

CHAPTER 3

Distributed-Lag Models

A *distributed-lag model* is a dynamic model in which the effect of a regressor x on y occurs over time rather than all at once. In the simple case of one explanatory variable and a linear relationship, we can write the model as

$$y_t = \alpha + \beta(L)x_t + u_t = \alpha + \sum_{s=0}^{\infty} \beta_s x_{t-s} + u_t, \quad (3.1)$$

where u_t is a stationary error term.¹ This form is very similar to the infinite-moving-average representation of an ARMA process, except that the lag polynomial on the right-hand side is applied to the explanatory variable x rather than to a white-noise process ε . The individual coefficients β_s are called *lag weights* and the collectively comprise the *lag distribution*. They define the pattern of how x affects y over time.

We cannot, of course, estimate an infinite number of β coefficients in (3.1). One practical method is to truncate the lag in (3.1) to finite length q , which is appropriate if the lag distribution is effectively zero beyond q periods. Another approach is to use a functional form that allows the lag distribution in (3.1) to decay gradually to zero. We saw above that a stationary autoregressive process can be expressed as an infinite moving average with declining lag weights, so a form with one or more lags of y on the right-hand side will allow infinite-length lag distributions while requiring estimation of only a small number of parameters.

One difficulty that is common to all distributed-lag models is choice of lag length, whether this be choosing the point q at which to truncate a finite lag distribution in (3.1) or choosing how many lagged dependent variables to include. We defer this question until later in the chapter, after various distributed-lag models have been introduced.

3.1 Dynamic effects of temporary and permanent changes

In cross-sectional models, we often used econometric methods to estimate the *marginal effect* of an independent variable x on the dependent variable y , holding all of the other independent variables constant: $\partial y / \partial x$. In time-series models, we must consider not only how

¹ We assume that y_t does not depend on future values of x , thus we exclude negative values of s from the summation. However, it is theoretically possible to have “negative lags” on the right-hand side, for example, people might change their behavior now if they know that a law is going to change in the future. Equation (3.1) can be modified appropriately for such circumstances.

much effect x has on y , but when it has the effect. Is the effect immediate? Does it emerge slowly? Is there an initial effect that goes away after a few periods? In order to answer these questions, we must estimate the lag distribution relating y to x .

Instead of a single expression $\partial y / \partial x$, the effects of x on y in a dynamic model are a sequence of expressions, $\partial y_{t+s} / \partial x_t$, $s = 0, 1, 2, \dots$, that describe the effect of a change in x in period t on the value of y in periods $t, t + 1, t + 2$, etc. When the model is written in the form of equation (3.1), the dynamic effects correspond to the lag weights of the (possibly) infinite moving-average representation:

$$\frac{\partial y_{t+s}}{\partial x_t} = \frac{\partial y_t}{\partial x_{t-s}} = \beta_s. \quad (3.2)$$

Note that the first equation in (3.2) requires that the time-series relationship between y and x be stationary, so we can think of β_s either as the effect of current x_t on future y_{t+s} or as the effect of past x_{t-s} on current y_t .

When reporting the results of a lag regression it is common to express the lag weights either in a table, on a graph, or both. Suppose that we estimate a finite distributed lag with weights of 4, 2, and 0.5. For this example, equation (3.1) becomes

$$y_t = \alpha + 4x_t + 2x_{t-1} + 0.5x_{t-2} + u_t. \quad (3.3)$$

We might show the lag weights for equation (3.3) in a graph similar to Figure 3-1.

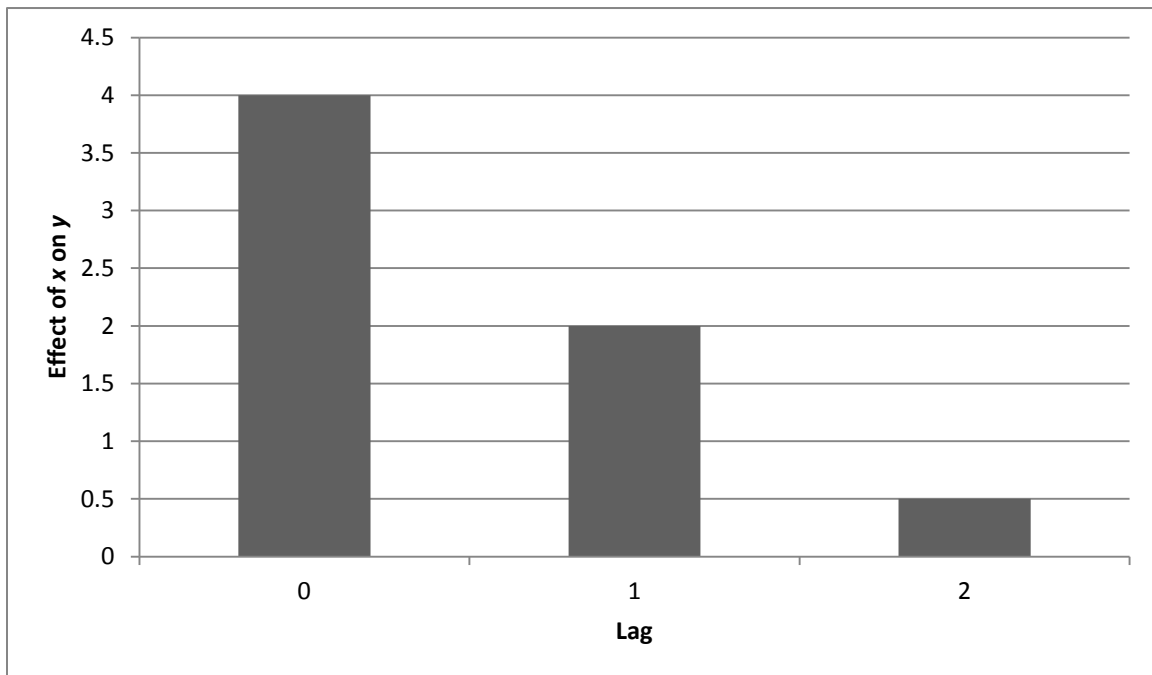


Figure 3-1. Graph of lag distribution

To see the interpretation of the lag weights, we consider two special cases: a temporary change in x and a permanent change in x . Suppose that x increases temporarily by one unit in period t , then returns to its original lower level for periods $t + 1$ and all future periods. For the temporary change, the time path of the changes in x looks like Figure 3-2: the change is zero except in period t , where it is one.

