3.5705689	Prob > F	=	0.0000
Residual	.036136249	42	.000860387	R-squared	=	0.9900
				Adj R-squared	=	0.9897
Total	3.60670515	43	.083876864	Root MSE	=	.02933

lc	Coef.	Std. Err.	t	$P > \|t\|$	[95% Conf.	Interval]
Z	.049768	.0007726	64.42	0.000	.0482089	.0513271
_cons	−.5086255	.1591019	–3.20	0.003	−.8297061	−.1875449

Table 6.2 Almon Polynomial, $r = 2, s = 5$ and Near End-Point Constraint

Dependent Variable = LNC
Method: Least Squares
Sample (adjusted): 1964 2007
Included observations: 44 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	–0.770611	0.201648	3.821563	0.0004
PDL01	0.342152	0.056727	6.031589	0.0000
PDL02	0.067215	0.012960	–5.186494	0.0000

R-squared	0.990054	Mean dependent var	9.736786
Adjusted R-squared	0.989568	S.D. dependent var	0.289615
S.E. of regression	0.029580	Akaike info criterion	–4.137705
Sum squared resid	0.035874	Schwarz criterion	–4.016055
Log likelihood	94.02950	Hannan-Quinn criter.	–4.092591
F-statistic	2040.559	Durbin-Watson stat	0.382851
Prob(F-statistic)	0.000000		

Lag Distribution of LNY	i	Coefficient	Std. Error	t-Statistic
. ⋆ |	0	0.27494	0.04377	6.28161
. ⋆|	1	0.41544	0.06162	6.74167
. ⋆|	2	0.42152	0.05358	7.86768
. ⋆ |	3	0.29317	0.01976	14.8332
⋆ |	4	0.03039	0.04056	0.74937
⋆ . |	5	0.36682	0.12630	2.90445
	Sum of Lags	1.06865	0.01976	54.0919

Residual Sum of Squares (URSS) is obtained by regressing C_t on $Y_t, Y_{t-1}, \ldots, Y_{t-5}$ and a constant. This yields $URSS = 0.016924$. The RRSS is given in Table 6.1 as 0.036136 and it involves imposing 5 restrictions given in (6.2). Therefore,

$$F = \frac{(0.036136249 - 0.016924337)/5}{0.016924337/37} = 8.40$$

and this is distributed as $F_{5,37}$ under the null hypothesis. This rejects the linear arithmetic lag restrictions.

Next we impose an Almon lag based on a second degree polynomial as described in equation (6.4). Table 6.2 reports the EViews output for $s = 5$ imposing the *near end point* constraint. To do this using EViews, one replaces the regressor Y by PDL$(Y, 5, 2, 1)$ indicating a request to fit a five year Almon lag on Y that is of the second-order degree, with a *near end point* constraint. In this case, the estimated regression coefficients rise and then fall becoming negative: $\widehat{\beta}_0 = 0.275, \widehat{\beta}_1 = 0.415, \ldots, \widehat{\beta}_5 = -0.367$. Note that $\widehat{\beta}_4$ is statistically insignificant. The Almon lag restrictions can be jointly tested using Chow's F-statistic. The URSS is obtained from the unrestricted regression of C_t on $Y_t, Y_{t-1}, \ldots, Y_{t-5}$ and a constant. This was reported above as $URSS = 0.016924$.

Table 6.3 Almon Polynomial, $r = 2, s = 5$ and Far End-Point Constraint

Dependent Variable = LNC
Method: Least Squares
Sample (adjusted): 1964 2007
Included observations: 44 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	–1.107052	0.158405	–6.988756	0.0000
PDL01	0.134206	0.011488	11.68247	0.0000
PDL02	–0.259490	0.036378	–7.133152	0.0000

R-squared	0.994467	Mean dependent var	9.736786
Adjusted R-squared	0.994197	S.D. dependent var	0.289615
S.E. of regression	0.022062	Akaike info criterion	–4.724198
Sum squared resid	0.019956	Schwarz criterion	–4.602548
Log likelihood	106.9324	Hannan-Quinn criter.	–4.679084
F-statistic	3684.610	Durbin-Watson stat	0.337777
Prob(F-statistic)	0.000000		

Lag Distribution of LNY	i	Coefficient	Std. Error	t-Statistic
. *|	0	0.87912	0.10074	8.72642
. * |	1	0.45018	0.03504	12.8475
. * |	2	0.13421	0.01149	11.6825
*. |	3	–0.06880	0.03787	–1.81686
* . |	4	–0.15884	0.04483	–3.54337
* . |	5	–0.13590	0.03221	–4.21962
	Sum of Lags	1.09997	0.01547	71.0964

The RRSS, given in Table 6.2, is 0.035874 and involves four restrictions. Therefore,

$$F = \frac{(0.03587367 - 0.016924337)/4}{0.016924337/37} = 10.357$$

and this is distributed as $F_{4,37}$ under the null hypothesis. This rejects the second degree polynomial Almon lag specification with a near end point constraint.

