Journal of Econometrics 53 (1992) 123–139. North-Holland

Estimation of polynomial distributed lags and leads with end point constraints

Donald W.K. Andrews and Ray C. Fair*
Yale University, New Haven, CT 06520, USA

Received November 1989, final version received March 1991

This paper considers the use of the polynomial distributed lag (PDL) technique when the lag length is estimated rather than fixed. We focus on the case where the degree of the polynomial is fixed, the polynomial is constrained to be zero at a certain lag length q, and q is estimated along with the other parameters. We extend the traditional PDL setup by allowing q to be real-valued rather than integer-valued, and we derive the asymptotic covariance matrix of all the parameter estimates, including the estimate of q. The paper also considers the estimation of distributed leads rather than lags, a case that can arise if expectations are assumed to be rational.

1. Introduction

This paper considers the use of the polynomial distributed lag (PDL) technique of Almon (1965) when the lag length is estimated rather than fixed. We focus on the case where the degree of the polynomial is fixed, the polynomial is constrained to be zero at a certain lag length q, and q is estimated along with the other parameters. We extend the traditional PDL setup by allowing q to be real-valued rather than integer-valued. This extension plus a minor (and quite natural) modification of the PDL yields a regression function that is twice differentiable in q. Consequently, the model is simply a nonlinear regression model, and under standard assumptions the least squares estimate of q and various functions of q and the other parameters, such as the sum of the PDL coefficients, are consistent and asymptotically normal. Furthermore, if the errors are iid and normally

* The authors thank Whitney K. Newey and two referees for very helpful comments. The first author gratefully acknowledges support from the National Science Foundation via grant number SES-8821021.

distributed, these estimates are also asymptotically efficient. Estimates of their asymptotic variances and covariances are provided.

The paper also considers the estimation of distributed leads rather than lags. If expectations are rational and if the coefficients of the lead variables are assumed to lie on a polynomial, the PDL technique can be combined with Hansen's (1982) method of moments estimator to produce consistent and asymptotically normal estimates of all the parameters, including the lead length.

Considerable attention has been paid in the literature to the adverse effects of incorrectly specifying the lag length of PDLs, e.g., Schmidt and Waud (1973), Trevedi and Pagan (1979), Hendry, Pagan, and Sargan (1984), and references therein. For a fixed lag length the parameter estimates are usually inconsistent if the lag length is misspecified. For example, if the correct specification is for an explanatory variable to enter an equation only contemporaneously and if the lag length is specified to be greater than one, then the effect of the explanatory variable on the dependent variable will not be estimated consistently.

This misspecification problem does not arise if the lag length is estimated consistently. In consequence, several papers have considered estimating the lag length, e.g., Schmidt and Waud (1973), Sargan (1980), and Pagano and Hartley (1981). In each of these papers, however, no estimated standard error is obtained for the estimated lag length, and the estimated standard errors for the other parameter estimates are computed as though the estimated lag length is fixed. As has been recognized for some time – see Schmidt (1973) and Frost (1975) – such estimated standard errors understate the true variability of the parameter estimates. In contrast, this paper provides a standard error estimate for the lag length, and the estimated standard errors for the other parameter estimates take into account the estimation of the lag length. Note that if the standard error of the lag length estimate is large in a particular empirical application, one can argue using the asymptotic efficiency result mentioned above that this is a consequence of the difficulty of estimating the lag length, not of the method.

As mentioned above, the results of this paper apply to PDLs with a zero end point constraint. Of course, the use of such a constraint is only warranted if one believes a priori that the lag coefficients decline smoothly, rather than jump abruptly, to zero. Sargan (1980, p. 119) argues that the aggregation of micro-units leads to long-tailed lag distributions, where end point constraints are often of interest. He also presents some empirical evidence that end point constraints are appropriate in some applications. We do not discuss this issue further here except to note that one may wish to test whether the end point constraint is rejected by the data. See Sargan (1980) for a discussion of such a test.

It should be noted that the lag length estimator considered here is nearly the same as that obtained using Akaike's (1974), Schwarz's (1978), or Mallows' (1973) C_p criterion. In a normal regression model for a fixed order of the polynomial, all these criteria choose the lag length that minimizes the sum of squared residuals (SSR). This occurs because the number of parameters in the model is the same regardless of the lag length chosen. Hence, the different penalties that these model selection procedures place on additional parameters are irrelevant. The only difference between the Akaike, Schwarz, and C_p criteria and the criterion considered here is that the former typically minimize the SSR over integer-valued lag lengths, whereas ours minimizes it over real-valued lag lengths.