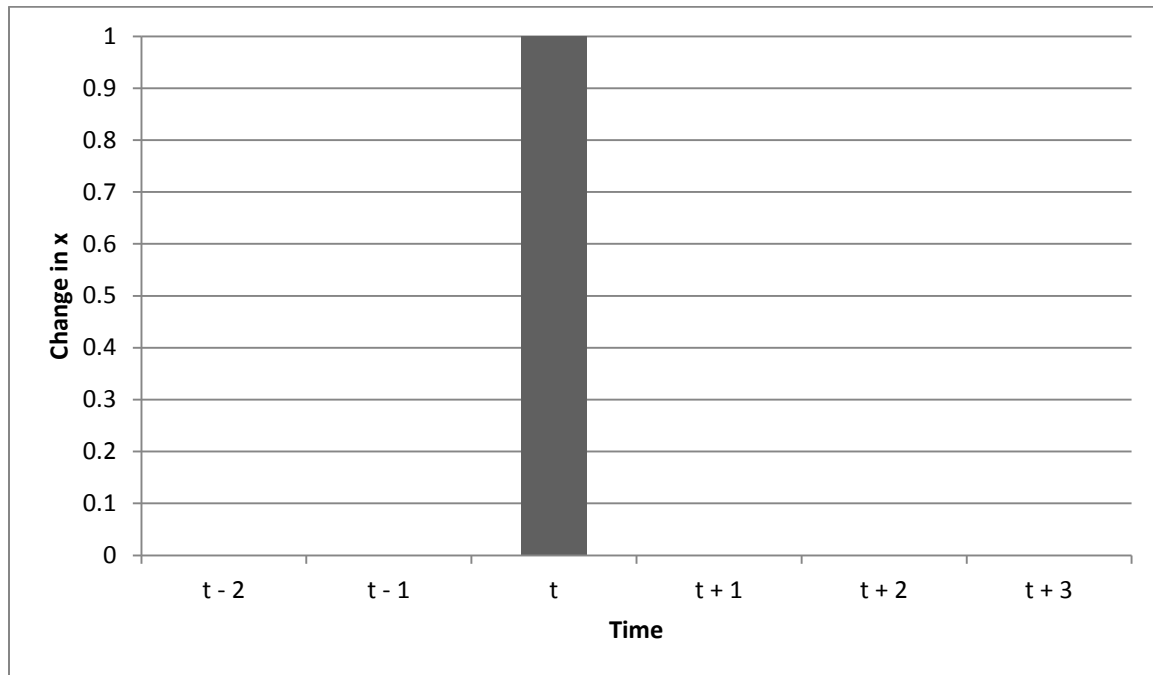


Figure 3-2. Temporary change in x

In period t , x_t is increased by one unit but all earlier x values are unchanged. From equation (3.3), we see that y_t increases by 4 when x_t goes up by one with x_{t-1} , x_{t-2} , and the disturbance term u_t unchanged. We call this immediate effect $\partial y_t / \partial x_t = \beta_0$ the **impact effect of x on y** .

In period $t + 1$, equation (3.3) is advanced one period to become

$$y_{t+1} = \alpha + 4x_{t+1} + 2x_t + 0.5x_{t-1} + u_{t+1}. \quad (3.4)$$

Given that x_{t+1} , x_{t-1} , and the disturbance are unchanged, the change in y in period $t + 1$ is 2, the coefficient on the first lag of x (β_1). This is the **dynamic marginal effect of x on y at one lag**. By similar analysis, we can see that the effect of the temporary change in x at time t on y_{t+2} is $\beta_2 = 0.5$.

From this example we can see that the pattern of dynamic marginal effects of a temporary change in x on y is given by the coefficients of the lag distribution β_s , that are shown in Figure 3-1.

Now consider the case of a permanent increase in x at time t : x increases by one unit in period t and remains higher in all periods after t than it was before t . This change is graphed in Figure 3-3.

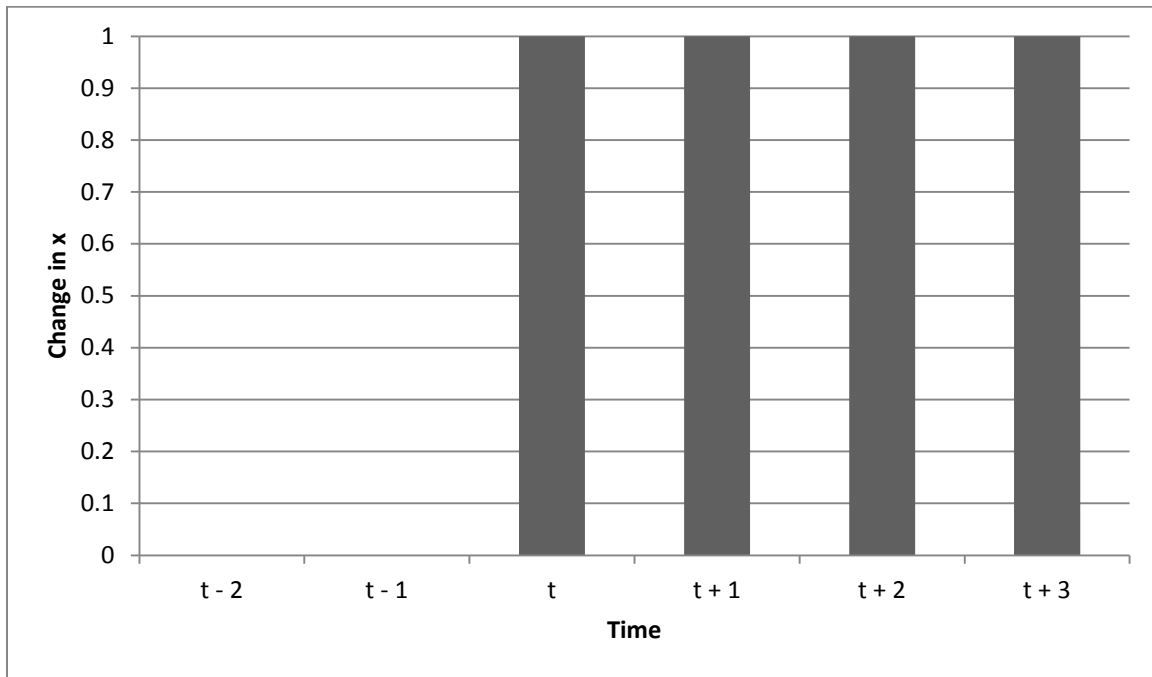


Figure 3-3. Permanent change in x

Once again, from equation (3.3) we see that the effect of the change in x_t on y_t is 4, because x_{t-1} , x_{t-2} , and u_t are assumed unchanged. Moving forward to period $t + 1$, the analysis of equation (3.4) is different from the temporary change because now x_{t+1} as well as x_t is increased by one unit. The cumulative effect on y_{t+1} is $4 + 2 = \beta_0 + \beta_1 = 6$. We call this the **cumulate effect of x on y at one lag**. At two lags, we advance equation (3.4) one more period forward and now see the effects of the increased levels of x_{t+2} , x_{t+1} , and x_t raising y_{t+2} by $4 + 2 + 0.5 = 6.5$ units.