Table 6.3 reports the EViews output for $s = 5$, imposing the far end point constraint. To do this using EViews, one replaces the regressor Y by PDL$(Y, 5, 2, 2)$ indicating a request to fit a five year Almon lag on Y that is of the second-order degree, with a *far* end point constraint. In this case, the $\widehat{\beta}$'s are positive, then becoming negative, $\widehat{\beta}_0 = 0.879, \widehat{\beta}_1 = 0.450, \ldots, \widehat{\beta}_5 = -0.136$, all being statistically significant. This second degree polynomial Almon lag specification with a far end point constraint can be tested against the unrestricted lag model using Chow's F-statistic. The RRSS, given in Table 6.3, is 0.019955101 and involves four restrictions. Therefore,

$$F = \frac{(0.019955101 - 0.016924337)/4}{0.016924337/37} = 1.656$$

and this is distributed as $F_{4,37}$ under the null hypothesis. This does not reject the restrictions imposed by this model.

6.2 Infinite Distributed Lag

So far we have been dealing with a finite number of lags imposed on X_t. Some lags may be infinite. For example, the investment in building highways and roads several decades ago may still have an effect on today's growth in GNP. In this case, we write equation (6.1) as

$$Y_t = \alpha + \textstyle\sum_{i=0}^{\infty} \beta_i X_{t-i} + u_t \quad t = 1, 2, \ldots, T. \tag{6.7}$$

There are an infinite number of β_i's to estimate with only T observations. This can only be feasible if more structure is imposed on the β_i's. First, we normalize these β_i's by their sum, i.e., let $w_i = \beta_i/\beta$ where $\beta = \sum_{i=0}^{\infty} \beta_i$. If all the β_i's have the *same sign*, then the β_i's take the sign of β and $0 \le w_i \le 1$ for all i, with $\sum_{i=0}^{\infty} w_i = 1$. This means that the w_i's can be interpreted as probabilities. In fact, Koyck (1954) imposed the geometric lag on the w_i's, i.e., $w_i = (1 - \lambda)\lambda^i$ for $i = 0, 1, \ldots, \infty$[1]. Substituting

$$\beta_i = \beta w_i = \beta(1 - \lambda)\lambda^i$$

in (6.7) we get

$$Y_t = \alpha + \beta(1 - \lambda) \textstyle\sum_{i=0}^{\infty} \lambda^i X_{t-i} + u_t \tag{6.8}$$

Equation (6.8) is known as the *infinite distributed* lag form of the *Koyck lag*. The short-run effect of a unit change in X_t on Y_t is given by $\beta(1 - \lambda)$; whereas the long-run effect of a unit change in X_t on Y_t is $\sum_{i=0}^{\infty} \beta_i = \beta \sum_{i=0}^{\infty} w_i = \beta$. Implicit in the Koyck lag structure is that the effect of a unit change in X_t on Y_t declines the further back we go in time. For example, if $\lambda = 1/2$, then $\beta_0 = \beta/2$, $\beta_1 = \beta/4$, $\beta_2 = \beta/8$, etc. Defining $LX_t = X_{t-1}$, as the lag operator, we have $L^i X_t = X_{t-i}$, and (6.8) reduces to

$$Y_t = \alpha + \beta(1 - \lambda) \textstyle\sum_{i=0}^{\infty} (\lambda L)^i X_t + u_t = \alpha + \beta(1 - \lambda) X_t/(1 - \lambda L) + u_t \tag{6.9}$$

where we have used the fact that $\sum_{i=0}^{\infty} c^i = 1/(1-c)$. Multiplying the last equation by $(1-\lambda L)$ one gets

$$Y_t - \lambda Y_{t-1} = \alpha(1-\lambda) + \beta(1-\lambda)X_t + u_t - \lambda u_{t-1}$$

or

$$Y_t = \lambda Y_{t-1} + \alpha(1-\lambda) + \beta(1-\lambda)X_t + u_t - \lambda u_{t-1} \tag{6.10}$$

This is the *autoregressive form* of the infinite distributed lag. It is autoregressive because it includes the lagged value of Y_t as an explanatory variable. Note that we have reduced the problem of estimating an infinite number of β_i's into estimating λ and β from (6.10). However, OLS would lead to biased and inconsistent estimates, because (6.10) contains a lagged dependent variable as well as serially correlated errors. In fact the error in (6.10) is a Moving Average process of order one, i.e., MA(1), see Chapter 14. We digress at this stage to give two econometric models which would lead to equations resembling (6.10).

6.2.1 Adaptive Expectations Model (AEM)

Suppose that output Y_t is a function of expected sales X_t^* and that the latter is unobservable, i.e.,

$$Y_t = \alpha + \beta X_t^* + u_t$$

where expected sales are updated according to the following method

$$X_t^* - X_{t-1}^* = \delta(X_t - X_{t-1}^*) \tag{6.11}$$

that is, expected sales at time t is a weighted combination of expected sales at time $t-1$ and actual sales at time t. In fact,

$$X_t^* = \delta X_t + (1-\delta)X_{t-1}^* \tag{6.12}$$

Equation (6.11) is also an error learning model, where one learns from past experience and adjust expectations after observing current sales. Using the lag operator L, (6.12) can be rewritten as $X_t^* = \delta X_t/[1-(1-\delta)L]$. Substituting this last expression in the above relationship, we get

$$Y_t = \alpha + \beta\delta X_t/[1-(1-\delta)L] + u_t \tag{6.13}$$

Multiplying both sides of (6.13) by $[1-(1-\delta)L]$, we get

$$Y_t - (1-\delta)Y_{t-1} = \alpha[(1-(1-\delta)] + \beta\delta X_t + u_t - (1-\delta)u_{t-1} \tag{6.14}$$

(6.14) looks exactly like (6.10) with $\lambda = (1-\delta)$.