2. Estimation of distributed lags

2.1. A simple example

It will be useful to begin with a simple example. Assume that the polynomial is linear and that there is one distributed lag variable:

$$\begin{aligned} Y_t &= X_{1t}'\beta + \int_0^q \alpha_j X_{2t-[j]}\,\mathrm{d}j + u_t, \qquad t = 1,\ldots,T, \\ &= X_{1t}'\beta + \sum_{j=0}^{[q]} \alpha_j^* X_{2t-j} + u_t, \qquad t = 1,\ldots,T, \end{aligned} \tag{1}$$

$$\alpha_j = \gamma_0 + \gamma_1 j, \qquad j \in [0, q], \tag{2}$$

$$\alpha_q = 0, \tag{3}$$

$$\alpha_j^* = \begin{cases} \int_j^{j+1} \alpha_s\,\mathrm{d}s & \text{for } j = 0,\ldots,[q]-1, \\ \int_j^q \alpha_s\,\mathrm{d}s & \text{for } j = [q], \end{cases} \tag{4}$$

where X_{1t} is a k-dimensional vector of explanatory variables other than X_{2t} and its lags, q is a real number greater than or equal to one, and $[q]$ is the

integer part of q. Y_t and the X_{2t-j} are scalars. Eqs. (2) and (3) imply that

$$\alpha_j = \alpha_j(q) = -\gamma_1(q-j), \tag{5}$$

$$\alpha_j^* = \alpha_j^*(q) = \begin{cases} -\gamma_1\left(q-j-\frac{1}{2}\right) & \text{for } j=0,\ldots,[q]-1, \\ -\gamma_1(q-[q])^2/2 & \text{for } j=[q]. \end{cases} \tag{6}$$

Since the constraints (2) and (3) imply that α_j and α_j^* are functions of q, they have been written as $\alpha_j(q)$ and $\alpha_j^*(q)$ in (5) and (6). (When q is integer-valued, the above model is a minor modification of the standard PDL model since the coefficient of X_{2t-j} is α_j^* rather than α_j.)

Let

$$\theta = (\beta', \gamma_1, q)', \qquad X_t = (X_{1t}', X_{2t}, X_{2t-1}, \ldots, X_{2t-[q]})'. \tag{7}$$

Given (5)–(7), eq. (1) can be written as

$$\begin{aligned} Y_t = g(X_t, \theta) &= X_{1t}'\beta - \gamma_1 \int_0^q (q-j) X_{2t-[j]}\, \mathrm{d}j + u_t \\ &= X_{1t}'\beta - \gamma_1 \left\{ \sum_{j=0}^{[q]-1} \left(q-j-\tfrac{1}{2}\right) X_{2t-j} + \tfrac{1}{2}(q-[q])^2 X_{2t-[q]} \right\} + u_t \\ &= X_{1t}'\beta - \gamma_1 Q_{1t} + u_t. \end{aligned} \tag{8}$$

An estimate of θ, denoted $\hat{\theta}$, can be obtained by minimizing the sum of squared residuals $u'u$, where $u' = (u_1, \ldots, u_T)$. One way this minimization can be done in practice is by searching over values of q. Given a value of q, Q_{1t} can be computed, and given Q_{1t}, eq. (8) is linear in parameters and can thus be estimated by ordinary least squares. Thus, one can search over q by running least squares regressions to find the value that leads to the smallest overall sum of squared residuals. Alternatively, a gradient method can be used to compute the estimates, where the gradient is given in (9) below.

By writing the nonlinear regression function $g(X_t, \theta)$ in terms of an integral, as in (8), it is easy to see that it is a twice differentiable function of q and the other parameters. Thus, under standard conditions the nonlinear least squares estimator $\hat{\theta}$ is consistent and asymptotically normal [e.g., see Hansen (1982), Gallant (1987, chs. 1, 2), or Andrews and Fair (1988)]. Note that q is identified only if $q \geq 1$, and it is an interior point of its parameter space, as is required for asymptotic normality, only if $q > 1$.