Moving one more period into the future, the effect on y_{t+3} will be the same as the effect on y_{t+2} because once again all of the x terms on the right-hand side are increased by one unit. The time path of the cumulative effects is shown in Figure 3-4, with the cumulative effect staying at 6.5 for all lags starting at 2. The limit of the cumulative effect as the lag length goes to infinity as called the **long-run cumulative effect of x on y** . It measures how much y will eventually change in response to a permanent change in x .

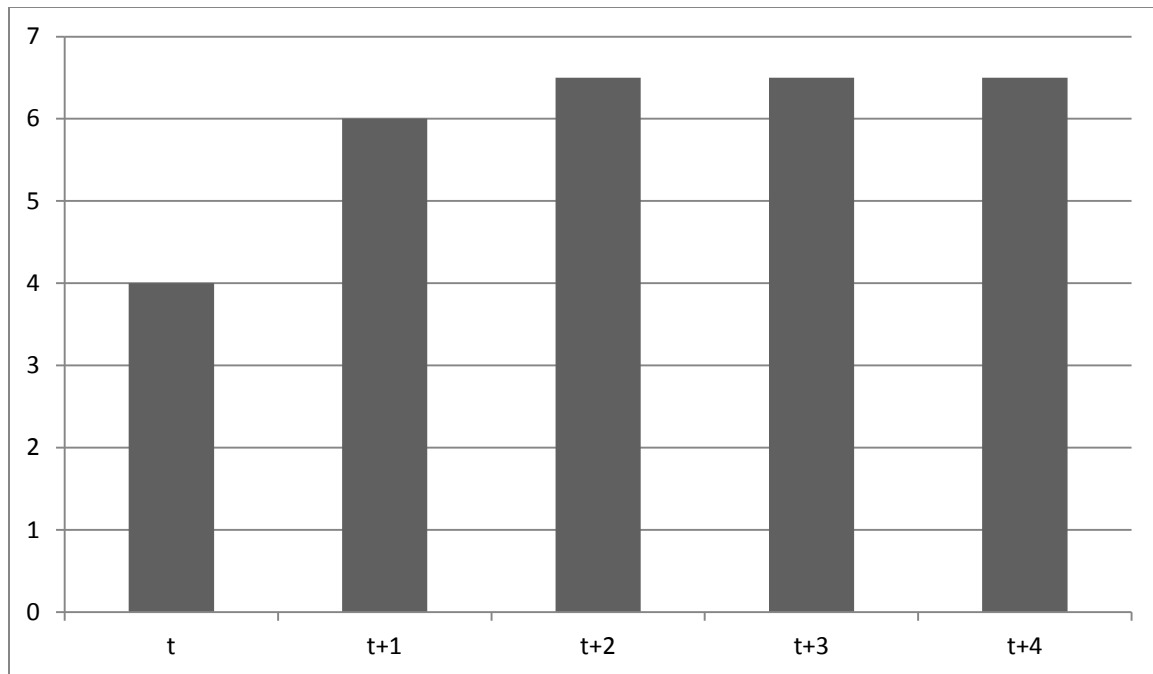


Figure 3-4. Cumulative effects of x on y

For a distributed lag with a finite moving-average representation of length q , equation (3.1) can be written as

$$y_t = \alpha + \sum_{s=0}^q \beta_s x_{t-s} + u_t. \quad (3.5)$$

In this case, the long-run cumulative effect is $\sum_{s=0}^q \beta_s$. If the moving-average representation converges to zero slowly as s goes to infinity rather than truncating at finite q , then the long-run cumulative effect is $\lim_{q \rightarrow \infty} \sum_{s=0}^q \beta_s$ or $\sum_{s=0}^{\infty} \beta_s$.

The pattern of the dynamic marginal effects and cumulative effects tells us about both the magnitude and the timing of the effect of x on y . In the example we studied above, permanent increases in x lead to permanent increases in y that get larger over the first three periods of the change. Temporary changes in x , by contrast, lead to temporary changes in y that die away after three periods.

Another pattern that is plausible for some economic relationships is that permanent changes in x may lead to only temporary changes in y . For example, standard macroeconomic theory tells us that changes in the rate of monetary growth have only temporary effects on real output growth. In such a situation, the positive marginal effects at short lags (β_0 , β_1 , and β_2 , perhaps) would be offset by negative marginal effects at longer lags so that the long-run cumulative effect (the sum of all the β coefficients) is zero.

Is it possible to have a dynamic relationship in which temporary changes in x lead to permanent changes in y ? It is possible, but this creates difficulty for an equation such as (3.1). For a temporary change in x to have a permanent effect on y , the β coefficients in (3.1) could not approach zero as s gets large. This obviously rules out any finite lag distribution, and indeed creates trouble for any specification because the dynamic relationship between the two variables is effectively non-stationary. To cope with this situation, it is usually best to redefine either x or y to eliminate the problem.

For example, suppose that we are looking at the effect of changing the legal drinking age on traffic fatalities. If we were to model this relationship with x being the change from year to year in the drinking age and y being fatalities, then a once-and-for-all increase in the drinking age leads to a positive value of x in the year of the increase and $x = 0$ in subsequent years when the drinking age stays high. We might expect a permanent reduction in traffic fatalities to result, which would mean negative β values extending into the indefinite future.

We can solve this problem by redefining x to be the legal drinking age itself, rather than the change in the drinking age. If we do this, then the once-and-for-all increase leads to a permanently higher value of x . We saw above that permanent changes in x may have a permanent effect on y even if the β coefficients eventually converge to zero.

Another way of resolving this difficulty is by redefining y to be the *change* in traffic fatalities leaving x as the change in drinking age. In this case, the temporary change in x associated with the rise in the drinking age would be associated with a temporary (though possibly lagged) change in y measuring the change in traffic deaths. We would expect the increases in fatalities eventually to return to zero after the one-time change in drinking age, so again the β coefficients would converge to zero over time.

In specifying dynamic econometric models, it is crucial to think very carefully about the nature of the dynamic relationship among the variables. We must decide how we would expect y to respond over time to a one-time change in x , then define the variables as levels or changes in order to represent the expected relationship with a stationary lag function.

3.2 Finite distributed lag models: Estimation and interpretation

3.2.1 Unrestricted distributed lags

In cases where the effect of x on y dies out quickly, it may be feasible to estimate equation (3.5) directly. The finite distributed lag model has several advantages. The coefficients can be estimated by OLS or GLS, assuming that x is strictly exogenous. Interpretation of the β_s coefficients is straightforward using the analysis above. Since there are no restrictions imposed on the $q + 1$ lag coefficients, any finite pattern of lag weights can be estimated.

There are two disadvantages to the finite distributed lag model. The first is multicollinearity. Even if x is stationary, it may be highly autocorrelated, meaning that x_t and x_{t-1} are strongly correlated, as are x_{t-1} and x_{t-2} , x_{t-2} and x_{t-3} , etc. High levels of correlation among the regressors imply multicollinearity, which leads to unreliable coefficient estimates with large variances and standard errors. Estimation of finite distributed lag models with strongly autocorrelated regressors often leads to lag distributions in which the sequence of lag coefficients bounces around between large and small—and sometimes positive and negative—values in ways that are not consistent with economic theory. When this happens, econometricians often try to restrict the estimated β coefficients to satisfy prior assumptions of smoothness.