6.2.2 Partial Adjustment Model (PAM)

Under this model there is a cost of being out of equilibrium and a cost of adjusting to that equilibrium, i.e.,

$$Cost = a(Y_t - Y_t^*)^2 + b(Y_t - Y_{t-1})^2 \tag{6.15}$$

where Y_t^* is the target or equilibrium level for Y, whereas Y_t is the current level of Y. The first term of (6.15) gives a quadratic loss function proportional to the distance of Y_t from the equilibrium level Y_t^*. The second quadratic term represents the cost of adjustment. Minimizing this quadratic cost function with respect to Y, we get $Y_t = \gamma Y_t^* + (1-\gamma)Y_{t-1}$, where $\gamma = a/(a+b)$. Note that if the cost of adjustment was zero, then $b = 0$, $\gamma = 1$, and the target is reached immediately. However, there are costs of adjustment, especially in building the desired capital stock. Hence,

$$Y_t = \gamma Y_t^* + (1-\gamma)Y_{t-1} + u_t \tag{6.16}$$

where we made this relationship stochastic. If the true relationship is $Y_t^* = \alpha + \beta X_t$, then from (6.16)

$$Y_t = \gamma\alpha + \gamma\beta X_t + (1-\gamma)Y_{t-1} + u_t \tag{6.17}$$

and this looks like (6.10) with $\lambda = (1-\gamma)$, except for the error term, which is not necessarily MA(1) with the Moving Average parameter λ.

6.3 Estimation and Testing of Dynamic Models with Serial Correlation

Both the AEM and the PAM give equations resembling the autoregressive form of the infinite distributed lag. In all cases, we end up with a lagged dependent variable and an error term that is either Moving Average of order one as in (6.10), or just classical or autoregressive as in (6.17). In this section we study the testing and estimation of such *autoregressive* or *dynamic* models.

If there is a Y_{t-1} in the regression equation and the u_t's are classical disturbances, as may be the case in equation (6.17), then Y_{t-1} is said to be *contemporaneously uncorrelated* with the disturbance term u_t. In fact, the disturbances satisfy assumptions 1–4 of Chapter 3 and $E(Y_{t-1}u_t) = 0$ even though $E(Y_{t-1}u_{t-1}) \neq 0$. In other words, Y_{t-1} is not correlated with the current disturbance u_t but it is correlated with the lagged disturbance u_{t-1}. In this case, as long as the disturbances are not serially correlated, OLS will be biased, but remains consistent and asymptotically efficient. This case is unlikely with economic data given that most macro time-series variables are highly trended. More likely, the u_t's are serially correlated. In this case, OLS is biased and inconsistent. Intuitively, Y_t is related to u_t, so Y_{t-1} is related to u_{t-1}. If u_t and u_{t-1} are correlated, then Y_{t-1} and u_t are correlated. This means that one of the regressors, lagged Y, is correlated with u_t and we have the problem of endogeneity. Let us demonstrate what happens to OLS for the simple autoregressive model with no constant

$$Y_t = \beta Y_{t-1} + \nu_t \quad |\beta| < 1 \quad t = 1, 2, \ldots, T \tag{6.18}$$

with $\nu_t = \rho\nu_{t-1} + \epsilon_t$, $|\rho| < 1$ and $\epsilon_t \sim \text{IIN}(0, \sigma_\epsilon^2)$. One can show, see problem 3, that

$$\widehat{\beta}_{OLS} = \textstyle\sum_{t=2}^T Y_t Y_{t-1} / \sum_{t=2}^T Y_{t-1}^2 = \beta + \sum_{t=2}^T Y_{t-1}\nu_t / \sum_{t=2}^T Y_{t-1}^2$$

with $\text{plim}(\widehat{\beta}_{OLS} - \beta) = \text{asymp. bias}(\widehat{\beta}_{OLS}) = \rho(1-\beta^2)/(1+\rho\beta)$. This asymptotic bias is positive if $\rho > 0$ and negative if $\rho < 0$. Also, this asymptotic bias can be large for small values of β and large values of ρ. For example, if $\rho = 0.9$ and $\beta = 0.2$, the asymptotic bias for β is 0.73. This is more than 3 times the value of β.

Also, $\widehat{\rho} = \sum_{t=2}^{T} \widehat{\nu}_t \widehat{\nu}_{t-1} / \sum_{t=2}^{T} \widehat{\nu}_{t-1}^2$ where $\widehat{\nu}_t = Y_t - \widehat{\beta}_{OLS} Y_{t-1}$ has

$$\text{plim}(\widehat{\rho} - \rho) = -\rho(1-\beta^2)/(1+\rho\beta) = -\text{ asymp.bias}(\widehat{\beta}_{OLS})$$

This means that if $\rho > 0$, then $\widehat{\rho}$ would be negatively biased. However, if $\rho < 0$, then $\widehat{\rho}$ is positively biased. In both cases, $\widehat{\rho}$ is biased towards zero. In fact, the asymptotic bias of the D.W. statistic is twice the asymptotic bias of $\widehat{\beta}_{OLS}$, see problem 3. This means that the D.W. statistic is biased towards not rejecting the null hypothesis of zero serial correlation. Therefore, if the D.W. statistic rejects the null of $\rho = 0$, it is doing that when the odds are against it, and therefore confirming our rejection of the null and the presence of serial correlation. If on the other hand it does not reject the null, then the D.W. statistic is uninformative and has to be replaced by another conclusive test for serial correlation. Such an alternative test in the presence of a lagged dependent variable has been developed by Durbin (1970), and the statistic computed is called Durbin's h. Using (6.10) or (6.17), one computes OLS ignoring its possible bias and $\widehat{\rho}$ from OLS residuals as shown above. Durbin's h is given by

$$h = \widehat{\rho}[n/(1 - n\ \widehat{\text{var}}(\text{coeff. of } Y_{t-1}))]^{1/2}. \tag{6.19}$$