The estimation of the covariance matrix of $\hat{\theta}$ is straightforward. Let G be a $T \times (k+2)$ matrix whose tth row is

$$\frac{\partial}{\partial \theta'} g(X_t, \theta) = \left(X_{1t}' : -Q_{1t} : -\gamma_1 \left\{ \sum_{j=0}^{[q]-1} X_{2t-j} + (q - [q]) X_{2t-[q]} \right\} \right). \tag{9}$$

An estimate of the covariance matrix of $\hat{\theta}$ is

$$\hat{V} = \hat{\sigma}^2 (\hat{G}'\hat{G})^{-1}, \tag{10}$$

where $\hat{\sigma}^2 = \hat{u}'\hat{u}/T$, $\hat{u}$ is the vector of estimated residuals from (8), and $\hat{G}$ is G evaluated at $\theta = \hat{\theta}$. The estimate $\hat{V}$ is appropriate when the errors $\{u_t: t \geq 1\}$ are independent, mean zero, variance σ^2 random variables conditional on $\{X_t: t \geq 1\}$. $\hat{V}$ is easy to compute in practice, since G is simply the matrix of regressors expanded by one column to include the derivative of $g(X_t, \theta)$ with respect to q.[1]

In most PDL applications one is interested in the sum λ of the lag coefficients. In the present context λ is given by

$$\lambda = \int_0^q \alpha_j(q)\, \mathrm{d}j = -\gamma_1 \int_0^q (q - j)\, \mathrm{d}j = -\gamma_1 q^2/2. \tag{11}$$

The least squares estimate of λ, $-\hat{\gamma}_1 \hat{q}^2/2$, has asymptotic variance

$$\sigma^2(\hat{\lambda}) = (\partial\lambda/\partial\gamma_1, \partial\lambda/\partial q) V_2 (\partial\lambda/\partial\gamma_1, \partial\lambda/\partial q)', \tag{12}$$

where V_2 is the 2×2 covariance matrix of $(\hat{\gamma}_1, \hat{q})'$, i.e., the lower right 2×2 block of the covariance matrix of $\hat{\theta}$, and

$$\partial\lambda/\partial\gamma_1 = -q^2/2, \tag{13}$$

$$\partial\lambda/\partial q = -\gamma_1 q. \tag{14}$$

$\sigma^2(\hat{\lambda})$ can be estimated using the lower right 2×2 block of $\hat{V}$ in (10) and evaluating (13) and (14) at $q = \hat{q}$ and $\gamma_1 = \hat{\gamma}_1$.

We now consider various extensions of model (1).

[1] If $\hat{q}$ is equal to one, then $\hat{V}$ in (10) is singular, and so the variance of $\hat{q}$ cannot be computed. Also, values of $\hat{q}$ close to one are likely to result in very large estimates of its variance (since $\hat{V}$ would be nearly singular). In practice, of course, if $\hat{q}$ is equal to or close to one, the method proposed in this paper is not needed.

2.2. Endogenous explanatory variables

If X_{2t} or some of the variables in X_{1t} are endogenous and if a matrix Z of first-stage regressors is available, eq. (1) can be estimated by two-stage least squares (2SLS). $\hat{\theta}$ is obtained by minimizing $u'Z(Z'Z)^{-1}Z'u$, and the estimated covariance matrix is

$$\hat{V} = \hat{\sigma}^2\left(\hat{G}'Z(Z'Z)^{-1}Z'\hat{G}\right)^{-1}. \tag{15}$$

Again, $\hat{\theta}$ can be computed by searching over values of q. Given q, the problem is a standard 2SLS estimation problem. Alternatively, a gradient method can be used.

2.3. Quadratic polynomials

If the polynomial is quadratic:

$$\alpha_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2, \qquad j \in [0, q], \tag{16}$$

$$\alpha_q = 0, \tag{17}$$

and so

$$\alpha_j = \alpha_j(q) = -\gamma_1(q-j) - \gamma_2(q^2 - j^2), \qquad j \in [0, q], \tag{18}$$

and

$$\alpha_j^* = \alpha_j^*(q) = \begin{cases} -\gamma_1\left(q - j - \frac{1}{2}\right) - \gamma_2\left(q^2 - j^2 - j - \frac{1}{3}\right) \\ \qquad\qquad \text{for} \quad j = 0, \ldots, [q] - 1, \\ -\gamma_1(q - [q])^2/2 - \gamma_2\left(\frac{2}{3}\right)(q - [q])^2(q + [q]/2) \\ \qquad\qquad \text{for} \quad j = [q]. \end{cases} \tag{19}$$

In this case, $q \geq 2$ is needed for identification, θ contains an extra element γ_2, i.e., $\theta' = (\beta', \gamma_1, \gamma_2, q)$, and eq. (8) becomes

$$Y_t = X_{1t}'\beta - \gamma_1 Q_{1t} - \gamma_2 Q_{2t} + u_t, \tag{20}$$

where Q_{1t} is as in (8) and

$$Q_{2t} = \sum_{j=0}^{[q]-1} \left(q^2 - j^2 - j - \tfrac{1}{3}\right) X_{2t-j} + \tfrac{2}{3}(q - [q])^2 (q + [q]/2) X_{2t-[q]}. \tag{21}$$