The second disadvantage of finite distributed lags is that they can be problematic when the lag length is long, especially in small samples. If we have data for observations from $t = 1$ through T , then the earliest observation that can be included in the estimating sample is $t = q + 1$, because we need to have data for q periods before the beginning of the estimating sample for the lagged terms on the right-hand side. Thus, we have $T - q$ observations available to estimate $q + 2$ coefficients (including the constant, assuming that there is only one regressor). This affords us $T - 2q - 2$ degrees of freedom. Each time we lengthen the lag by one period, we lose two degrees of freedom—one because we must estimate another coefficient and one because we must reduce our sample by one period. Unless T is very large compared to plausible values of q , degrees of freedom can be depleted very quickly as the lag length increases. Note also that this effect is magnified if there are two or more regressors for which we must estimate lag distributions.

In summary, the finite distributed lag model is most suitable to estimating dynamic relationships when lag weights decline to zero relatively quickly, when the regressor is not highly autocorrelated, and when the sample is long relative to the length of the lag distribution.

3.2.2 Restricted finite distributed lag models

We may believe strongly (perhaps based on economic theory) that the lag weights β_s should be a smooth function of s . If the unrestricted finite distributed lag estimates contradict this smoothness, we may choose to restrict the model to impose smooth lag weights. Restricting the lag coefficients can not only impose smoothness, but also reduces the number of parameters that must be estimated. There are many patterns of smoothness that we can choose to impose a restrictive structure on the weights. We shall discuss a simple example in detail, then consider other possible restrictions.

One simple restriction on the lag weights is that they decline linearly from an initial positive or negative impact effect to zero at a lag of length $q + 1$. [Might want a graph here.] In other words, each of the lag weights $\beta_1, \beta_2, \dots, \beta_q$ are linearly declining fractions of the impact effect β_0 according to Table 3-1. Each lag weight β_s is smaller than its predecessor β_{s-1} by the fixed amount $1/(1 + q)$ until the effect dies to zero at $s = q + 1$. The formula for the lag weights in Table 3-1 is

$$\beta_s = \frac{q+1-s}{q+1} \beta_0, \quad s=1, 2, \dots, q. \quad (3.6)$$

Table 3-1. Linearly declining lag weights

Lag s	Lag weight β_s
0	β_0
1	$q/(q+1) \beta_0$
2	$(q-1)/(q+1) \beta_0$
...	...
$q-1$	$2/(q+1) \beta_0$
q	$1/(q+1) \beta_0$
$q+1$	0

To estimate the linear-declining lag model for given q , we substitute for each β_s from equation (3.6) into equation (3.5) to get

$$y_t = \alpha + \sum_{s=0}^q \frac{q+1-s}{q+1} \beta_0 x_{t-s} + u_t = \alpha + \beta_0 \left[\sum_{s=0}^q \frac{q+1-s}{q+1} x_{t-s} \right] + u_t. \quad (3.7)$$

Equation (3.7) has two parameters to be estimated, α and β_0 , with the single regressor in the model being the bracketed term in (3.7). The remainder of the lag distribution is determined by the choice of q and the linear restriction. Estimation of (3.7) will have only $T-q$ observations because we still need to exclude the first q data points to construct the weighted-sum lag variable, but it has only 2 parameters to be estimating, giving $T-q-2$ degrees of freedom. Relative to unrestricted estimation, we have saved q degrees of freedom.

The impact effect is β_0 in this model, and the long-run cumulative effect is

$$\begin{aligned} \beta_0 \sum_{s=0}^q \frac{q+1-s}{q+1} &= \frac{\beta_0}{q+1} \sum_{s=0}^q (s+1) \\ &= \frac{\beta_0}{q+1} \left[(q+1) + \sum_{s=0}^q s \right] \\ &= \beta_0 \left(1 + \frac{q(q+1)/2}{q+1} \right) \\ &= \beta_0 \left(1 + \frac{q}{2} \right). \end{aligned}$$

For $q = 4$, the lag weights would be

s	β_s
0	β_0
1	$(4/5) \beta_0$
2	$(3/5) \beta_0$
3	$(2/5) \beta_0$
4	$(1/5) \beta_0$
Long-run cumulative effect	$3\beta_0$

The same procedure we used to estimate the linear-declining lag model can be applied to other shapes as well. For example, if the lag weights increase linearly to a peak at lag m , then decline to zero in a symmetric “tent” shape, then we can model β_s as

$$\beta_s = \beta_m \left(1 - \frac{|m-s|}{m+1} \right), \quad s = 0, 1, \dots, 2m.$$

Once again, this formula can be substituted into (3.5) to get an estimating form that allows β_m to be estimated. The long-run cumulative effect in this model is $(m+1)\beta_m$. For a lag of length 6 that reaches its maximum effect at 3 lags, the lag coefficients of the tent lag would be

s	β_s
0	$(1/4) \beta_3$
1	$(1/2) \beta_3$
2	$(3/4) \beta_3$
3	β_3
4	$(3/4) \beta_3$
5	$(1/2) \beta_3$
6	$(1/4) \beta_3$
Long-run cumulative effect	$4\beta_3$

A more common application of restricted distributed lags is the *polynomial distributed lag* first explored by Shirley Almon (1965). The most common application of the polynomial distributed lag is restricting the lag coefficients to lie on a quadratic function. This imposes smoothness on the coefficients, but allows for considerable flexibility in the shapes of the lag distributions that it permits. Depending on what part of the parabola lies in the range $(0, q)$, the commonly plausible shapes shown in Figure 3-5 can be estimated. The quadratic lag distribution also allows linear lags as a special case, including flat lags and the linear-declining lags discussed earlier.