This is asymptotically distributed $N(0,1)$ under null hypothesis of $\rho = 0$. If $n[\widehat{\text{var}}(\text{coeff. of } Y_{t-1})]$ is greater than one, then h cannot be computed, and Durbin suggests running the OLS residuals e_t on e_{t-1} and the regressors in the model (including the lagged dependent variable), and testing whether the coefficient of e_{t-1} in this regression is significant. In fact, this test can be generalized to higher order autoregressive errors. Let u_t follow an AR(p) process

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + .. + \rho_p u_{t-p} + \epsilon_t$$

then this test involves running e_t on $e_{t-1}, e_{t-2}, \ldots, e_{t-p}$ and the regressors in the model including Y_{t-1}. The test statistic for $H_0; \rho_1 = \rho_2 = .. = \rho_p = 0;$ is TR^2 which is distributed χ_p^2. This is the Lagrange multiplier test developed independently by Breusch (1978) and Godfrey (1978) and discussed in Chapter 5. In fact, this test has other useful properties. For example, this test is the same whether the null imposes an AR(p) model or an MA(p) model on the disturbances, see Chapter 14. Kiviet (1986) argues that even though these are large sample tests, the Breusch-Godfrey test is preferable to Durbin's h in small samples.

6.3.1 A Lagged Dependent Variable Model with AR(1) Disturbances

A model with a lagged dependent variable and an autoregressive error term is estimated using instrumental variables (IV). This method will be studied extensively in Chapter 11. In short, the IV method corrects for the correlation between Y_{t-1} and the error term by replacing Y_{t-1} with its predicted value $\widehat{Y}_{t-1}$. The latter is obtained by regressing Y_{t-1} on some exogenous variables, say a set of Z's, which are called a set of instruments for Y_{t-1}. Since these variables are exogenous and uncorrelated with u_t, $\widehat{Y}_{t-1}$ will not be correlated with u_t. Suppose the regression equation is

$$Y_t = \alpha + \beta Y_{t-1} + \gamma X_t + u_t \qquad t = 2, \ldots, T \tag{6.20}$$

and that at least one exogenous variable Z_t exists which will be our instrument for Y_{t-1}. Regressing Y_{t-1} on X_t, Z_t and a constant, we get

$$Y_{t-1} = \widehat{Y}_{t-1} + \widehat{\nu}_t = \widehat{a}_1 + \widehat{a}_2 Z_t + \widehat{a}_3 X_t + \widehat{\nu}_t. \tag{6.21}$$

Then $\widehat{Y}_{t-1} = \widehat{a}_1 + \widehat{a}_2 Z_t + \widehat{a}_3 X_t$ and is independent of u_t, because it is a linear combination of exogenous variables. But, Y_{t-1} is correlated with u_t. This means that $\widehat{\nu}_t$ is the part of Y_{t-1} that is correlated with u_t. Substituting $Y_{t-1} = \widehat{Y}_{t-1} + \widehat{\nu}_t$ in (6.20) we get

$$Y_t = \alpha + \beta \widehat{Y}_{t-1} + \gamma X_t + (u_t + \beta \widehat{\nu}_t) \tag{6.22}$$

$\widehat{Y}_{t-1}$ is uncorrelated with the new error term $(u_t + \beta\widehat{\nu}_t)$ because $\Sigma \widehat{Y}_{t-1}\widehat{\nu}_t = 0$ from (6.21). Also, X_t is uncorrelated with u_t by assumption. But, from (6.21), X_t also satisfies $\Sigma X_t \widehat{\nu}_t = 0$. Hence, X_t is uncorrelated with the new error term $(u_t + \beta\widehat{\nu}_t)$. This means that OLS applied to (6.22) will lead to consistent estimates of α, β and γ. The only remaining question is where do we find instruments like Z_t? This Z_t should be (i) uncorrelated with u_t, (ii) preferably predicting Y_{t-1} fairly well, but, not predicting it perfectly, otherwise $\widehat{Y}_{t-1} = Y_{t-1}$. If this happens, we are back to OLS which we know is inconsistent, (iii) $\Sigma z_t^2/T$ should be finite and different from zero. Recall that $z_t = Z_t - \bar{Z}$. In this case, X_{t-1} seems like a natural instrumental variable candidate. It is an exogenous variable which would very likely predict Y_{t-1} fairly well, and satisfies $\Sigma x_{t-1}^2/T$ being finite and different from zero. In other words, (6.21) regresses Y_{t-1} on a constant, X_{t-1} and X_t, and gets $\widehat{Y}_{t-1}$. Additional lags on X_t can be used as instruments to improve the small sample properties of this estimator. Substituting $\widehat{Y}_{t-1}$ in equation (6.22) results in consistent estimates of the regression parameters. Wallis (1967) substituted these consistent estimates in the original equation (6.20) and obtained the residuals $\widetilde{u}_t$. Then he computed

$$\widehat{\rho} = [\textstyle\sum_{t=2}^{T} \widetilde{u}_t \widetilde{u}_{t-1}/(T-1)]/[\sum_{t=1}^{T} \widetilde{u}_t^2/T] + (3/T)$$

where the last term corrects for the bias in $\widehat{\rho}$. At this stage, one can perform a Prais-Winsten procedure on (6.20) using $\widehat{\rho}$ instead of ρ, see Fomby and Guilkey (1983).