Eq. (9) becomes

$$\frac{\partial}{\partial \theta'} g(X_t, \theta) = \left(X'_{1t} : -Q_{1t} : -Q_{2t} : -(\gamma_1 + 2q\gamma_2) \left\{ \sum_{j=0}^{[q]-1} X_{2t-j} + (q - [q]) X_{2t-[q]} \right\} \right), \tag{22}$$

and eqs. (11)–(14) become

$$\lambda = \int_0^q \alpha_j(q)\, \mathrm{d}j = -\gamma_1 q^2/2 - 2\gamma_2 q^3/3, \tag{23}$$

$$\sigma^2(\hat{\lambda}) = (\partial\lambda/\partial\gamma_1, \partial\lambda/\partial\gamma_2, \partial\lambda/\partial q) V_3 (\partial\lambda/\partial\gamma_1, \partial\lambda/\partial\gamma_2, \partial\lambda/\partial q)', \tag{24}$$

$$\partial\lambda/\partial\gamma_1 = -q^2/2, \tag{25}$$

$$\partial\lambda/\partial\gamma_2 = -2q^3/3, \tag{26}$$

$$\partial\lambda/\partial q = -\gamma_1 q - 2\gamma_2 q^2, \tag{27}$$

where V_3 is the 3×3 covariance matrix of $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{q})'$.

2.4. *Multiple distributed lag variables*

If model (1) contains a second distributed lag variable, say X_{3t-j}, two cases need to be considered, one in which the lag lengths for X_2 and X_3 are the same and the other in which they are not. If they are the same, the new term in (1) is $\int_0^q \eta_j X_{3t-[j]}\, \mathrm{d}j$, where (assuming a linear polynomial) $\eta_j = \delta_0 + \delta_1 j$, $j \in [0, q]$, and $\eta_q = 0$. θ now contains an extra element δ_1, i.e., $\theta' =$

$(\beta', \gamma_1, \delta_1, q)$, and eq. (8) becomes

$$Y_t = X_{1t}'\beta - \gamma_1 Q_{1t} - \delta_1 R_{1t} + u_t, \tag{28}$$

where

$$R_{1t} = \sum_{j=0}^{[q]-1} \left(q - j - \tfrac{1}{2}\right) X_{3t-j} + \tfrac{1}{2}(q - [q])^2 X_{3t-[q]}. \tag{29}$$

Eq. (9) becomes

$$\frac{\partial}{\partial \theta'} g(X_t, \theta) = \Bigg(X_{1t}' : -Q_{1t} : -R_{1t} : \\ -\gamma_1 \Bigg\{ \sum_{j=0}^{[q]-1} X_{2t-j} + (q - [q]) X_{2t-[q]} \Bigg\} \\ -\delta_1 \Bigg\{ \sum_{j=0}^{[q]-1} X_{3t-j} + (q - [q]) X_{3t-[q]} \Bigg\} \Bigg). \tag{30}$$

If the lag lengths are not equal, the new term in (1) is $\int_0^r \eta_j X_{3t-[j]}\, dj$, where (assuming a linear polynomial) $\eta_j = \delta_0 + \delta_1 j$, $j \in [0, r]$, and $\eta_r = 0$. θ now contains two extra elements δ_1 and r, i.e., $\theta' = (\beta', \gamma_1, \delta_1, q, r)$. Eq. (8) becomes eq. (28) except that r replaces q in the definition of R_{1t} given in (29). Eq. (9) becomes

$$\frac{\partial}{\partial \theta'} g(X_t, \theta) = \Bigg(X_{1t}' : -Q_{1t} : -R_{1t} : \\ -\gamma_1 \Bigg\{ \sum_{j=0}^{[q]-1} X_{2t-j} + (q - [q]) X_{2t-[q]} \Bigg\} : \\ -\delta_1 \Bigg\{ \sum_{j=0}^{[r]-1} X_{3t-j} + (r - [r]) X_{3t-[r]} \Bigg\} \Bigg). \tag{31}$$

When there are two lag lengths rather than one, the computational burden of searching over different lag lengths is more burdensome, and a gradient method is likely to be much faster.

The extension to models with quadratic polynomials and more than two lagged variables is straightforward. In addition, the extension is straightforward to models with a PDL on both the dependent variable and various independent variables, as in the class of autoregressive distributed lag models considered in Hendry, Pagan, and Sargan (1984).