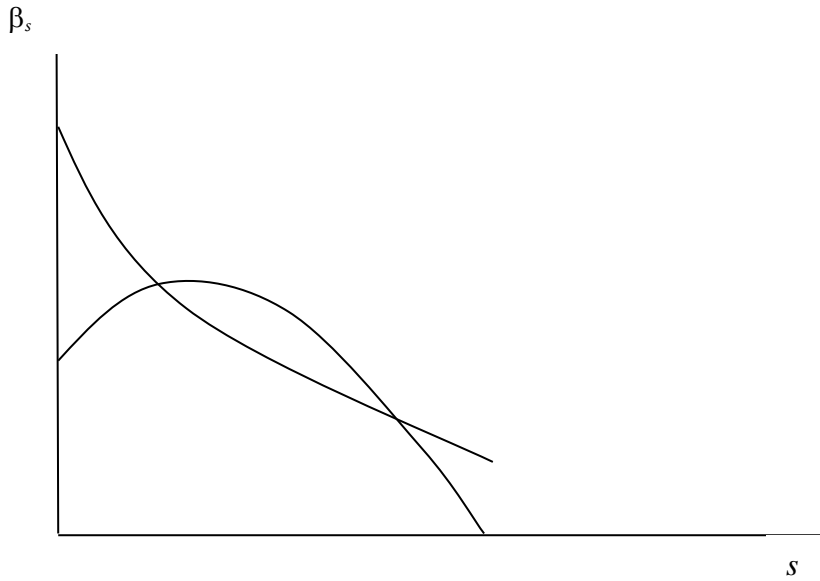


Figure 3-5. Quadratic lag distributions

A quadratic lag function restricts the lag coefficients to lie on a parabola:

$$\beta_s = \xi_0 + \xi_1 s + \xi_2 s^2, \quad s = 0, 1, \dots, q, \quad (3.8)$$

where ξ_0 , ξ_1 , and ξ_2 are the parameters of the quadratic function describing the lag weights. Substituting into (3.5) yields

$$\begin{aligned} y_t &= \alpha + \sum_{s=0}^q (\xi_0 + \xi_1 s + \xi_2 s^2) x_{t-s} + u_t \\ &= \alpha + \xi_0 \sum_{s=0}^q x_{t-s} + \xi_1 \sum_{s=0}^q s x_{t-s} + \xi_2 \sum_{s=0}^q s^2 x_{t-s} + u_t, \end{aligned}$$

or

$$y_t = \alpha + \xi_0 z_t^0 + \xi_1 z_t^1 + \xi_2 z_t^2 + u_t, \quad (3.9)$$

where

$$z_t^0 \equiv \sum_{s=0}^q x_{t-s}, \quad z_t^1 \equiv \sum_{s=0}^q s x_{t-s}, \quad z_t^2 \equiv \sum_{s=0}^q s^2 x_{t-s}.$$

The z variables can be constructed by simple transformations once you have chosen a value of q , which allows equation (3.9) to be estimated by standard linear methods. Once we have obtained estimates for the parameters of (3.9), we can obtain the implied estimates of the lag weights from (3.8). Since the functions in (3.8) relating the lag weights β to the quadratic pa-

parameters ξ are linear, calculation of the standard errors of the lag weights is straightforward using the usual procedure for linear functions of coefficients.

Of course, quadratic lag distributions can also come out looking counterintuitive, for example with lags that diverge from zero at the far endpoint or that swoop into negative values in the middle. Implausible estimated lag distributions may be evidence of misspecification of the model, so they should not be ignored. It is possible to force the quadratic lag distribution to have certain properties, such as having the weights converge smoothly to a zero value at lag $q + 1$ (as we did for the linear-declining lags). For the quadratic case, we want to restrict β_{q+1} to have the value zero, so $\beta_{q+1} = \xi_0 + \xi_1(q+1) + \xi_2(q+1)^2 = 0$. This means that the parameters of (3.9) have the linear restriction $\xi_0 = -\xi_1(q+1) - \xi_2(q+1)^2$, which can be imposed in estimation to assure that the quadratic lag distribution dies to zero smoothly at the end of the lag.

The method of polynomial distributed lags can be used with cubic or even higher-order polynomials as well. The higher the order of the polynomial, the less “smooth” the lag distribution is allowed to be and the more parameters must be estimated. Of course, just as we can always find a line passing through any two points and a parabola passing through any three points, any unrestricted lag distribution with $q + 1$ lag weights lies on a polynomial lag distribution of order q , so a q -order polynomial distributed lag is equivalent to the unrestricted distributed lag model.

The restrictions implied by a polynomial distributed lag model can always be tested as linear restrictions on the unrestricted polynomial estimates. This tells us whether the data can reject the smoothness imposed by the polynomial lag model. If non-smoothness of the unrestricted lag weights was due to multicollinearity (which makes the point estimates unreliable), then the smoothness restrictions may not be rejected. If the restrictions are rejected, then the data contain strong evidence that the lag distribution does not follow the smoothed model we have imposed, and we impose the smoothness model at our peril.

The easiest way to implement the test of the polynomial restrictions is in terms of “differences” of coefficients. For the quadratic lag, we difference (3.8) in terms of s to get

$$\begin{aligned}\beta_s - \beta_{s-1} &= (\xi_0 + \xi_1 s + \xi_2 s^2) - (\xi_0 + \xi_1(s-1) + \xi_2(s-1)^2) \\ &= \xi_1 + \xi_2 [s^2 - (s-1)^2] = \xi_1 + \xi_2 [s^2 - (s^2 - 2s + 1)] \\ &= \xi_1 + \xi_2 (2s - 1), \quad s = 1, 2, \dots, q,\end{aligned}$$

then difference again to get

$$\begin{aligned}(\beta_s - \beta_{s-1}) - (\beta_{s-1} - \beta_{s-2}) &= [\xi_1 + \xi_2 (2s - 1)] - [\xi_1 + \xi_2 (2(s-1) - 1)] \\ &= \xi_2 [2s - 1 - (2s - 2 - 1)] \\ &= 2\xi_2, \quad s = 2, 3, \dots, q.\end{aligned}$$

Thus, the second difference of the lag coefficients (differencing with respect to the lag) is a constant. That means that we can test the following $q - 2$ restrictions on $\beta_0, \beta_1, \dots, \beta_q$ to test the quadratic restriction:

$$\begin{aligned}\beta_2 - 2\beta_1 - \beta_0 &= \beta_3 - 2\beta_2 - \beta_1 \\ \beta_3 - 2\beta_2 - \beta_1 &= \beta_4 - 2\beta_3 - \beta_2 \\ &\vdots \\ \beta_{q-1} - 2\beta_{q-2} - \beta_{q-3} &= \beta_q - 2\beta_{q-1} - \beta_{q-2}.\end{aligned}$$

3.3 Models with lagged dependent variables

In an earlier chapter we introduced autoregressive-moving-average (ARMA) time-series processes as a way of modeling univariate time series. The autoregressive component of the ARMA model involves using one or more lagged values of y as determinants of the current value y_t . We can apply the same method in a distributed-lag context by adding y_{t-1} and possibly additional lags to the right-hand side. The simplest model is the *Koyck lag*, which has one lag of y on the right-hand side with only the current value of x . Adding lagged values of x in addition to lagged y variables leads to the *rational lag* model.

3.3.1 Koyck lag

The first-order autoregressive lag model is often called the Koyck lag in recognition of the seminal application of the model to the macroeconomic investment function by L. M. Koyck (1954). With a single explanatory variable x , the model is written

$$y_t = \delta + \phi_1 y_{t-1} + \theta_0 x_t + u_t. \quad (3.10)$$

Estimation of equation (3.10) presents challenges because y_{t-1} is by definition not strictly exogenous and, unless u_t is white noise, is not even weakly exogenous. We postpone these estimation concerns for the moment and discuss the interpretation of the coefficients in the Koyck model.