An alternative two-step procedure has been proposed by Hatanaka (1974). After estimating (6.22) and obtaining the residuals $\widetilde{u}_t$ from (6.20), Hatanaka (1974) suggests running $Y_t^* = Y_t - \widetilde{\rho} Y_{t-1}$ on $Y_{t-1}^* = Y_{t-1} - \widetilde{\rho} Y_{t-2}$, $X_t^* = X_t - \widetilde{\rho} X_{t-1}$ and $\widetilde{u}_{t-1}$. Note that this is the Cochrane-Orcutt transformation which ignores the first observation. Also, $\widetilde{\rho} = \sum_{t=3}^{T} \widetilde{u}_t \widetilde{u}_{t-1} / \sum_{t=3}^{T} \widetilde{u}_t^2$ ignores the small sample bias correction factor suggested by Wallis (1967). Let $\widetilde{\delta}$ be the coefficient of $\widetilde{u}_{t-1}$, then the efficient estimator of ρ is given by $\widetilde{\widetilde{\rho}} = \widetilde{\rho} + \widetilde{\delta}$. Hatanaka shows that the resulting estimators are asymptotically equivalent to the MLE in the presence of Normality.

Empirical Example: Consider the Consumption-Income data from the Economic Report of the President over the period 1959–2007 given in Table 5.3. Problem 5 asks the reader to verify that Durbin's h obtained from the lagged dependent variable model described in (6.20) yields a value of 3.50. This is asymptotically distributed as $N(0, 1)$ under the null hypothesis of no serial correlation of the disturbances. This null is soundly rejected. The Bruesch and Godfrey test runs the regression of OLS residuals on their lagged values and the regressors in the model. This yields a $TR^2 = 13.37$. This is distributed as χ_1^2 under the null and has a p-value of 0.0003. Therefore, we reject the hypothesis of no first-order serial correlation. Next, we estimate (6.20) using current and lagged values of income (Y_t, Y_{t-1} and Y_{t-2}) as a set of instruments for lagged consumption (C_{t-1}). The regression given by (6.22), yields:

$$\begin{aligned} C_t = &-0.831 + 0.104\ \widehat{C}_{t-1} + 1.177\ Y_t + \text{residuals} \\ &(0.280) \quad (0.303) \qquad\quad (0.326) \end{aligned}$$

Substituting these estimates in (6.20), one gets the residuals $\widetilde{u}_t$. Based on these $\widetilde{u}_t$'s, the Wallis (1967) estimate of ρ yields $\widehat{\rho} = 0.907$ and Hatanaka's (1974) estimate of ρ yields $\widetilde{\rho} = 0.828$. Running the Hatanaka regression gives

$$\begin{aligned} C_t^* = &-0.142 + 0.233\ C_{t-1}^* + 0.843\ Y_t^* + 0.017\ \widetilde{u}_{t-1} + \text{residuals} \\ &(0.053) \quad (0.083) \qquad\quad (0.095) \qquad (0.058) \end{aligned}$$

where $C_t^* = C_t - \widetilde{\rho} C_{t-1}$. The efficient estimate of ρ is given by $\widetilde{\widetilde{\rho}} = \widetilde{\rho} + 0.017 = 0.846$.

6.3.2 A Lagged Dependent Variable Model with MA(1) Disturbances

Zellner and Geisel (1970) estimated the Koyck autoregressive representation of the infinite distributed lag, given in (6.10). In fact, we saw that this could also arise from the AEM, see (6.14). In particular, it is a regression with a lagged dependent variable and an MA(1) error term with the added restriction that the coefficient of Y_{t-1} is the same as the MA(1) parameter. For simplicity, we write

$$Y_t = \alpha + \lambda Y_{t-1} + \beta X_t + (u_t - \lambda u_{t-1}) \tag{6.23}$$

Let $w_t = Y_t - u_t$, then (6.23) becomes

$$w_t = \alpha + \lambda w_{t-1} + \beta X_t \tag{6.24}$$

By continuous substitution of lagged values of w_t in (6.24) we get

$$w_t = \alpha(1 + \lambda + \lambda^2 + .. + \lambda^{t-1}) + \lambda^t w_0 + \beta(X_t + \lambda X_{t-1} + .. + \lambda^{t-1} X_1)$$

and replacing w_t by $(Y_t - u_t)$, we get

$$Y_t = \alpha(1 + \lambda + \lambda^2 + .. + \lambda^{t-1}) + \lambda^t w_0 + \beta(X_t + \lambda X_{t-1} + .. + \lambda^{t-1} X_1) + u_t \tag{6.25}$$

knowing λ, this equation can be estimated via OLS assuming that the disturbances u_t are not serially correlated. Since λ is not known, Zellner and Geisel (1970) suggest a search procedure over λ, where $0 < \lambda < 1$. The regression with the minimum residual sums of squares gives the optimal λ, and the corresponding regression gives the estimates of α, β and w_0. The last coefficient $w_o = Y_o - u_o = E(Y_o)$ can be interpreted as the expected value of the initial observation on the dependent variable. Klein (1958) considered the direct estimation of the infinite Koyck lag, given in (6.8) and arrived at (6.25). The search over λ results in MLEs of the coefficients. Note, however, that the estimate of w_o is not consistent. Intuitively, as t tends to infinity, λ^t tends to zero implying no new information to estimate w_o. In fact, some applied researchers ignore the variable λ^t in the regression given in (6.25). This practice, known as truncating the remainder, is not recommended since the Monte Carlo experiments of Maddala and Rao (1971) and Schmidt (1975) have shown that even for $T = 60$ or 100, it is not desirable to omit λ^t from (6.25).