In terms of the lag operator, we can write (3.10) as

$$(1 - \phi_1 L)y_t = \delta + \theta_0 x_t + u_t.$$

This suggests solving the model for y_t as

$$y_t = \frac{\delta}{1 - \phi_1 L} + \frac{\theta_0}{1 - \phi_1 L} x_t + \frac{1}{1 - \phi_1 L} u_t,$$

which, by methods analogous to those used to examine autoregressive processes, can be written in an infinite-distributed-lag form as

$$y_t = \frac{\delta}{1 - \phi_1} + \theta_0 \sum_{s=0}^{\infty} \phi_1^s x_{t-s} + \sum_{s=0}^{\infty} \phi_1^s u_{t-s}, \quad (3.11)$$

as long as $|\phi_1| < 1$.

Equation (3.11) has the form of the infinite distributed lag (3.1), with

$$\alpha = \frac{\delta}{1 - \phi_1},$$

$$\beta_s = \theta_0 \phi_1^s,$$

and the disturbance term having an infinite-moving-average form. The dynamic marginal effects of x on y in the Koyck model are

$$\frac{\partial y_{t+s}}{\partial x_t} = \beta_s = \theta_0 \phi_1^s. \quad (3.12)$$

If, as is usually the case, $0 < \phi_1 < 1$, the lag weights decline exponentially to zero from an initial value of θ_0 . The long-run cumulative effect of x on y is

$$\theta_0 \sum_{s=0}^{\infty} \phi_1^s = \frac{\theta_0}{1 - \phi_1}.$$

This exponentially declining lag distribution seems to fit many economic relationships well. Moreover, some theoretical models, such as exponential convergence models in economic growth and models with quadratic adjustment costs, predict exponentially declining lag weights.

The Koyck lag model can be used with more than one regressor in the equation, but it imposes a significant restriction on the lag distributions. Suppose that we have two regressors, x and z :

$$y_t = \delta + \phi_1 y_{t-1} + \theta_0 x_t + \lambda_0 z_t + u_t.$$

The dynamic marginal effects of x on y will be as in equation (3.12). The effects of z on y will be

$$\frac{\partial y_{t+s}}{\partial z_t} = \lambda_0 \phi_1^s.$$

Both the effects of x and the effects of z decline at the same exponential rate ϕ_1 as s increases, so with the Koyck lag it is not possible for y to respond more quickly to one explanatory variable than to another. This symmetry of dynamic responses may be appropriate in some applications. For example, if the reason for lagged response is costs of adjusting y to its optimal level, then it may be reasonable to expect that the same adjustment costs—and thus the same

lag structure—would apply regardless of which variable caused a change in the optimal y . In other cases, we might expect different patterns of dynamic response to different regressors. Consumption of asparagus may adjust more (or less) quickly to changes in the price of asparagus than to changes in income. The Koyck lag would not allow independent estimation of the lag structures of the two variables.²

The theoretical and empirical appeal of the Koyck lag has led to its frequent use. However, consistent estimation of the Koyck-lag model can be problematic. The lagged dependent variable as a regressor on the right-hand side is never strictly exogenous, so the small-sample assumptions needed for the Gauss-Markov Theorem cannot be satisfied.

Even the weak exogeneity required for consistency is often dubious. Lagging the model one period means that $y_{t-1} = \delta + \phi_1 y_{t-2} + \theta_0 x_t + u_{t-1}$, so y_{t-1} is clearly correlated with u_{t-1} . If u_t is not white noise so $\text{cov}(u_t, u_{t-1}) \neq 0$, so is not white noise, then y_{t-1} and u_t are correlated and the lagged dependent variable is not even weakly exogenous and the OLS estimator is inconsistent. Given the ubiquity of serial correlation in error terms of dynamic models, it is hard to maintain confidence in the consistent estimation of the Koyck model.

Avoiding inconsistency requires that we make sure that the error term is not serially correlated. GLS methods can be used to transform the model into one that is not serially correlated, but only with great caution. Because the OLS estimators are inconsistent, so are all test statistics and estimators based on the OLS residuals. For example, the Durbin-Watson test and other tests that rely on the correlation coefficient between \hat{u}_t and \hat{u}_{t-1} are not valid. Moreover, we must estimate the disturbance autocorrelation parameter using a method that does not rely on OLS residuals. Thus, the Hildreth-Lu search procedure is valid for estimating a model with an AR(1) error and a lagged dependent variable, but the Prais-Winsten and Cochrane-Orcutt methods are not.

3.3.2 Longer autoregressive lags

The Koyck lag treats y as a first-order autoregressive process. Although one lag of the dependent variable is often enough to capture the dynamic relationship between y and the regressors. However, longer autoregressive lags can be included as well. The general autoregressive lag model AR(p) would be written

$$\phi(L)y_t = \delta + \theta_0 x_t + u_t, \quad (3.13)$$

with $\phi(L)$ a p -order polynomial in the lag operator. In order for the relationship between y and x to be dynamically stable, the roots of $\phi(L)$ must lie outside the unit circle. This generalizes the condition $|\phi_1| < 1$ from the Koyck lag model. If the stability condition does not

² We shall see below that adding one or more lags of x and/or z could allow differences in the first few terms of the lag distribution. However, the “tail” of the lag distribution beyond the longest lag of x or z always depends only on ϕ_1 .

hold, then a one-time change in x will cause permanent or explosive changes in y , which suggests differencing y to make the order of integration the same on both sides of the equation.

The ϕ parameters of the autoregressive lags determine the shape of the lag distribution. As with the first-order Koyck lag, lag weights can decline smoothly according to an exponential pattern. But with higher-order lags, other patterns are possible. For example, if the roots of $\phi(L)$ are complex numbers (i.e., have non-zero imaginary terms), then lag weights may oscillate back and forth like a damped sine function, converging eventually to zero. (Complex roots on the unit circle would have non-damped oscillations like a pure sine wave; complex roots inside the unit circle would lead to explosive oscillations.)

For example, suppose that the model is $y_t = 0.8y_{t-1} - 0.4y_{t-2} + x_t + u_t$. The autoregressive polynomial $\phi(L) = 1 - 0.8L + 0.4L^2$, whose roots are $1 \pm 1.225i$. These roots lie outside the unit circle because $\sqrt{1^2 + 1.225^2} > 1$, so the lag distribution is stable and converges to zero as the lag length increases. Evaluating the lag distribution gives the pattern of oscillating convergence of lag weights shown in Figure 3-6. A one-unit temporary increase in x leads to an initial increase in y . But by the third, fourth, and fifth periods the effect on y has become slightly negative. The lag weights subsequently approach zero quickly, oscillating between positive and negative values every three to four periods.