In summary, we have learned how to estimate a dynamic model with a lagged dependent variable and serially correlated errors. In case the error is autoregressive of order one, we have outlined the steps to implement the Wallis Two-Stage estimator and Hatanaka's two-step procedure. In case the error is Moving Average of order one, we have outlined the steps to implement the Zellner-Geisel procedure.

6.4 Autoregressive Distributed Lag

So far, section 6.1 considered finite distributed lags on the explanatory variables, whereas section 6.2 considered an autoregressive relation including the first lag of the dependent variable and current values of the explanatory variables. In general, economic relationships may be generated by an *Autoregressive Distributed Lag* (ADL) scheme. The simplest form is the ADL (1,1) model which is given by

$$Y_t = \alpha + \lambda Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + u_t \tag{6.26}$$

where both Y_t and X_t are lagged once. By specifying higher order lags for Y_t and X_t, say an ADL (p, q) with p lags on Y_t and q lags on X_t, one can test whether the specification now is general enough to ensure White noise disturbances. Next, one can test whether some restrictions can be imposed on this general model, like reducing the order of the lags to arrive at a simpler ADL model, or estimating the simpler static model with the Cochrane-Orcutt correction for serial correlation, see problem 20 in Chapter 7. This general to specific modelling strategy is prescribed by David Hendry and is utilized by the econometric software PC-Give, see Gilbert (1986).

Returning to the ADL $(1, 1)$ model in (6.26) one can invert the autoregressive form as follows:

$$Y_t = \alpha(1 + \lambda + \lambda^2 + ..) + (1 + \lambda L + \lambda^2 L^2 + ..)(\beta_0 X_t + \beta_1 X_{t-1} + u_t) \tag{6.27}$$

provided $|\lambda| < 1$. This equation gives the effect of a unit change in X_t on future values of Y_t. In fact, $\partial Y_t/\partial X_t = \beta_0$ while $\partial Y_{t+1}/\partial X_t = \beta_1 + \lambda\beta_0$, etc. This gives the immediate short-run responses with the long-run effect being the sum of all these partial derivatives yielding $(\beta_0+\beta_1)/(1-\lambda)$. This can be alternatively derived from (6.26) at the long-run static equilibrium (Y^*, X^*) where $Y_t = Y_{t-1} = Y^*$, $X_t = X_{t-1} = X^*$ and the disturbance is set equal to zero, i.e.,

$$Y^* = \frac{\alpha}{1-\lambda} + \frac{\beta_0 + \beta_1}{1-\lambda} X^* \tag{6.28}$$

Replacing Y_t by $Y_{t-1} + \Delta Y_t$ and X_t by $X_{t-1} + \Delta X_t$ in (6.26) one gets

$$\Delta Y_t = \alpha + \beta_0 \Delta X_t - (1-\lambda) Y_{t-1} + (\beta_0 + \beta_1) X_{t-1} + u_t$$

This can be rewritten as

$$\Delta Y_t = \beta_0 \Delta X_t - (1-\lambda)\left[Y_{t-1} - \frac{\alpha}{1-\lambda} - \frac{\beta_0 + \beta_1}{1-\lambda} X_{t-1}\right] + u_t \tag{6.29}$$

Note that the term in brackets contains the long-run equilibrium parameters derived in (6.28). In fact, the term in brackets represents the deviation of Y_{t-1} from the long-run equilibrium term corresponding to X_{t-1}. Equation (6.29) is known as the *Error Correction Model* (ECM), see Davidson, Hendry, Srba and Yeo (1978). Y_t is obtained from Y_{t-1} by adding the short-run effect of the change in X_t and a long-run equilibrium adjustment term. Since, the disturbances are White noise, this model is estimated by OLS.

Note

1. Other distributions besides the geometric distribution can be considered. In fact, a Pascal distribution was considered by Solow (1960), a rational-lag distribution was considered by Jorgenson (1966), and a Gamma distribution was considered by Schmidt (1974, 1975). See Maddala (1977) for an excellent review.

Problems

1. Consider the Consumption-Income data given in Table 5.3 and provided on the Springer web site as CONSUMP.DAT. Estimate a Consumption-Income regression in logs that allows for a six year lag on income as follows:

   (a) Use the linear arithmetic lag given in equation (6.2). Show that this result can also be obtained as an Almon lag first-degree polynomial with a far end point constraint.

   (b) Use an Almon lag second-degree polynomial, described in equation (6.4), imposing the near end point constraint.

   (c) Use an Almon lag second-degree polynomial imposing the far end point constraint.

   (d) Use an Almon lag second-degree polynomial imposing both end point constraints.

   (e) Using Chow's F-statistic, test the arithmetic lag restrictions given in part (a).

   (f) Using Chow's F-statistic, test the Almon lag restrictions implied by the model in part (b).

   (g) Repeat part (f) for the restrictions imposed in parts (c) and (d).

2. Consider fitting an Almon lag third degree polynomial $\beta_i = a_0 + a_1 i + a_2 i^2 + a_3 i^3$ for $i = 0, 1, \ldots, 5$, on the Consumption-Income relationship in logarithms. In this case, there are five lags on income, i.e., $s = 5$.

   (a) Set up the estimating equation for the a_i's and report the estimates using OLS.

   (b) What is your estimate of β_3? What is the standard error? Can you relate the var$(\widehat{\beta}_3)$ to the variances and covariances of the a_i's?

   (c) How would the OLS regression in part (a) change if we impose the near end point constraint $\beta_{-1} = 0$?

   (d) Test the near end point constraint.

   (e) Test the Almon lag specification given in part (a) against an unrestricted five year lag specification on income.

3. For the simple dynamic model with AR(1) disturbances given in (6.18),

   (a) Verify that $\text{plim}(\widehat{\beta}_{OLS} - \beta) = \rho(1-\beta^2)/(1+\rho\beta)$. **Hint:** From (6.18), $Y_{t-1} = \beta Y_{t-2} + \nu_{t-1}$ and $\rho Y_{t-1} = \rho\beta Y_{t-2} + \rho\nu_{t-1}$. Subtracting this last equation from (6.18) and re-arranging terms, one gets $Y_t = (\beta+\rho)Y_{t-1} - \rho\beta Y_{t-2} + \epsilon_t$. Multiply both sides by Y_{t-1} and sum $\sum_{t=2}^T Y_t Y_{t-1} = (\beta+\rho)\sum_{t=2}^T Y_{t-1}^2 - \rho\beta\sum_{t=2}^T Y_{t-1}Y_{t-2} + \sum_{t=2}^T Y_{t-1}\epsilon_t$. Now divide by $\sum_{t=2}^T Y_{t-1}^2$ and take probability limits. See Griliches (1961).

   (b) For various values of $|\rho| < 1$ and $|\beta| < 1$, tabulate the asymptotic bias computed in part (a).

   (c) Verify that $\text{plim}(\widehat{\rho} - \rho) = -\rho(1-\beta^2)/(1+\rho\beta) = -\text{plim}(\widehat{\beta}_{OLS} - \beta)$.

(d) Using part (c), show that plim $d = 2(1-$ plim $\widehat{\rho}) = 2[1 - \dfrac{\beta\rho(\beta+\rho)}{1+\beta\rho}]$ where $d = \sum_{t=2}^{T}(\widehat{\nu}_t - \widehat{\nu}_{t-1})^2/\sum_{t=1}^{T}\widehat{\nu}_t^2$ denotes the Durbin-Watson statistic.

(e) Knowing the true disturbances, the Durbin-Watson statistic would be $d^* = \sum_{t=2}^{T}(\nu_t - \nu_{t-1})^2/\sum_{t=1}^{T}\nu_t^2$ and its plim $d^* = 2(1-\rho)$. Using part (d), show that plim $(d - d^*) = \dfrac{2\rho(1-\beta^2)}{1+\beta\rho} = 2\text{plim}(\widehat{\beta}_{OLS} - \beta)$ obtained in part (a). See Nerlove and Wallis (1966). For various values of $|\rho| < 1$ and $|\beta| < 1$, tabulate d^* and d and the asymptotic bias in part (d).

4. For the simple dynamic model given in (6.18), let the disturbances follow an MA(1) process $\nu_t = \epsilon_t + \theta\epsilon_{t-1}$ with $\epsilon_t \sim \text{IIN}(0, \sigma_\epsilon^2)$.

   (a) Show that $\text{plim}(\widehat{\beta}_{OLS} - \beta) = \dfrac{\delta(1-\beta^2)}{1+2\beta\delta}$ where $\delta = \theta/(1+\theta^2)$.

   (b) Tabulate this asymptotic bias for various values of $|\beta| < 1$ and $0 < \theta < 1$.

   (c) Show that $\text{plim}(\dfrac{1}{T}\sum_{t=2}^{T}\widehat{\nu}_t^2) = \sigma_\epsilon^2[1 + \theta(\theta - \theta^*)]$ where $\theta^* = \delta(1-\beta^2)/(1+2\beta\delta)$ and $\widehat{\nu}_t = Y_t - \widehat{\beta}_{OLS}Y_{t-1}$.

5. Consider the lagged dependent variable model given in (6.20). Using the Consumption-Income data from the Economic Report of the President over the period 1950–1993 which is given in Table 5.3.

   (a) Test for first-order serial correlation in the disturbances using Durbin's h given in (6.19).

   (b) Test for first-order serial correlation in the disturbances using the Breusch (1978) and Godfrey (1978) test.

   (c) Test for second-order serial correlation in the disturbances.

6. Using the U.S. gasoline data in Chapter 4, problem 15 given in Table 4.2 and obtained from the USGAS.ASC file, estimate the following two models:

$$\text{Static: } \log\left(\frac{QMG}{CAR}\right)_t = \gamma_1 + \gamma_2\log\left(\frac{RGNP}{POP}\right)_t + \gamma_3\log\left(\frac{CAR}{POP}\right)_t + \gamma_4\log\left(\frac{PMG}{PGNP}\right)_t + \epsilon_t$$

$$\text{Dynamic: } \log\left(\frac{QMG}{CAR}\right)_t = \gamma_1 + \gamma_2\log\left(\frac{RGNP}{POP}\right)_t + \gamma_3\left(\frac{CAR}{POP}\right)_t + \gamma_4\log\left(\frac{PMG}{PGNP}\right)_t + \lambda\log\left(\frac{QMG}{CAR}\right)_{t-1} + \epsilon_t$$

   (a) Compare the implied short-run and long-run elasticities for price (PMG) and income ($RGNP$).