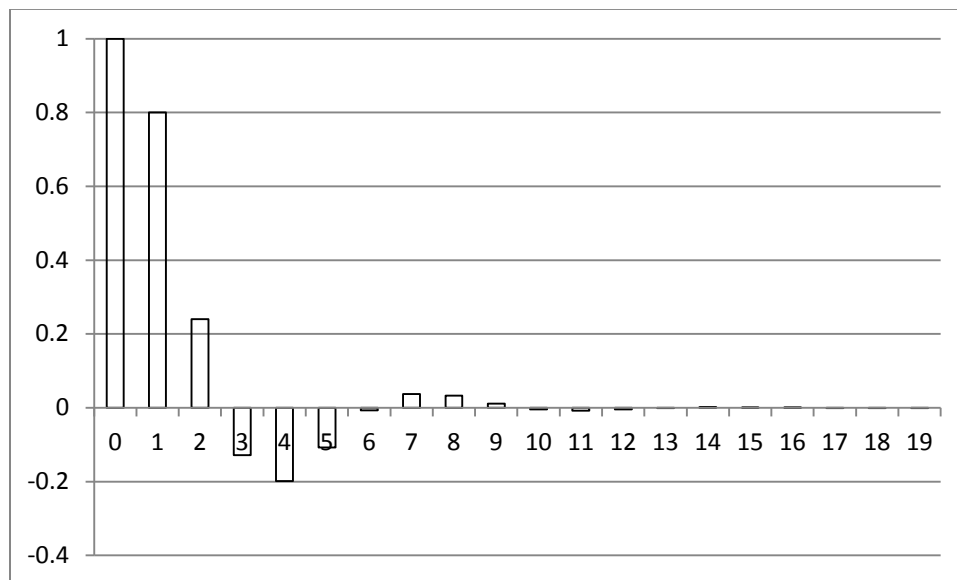


Figure 3-6. Oscillating convergence in a second-order AR lag

The oscillating pattern of Figure 3-6 may not be realistic for most economic applications, but there are situations where theory predicts a non-monotonic response with both positive and negative lag weights. For example, in a simple real-business-cycle model, productivity shocks may first increase the marginal return to capital (due to higher productivity), then

lower it (due to increases in the size of the capital stock), with the effect converging to zero in the long run.³

3.3.3 Autoregressive distributed lag (ARDL) models

Given the autoregressive lag relationship in equation (3.13), a logical extension is to allow lags of x on the right-hand side. The general *autoregressive distributed lag (ARDL) model* is written

$$\phi(L)y_t = \delta + \theta(L)x_t + u_t, \quad (3.14)$$

where $\phi(L)$ is an order- p polynomial that, for stability, has roots lying outside the unit circle and $\theta(L)$ is an order- q polynomial. Expanding the lag polynomials, equation (3.14) can be written as

$$y_t = \delta + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_0 x_t + \theta_1 x_{t-1} + \dots + \theta_q x_{t-q} + u_t.$$

With a sample of T observations, this model can be estimated for $T - \max\{p, q\}$ observations.

As we did with univariate ARMA models, we can divide both sides of (3.14) by the autoregressive polynomial to get

$$\begin{aligned} y_t &= \frac{\delta}{\phi(L)} + \frac{\theta(L)}{\phi(L)} x_t + \frac{1}{\phi(L)} u_t \\ &= \alpha + \frac{\theta(L)}{\phi(L)} x_t + v_t, \end{aligned} \quad (3.15)$$

where α and v are the constant and error term defined in (3.15). Because the lag distribution of the ARDL model can be represented by the ratio of two finite lag polynomials, it is sometimes called the *rational lag*. (Recall that rational numbers can be represented as the ratio of two integers.)

The analysis of ARDL models parallels that of univariate ARMA processes; the difference is that the lag structure on the right-hand side of (3.15) is applied to an explanatory variable x rather than to a white-noise error term ε as in equation **Error! Reference source not found.** As in the ARMA models, the coefficients of the order- p polynomial $\theta(L)$ affect only the first q lags of the dynamic lag distribution of the effect of x on y . The behavior of the “tail” of the lag distribution beyond q depends entirely on the autoregressive polynomial $\phi(L)$. The property that the dynamic effect is stable only if the roots of $\phi(L)$ lie outside the unit circle carries over from the autoregressive lag model of the previous section.

³ See Chapter 5 of Romer (2012).

3.3.4 Autocorrelated errors and distributed-lag models

It is easy to confuse models in which y is autocorrelated and models in which u is autocorrelated—and not just for students, but for statistical procedures as well. Consider the following ARDL(1,0) model:

$$\begin{aligned}y_t &= \delta + \phi_1 y_{t-1} + \theta_0 x_t + u_t, \\u_t &= \rho u_{t-1} + \varepsilon_t.\end{aligned}$$

We assume that both ϕ_1 and ρ are between -1 and 1 for stability and stationarity and that ε is white noise. Solving for u_t from the first line, $u_t = y_t - \delta - \phi_1 y_{t-1} - \theta_0 x_t$, so lagging once yields $u_{t-1} = y_{t-1} - \delta - y_{t-2} - \theta_0 x_{t-1}$. Substituting this into the second line and then plugging the resulting expression for u_t into the first line yields

$$\begin{aligned}y_t &= \delta + \phi_1 y_{t-1} + \theta_0 x_t + \rho(y_{t-1} - \delta - \phi_1 y_{t-2} - \theta_0 x_{t-1}) + \varepsilon_t \\&= (1 - \rho)\delta + (\phi_1 + \rho)y_{t-1} - \rho\phi_1 y_{t-2} + \theta_0 x_t - \rho\theta_0 x_{t-1} + \varepsilon_t.\end{aligned}\tag{3.16}$$

Equation (3.16) shows that the ARDL(1,0) model with a first-order autoregressive error can be reduced to an ARDL(2,1) model with a white-noise error. Given this apparent equivalence, how could we determine which is the appropriate model?

The equivalence is not quite perfect because there is one nonlinear restriction on the coefficients of (3.16). The ARDL(2,1) model in (3.16) has 5 linear coefficients (the constant, two lagged y coefficients, and two coefficients on current and lagged x), but they depend on only 4 underlying parameters: δ , ϕ_1 , θ_0 , and ρ . Thus, if we needed to, we could estimate the general ARDL(2,1) model (3.16) and test the nonlinear coefficient restriction:

$$\text{coef}(y_{t-1}) = \frac{\text{coef}(y_{t-2})}{\text{coef}(x_{t-1})} \text{coef}(x_t) - \frac{\text{coef}(x_{t-1})}{\text{coef}(x_t)}.\tag{3.17}$$

If the model is truly a Koyck lag with AR(1) error, then it is true that we can save one degree of freedom and thus gain a marginal bit of efficiency by imposing restriction (3.17) on the estimation of (3.16) (or by doing this implicitly through a two-step procedure such as Hildreth-Lu). However, in practice, it is much easier to estimate the unrestricted ARDL(2,1) model, which accounts (though not fully efficiently) for the possibility of an AR(1) error in the simpler Koyck model.