   (b) Compute the elasticities after $3, 5$ and 7 years. Do these lags seem plausible?

   (c) Can you apply the Durbin-Watson test for serial correlation to the dynamic version of this model? Perform Durbin's h-test for the dynamic gasoline model. Also, the Breusch-Godfrey test for first-order serial correlation.

7. Using the U.S. gasoline data in Chapter 4, problem 15, given in Table 4.2 estimate the following model with a six year lag on prices:

$$\log\left(\frac{QMG}{CAR}\right)_t = \gamma_1 + \gamma_2\log\left(\frac{RGNP}{POP}\right)_t + \gamma_3\log\left(\frac{CAR}{POP}\right)_t + \gamma_4\sum_{i=0}^{6}w_i\log\left(\frac{PMG}{PGNP}\right)_{t-i}$$

(a) Report the unrestricted OLS estimates.

(b) Now, estimate a second degree polynomial lag for the same model. Compare the results with part (a) and explain why you got such different results.

(c) Re-estimate part (b) comparing the six year lag to a four year, and eight year lag. Which one would you pick?

(d) For the six year lag model, does a third degree polynomial give a better fit?

(e) For the model outlined in part (b), reestimate with a far end point constraint. Now, reestimate with only a near end point constraint. Are such restrictions justified in this case?

References

This chapter is based on the material in Maddala (1977), Johnston (1984), Kelejian and Oates (1989) and Davidson and MacKinnon (1993). Additional references on the material in this chapter include:

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in B. Petrov and F. Csake, eds. *2nd. International Symposium on Information Theory*, Budapest: Akademiai Kiado.

Almon, S. (1965), "The Distributed Lag Between Capital Appropriations and Net Expenditures," *Econometrica*, 30: 407–423.

Breusch, T.S. (1978), "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17: 334–355.

Davidson, J.E.H., D.F. Hendry, F. Srba and S. Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal*, 88: 661–692.

Dhrymes, P.J. (1971), *Distributed Lags: Problems of Estimation and Formulation* (Holden-Day: San Francisco).

Durbin, J. (1970), "Testing for Serial Correlation in Least Squares Regression when Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38: 410–421.

Fomby, T.B. and D.K. Guilkey (1983), "An Examination of Two-Step Estimators for Models with Lagged Dependent and Autocorrelated Errors," *Journal of Econometrics*, 22: 291–300.

Gilbert, C.L. (1986), "Professor Hendry's Econometric Methodology," *Oxford Bulletin of Economics and Statistics*, 48: 283–307.

Godfrey, L.G. (1978), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables," *Econometrica*, 46: 1293–1302.

Griliches, Z. (1961), "A Note on Serial Correlation Bias in Estimates of Distributed Lags," *Econometrica*, 29: 65–73.

Hatanaka, M. (1974), "An Efficient Two-Step Estimator for the Dynamic Adjustment Model with Autocorrelated Errors," *Journal of Econometrics*, 2: 199–220.

Jorgenson, D.W. (1966), "Rational Distributed Lag Functions," *Econometrica*, 34: 135–149.

Kiviet, J.F. (1986), "On The Vigor of Some Misspecification Tests for Modelling Dynamic Relationships," *Review of Economic Studies*, 53: 241–262.

Klein, L.R. (1958), "The Estimation of Distributed Lags," *Econometrica*, 26: 553–565.

Koyck, L.M. (1954), *Distributed Lags and Investment Analysis* (North-Holland: Amsterdam).

Maddala, G.S. and A.S. Rao (1971), "Maximum Likelihood Estimation of Solow's and Jorgenson's Distributed Lag Models," *Review of Economics and Statistics*, 53: 80–88.

Nerlove, M. and K.F. Wallis (1967), "Use of the Durbin-Watson Statistic in Inappropriate Situations," *Econometrica*, 34: 235–238.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6: 461–464.

Schmidt, P. (1974), "An Argument for the Usefulness of the Gamma Distributed Lag Model," *International Economic Review*, 15: 246–250.

Schmidt, P. (1975), "The Small Sample Effects of Various Treatments of Truncation Remainders on the Estimation of Distributed Lag Models," *Review of Economics and Statistics*, 57: 387–389.

Schmidt, P. and R. N. Waud (1973), "The Almon lag Technique and the Monetary versus Fiscal Policy Debate," *Journal of the American Statistical Association*, 68: 11–19.

Solow, R.M. (1960), "On a Family of Lag Distributions," *Econometrica*, 28: 393–406.

Wallace, T.D. (1972), "Weaker Criteria and Tests for Linear Restrictions in Regression," *Econometrica*, 40: 689–698.

Wallis, K.F. (1967), "Lagged Dependent Variables and Serially Correlated Errors: A Reappraisal of Three-Pass Least Squares, " *Review of Economics and Statistics*, 49: 555–567.

Zellner, A. and M.Geisel (1970), "Analysis of Distributed Lag Models with Application to Consumption Function Estimation," *Econometrica*, 38: 865–888.