This example shows that lengthening the lags in the ARDL model can eliminate serial correlation in the error term. It is fairly straightforward to determine how many lags need to be included. In (3.16), if $\rho = 0$ then the coefficients of y_{t-2} and x_{t-1} are both zero, so testing the last lag terms and eliminating them if they are near zero will assure that we do not include an unneeded correction for autocorrelated disturbances. Testing the residuals for autocorrelation should reveal whether the remaining error term is white noise. If it is not, then adding more lags might be appropriate. Adding lags until the residual seems to be white

noise is the most common way that modern time-series econometricians deal with the possibility of serial correlation in the disturbance. Two-step GLS estimation using the Prais-Winsten or Hildreth-Lu procedures is used much less frequently.

3.4 Choosing the Lag Length

In all of the models we have studied, we must specify the length of the lag prior to estimation. Economic theory rarely gives us information about the lag length, so this must usually be determined empirically. Several methods are available to econometricians to gain information about the appropriate lag length, though they do not always give the same answer. There is no “right way” to identify the length of a lag. We are usually forced to make a judgment after looking at the evidence from several methods.

The methods we discuss for choosing lag length can apply either to the lagged x terms on the right-hand side—the lag length that we have been calling q —or to the number of lagged dependent variables to include in an autoregressive-lag or ARDL model—what we have called p . However, most of these methods cannot be applied in a straightforward way to determining the length of restricted lag models such as the linearly-declining lag or the polynomial distributed lag.

It is conventional in dynamic analysis that if we determine the appropriate lag length to be q , we usually include *all* lags between 0 and q . It would be very unusual to encounter an economic model in which, for example, x_{t-1} and x_{t-3} would affect y_t , but x_{t-2} would not. Therefore we would not usually omit x_{t-2} as a regressor even if its coefficient is not statistically significant. The only common exception to that rule is with seasonal data, where we might augment a short lag of 3 or 4 periods with a lag-12 coefficient when using monthly data to capture the effect of the same month in the preceding year.

3.4.1 Determining lag length by statistical significance

An obvious way to choose the length of a lag is to start with a long lag test the statistical significance of the coefficient at the longest lag—the “trailing lag”—and shorten the lag by one period if we cannot reject the null hypothesis that the effect at the longest lag is zero. We continue shortening the lag until the trailing lag coefficient is statistically significant.

Although this method has appeal, there are dangers as well. Remember that an insignificant t statistic on the trailing lag only *fails to reject* the hypothesis of a zero coefficient; it does not prove that the coefficient is zero! It is therefore quite possible that this procedure will choose a lag length that is too short.

An alternative that also relies on statistical tests of significance is to start with a very short lag and successively add lag terms, continuing to add lags that are statistically significant and stopping when the marginal lag coefficient is not. This method is similar to the one above and often, though not necessarily always, leads to a similar choice of lag length. To

see why they are not identical, consider what would happen if the first, second, and fourth lags were (always) statistically significant but the third lag and all lags longer than four are not. Starting from a long lag and working downward you would stop at four lags, eliminating the insignificant fifth lag but retaining the third lag by convention. Starting from a short lag length and working upward you would stop at two lags; you would never discover the significant fourth lag.

3.4.2 Determining lag length by information criteria

Information criteria are designed to measure the amount of information about the dependent variable contained in a set of regressors. They are goodness-of-fit measures of the same type as R^2 or \bar{R}^2 , but without the convenient interpretation as share of variance explained that we give to R^2 in an OLS regression with an intercept term. The two most commonly used criteria are the *Akaike information criterion (AIC)* and the *Schwartz/Bayesian information criterion (SBIC)*. They are usually calculated in log form by the formulas

$$\begin{aligned} \text{AIC} &= \ln \left(\frac{\sum_{t=1}^T \hat{u}_t^2}{T} \right) + \frac{2K}{T}, \\ \text{SBIC} &= \ln \left(\frac{\sum_{t=1}^T \hat{u}_t^2}{T} \right) + \frac{K \ln T}{T}. \end{aligned} \tag{3.18}$$

In equation (3.18), T is sample length, K is the total number of estimated coefficients, and \hat{u}_t are the residuals.

The “main ingredient” in both information criteria is the sum of squared residuals, which we want to make as small as possible. Thus, we want to *minimize* the criteria and choose the model with the *smallest* AIC or SBIC value.

When using the information criteria to choose lag length, we must be very careful to make sure that all candidate models among which we are choosing are estimated over exactly the same sample period. This requires particular caution in lag models because there will usually be more observations available for models with shorter lags (because with fewer lags we “lose” fewer observations at the beginning of the sample). Passively allowing Stata to set the sample by using all available observations will result in samples with different T for models with lags of different length, thus the information criteria calculated from them cannot be compared. You should always use an `in` or `if` clause in your estimation command to keep the sample the same across regressions being compared with AIC or SBIC, then verify that all of the observations being compared have identical sample lengths.

The first term of the information criteria (common to both) is the log of the standard error of the estimate (SEE), uncorrected for degrees of freedom. This measures how well the

model explains the dependent variable. The second term is a “penalty term” that depends positively on the number of estimated parameters K . Increasing lag length will lower the first term by marginally improving the fit, but will also increase the second term because the number of parameters will be larger. The information criteria provide alternative ways of “trading off” improved fit against more parameters to estimate.

From (3.18), the SBIC penalizes additional parameters more strongly than the AIC (assuming, plausibly, that $\ln T > 2$). Thus, the SBIC always chooses a lag length that is shorter than (or the same length as) the one that minimizes the AIC. Neither is “better,” so one might consider the AIC as a lower bound and the SBIC as an upper bound for the appropriate lag length. In the case that they happen to agree, the choice is clear.

3.4.3 Determining lag length by residual autocorrelation

As discussed above, adding lags of x and/or y to the right-hand side of a distributed-lag regression usually lessens the degree of autocorrelation in the error term. Estimators of some models (especially those with lagged dependent variables) are sensitive to autocorrelated errors, so another criterion that is sometimes used for choosing lag length is the elimination of autocorrelation in the residuals.

When using residual autocorrelation to determine lag length, one adds lags until the residuals appear to be white noise. After running the distributed-lag regression, one extracts the residuals and uses a Breusch-Godfrey LM test or a Box-Ljung Q test to test the null hypothesis that the residuals are white noise. Rejecting the white-noise null hypothesis means that more lags should be added to the regression according to this criterion.

References

- Almon, Shirley. 1965. The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica* 33 (1):178-196.
- Box, George E. P., and Gwilym M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Granger, C. W. J., and Paul Newbold. 1974. Spurious Regressions in Econometrics. *Journal of Econometrics* 2 (2):111-120.
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton, N.J.: Princeton University Press.
- Koyck, L. M. 1954. *Distributed Lags and Investment Analysis*. Amsterdam: North-Holland.
- Romer, David. 2012. *Advanced Macroeconomics*. 4th ed. New York: McGraw-Hill.