2.5. *Nonlinearity*

Finally, eq. (1) – and thus $g(X_t, \theta)$ in (8) – can be nonlinear in parameters other than just q. Given q, the minimization of $u'u$ need not be an ordinary least squares problem, and the derivatives of $g(X_t, \theta)$ with respect to θ can be more involved than those in (9). This means, among other things, that the case in which u_t is nth order autoregressive can be handled easily. Eq. (1) can be quasi-differenced using the autoregressive parameters in order to eliminate the autoregressive part of the error, and the autoregressive coefficients can be incorporated into θ. This merely converts the problem into one in which $g(X_t, \theta)$ is more nonlinear in parameters than otherwise.

2.6. *Estimation and testing of the degree of the polynomial*

Thus far we have considered the case where the degree of the polynomial is fixed. It is possible, however, to estimate both the lag length and the degree of the polynomial and to test the adequacy of a specified polynomial degree. With the lag length q treated as a real-valued parameter to be estimated, a sequence of models with PDLs of increasing degrees is a sequence of nested nonlinear regression models. Therefore, any of a number of standard consistent model selection procedures can be applied to estimate the polynomial degree. For example, one can use a downward sequential t- or F-testing procedure, as in Pagano and Hartley (1981), or one can use Akaike's information criterion, Schwartz's criterion, Mallow's C_p criterion, cross-validation, generalized cross-validation, or a posterior odds procedure, etc. With a consistent model selection procedure, the asymptotic variances given above are still valid (because the correct model is selected with probability that goes to one as T goes to infinity), but the accuracy of the asymptotic approximation is likely to suffer.

The adequacy of a given choice of polynomial degree can be tested using an asymptotic t- or F-test as in Pagano and Hartley (1981). A RESET or RASET specification test can be used to test whether the degree of the polynomial is correct and whether the PDL restriction itself is appropriate [see Harper (1977)].

This completes the discussion of distributed lags. The cases considered in this section can be easily extended and combined, and in each case it is

straightforward to treat the lag length or lengths as parameters to be estimated and to estimate their standard errors.

3. Estimation of distributed leads

Suppose that X_{2t-j} in (1) is replaced by X^e_{2t+j}, where the latter is the expected value of X_{2t+j} and all expectations are assumed to be formed at the end of period $t-1$, before information for period t is available. Let the expectation error for X^e_{2t+j} be

$$ {}_{t-1}\varepsilon_{t+j} = X_{2t+j} - X^e_{2t+j}, \qquad j = 0, 1, \ldots, [q]. \tag{32} $$

Eq. (1) in this case is

$$ \begin{aligned} Y_t &= X_{1t}'\beta + \int_0^q \alpha_j X^e_{2t+[j]}\,\mathrm{d}j + u_t, \qquad t = 1, \ldots, T, \\ &= X_{1t}'\beta + \sum_{j=0}^{[q]} \alpha_j^* X_{2t+j} + v_t, \qquad t = 1, \ldots, T, \end{aligned} \tag{33} $$

where

$$ v_t = -\int_0^q \alpha_{jt-1}\varepsilon_{t+[j]}\,\mathrm{d}j + u_t. \tag{34} $$

Given (2)–(7), a new eq. (8) can be derived:

$$ \begin{aligned} Y_t &= g(X_t, \theta) \\ &= X_{1t}'\beta - \gamma_1 \int_0^q (q-j) X_{2t+[j]}\,\mathrm{d}j + v_t \\ &= X_{1t}'\beta - \gamma_1 \left\{ \sum_{j=0}^{[q]-1} \left(q - j - \tfrac{1}{2}\right) X_{2t+j} + \tfrac{1}{2}(q-[q])^2 X_{2t+[q]} \right\} + v_t \\ &= X_{1t}'\beta - \gamma_1 Q_{1t} + v_t, \end{aligned} \tag{35} $$

where X_t now denotes $(X_{1t}', X_{2t}, X_{2t+1}, \ldots, X_{2t+[q]})'$.

Consider first 2SLS estimation of (35). Let Z_t be a vector of first-stage regressors. A necessary condition for consistency is that Z_t and v_t be uncorrelated. This will be true if both u_t and the ${}_{t-1}\varepsilon_{t+j}$ are mean zero and uncorrelated with Z_t. The assumption that u_t is mean zero and uncorrelated

with Z_t is the usual 2SLS assumption. The assumption that the ${}_{t-1}\varepsilon_{t+j}$ are mean zero and uncorrelated with Z_t is the rational expectations assumption. If expectations are formed rationally and if the variables in Z_t are used (perhaps along with others) in forming the expectations of the X_{2t+j}, then Z_t and the ${}_{t-1}\varepsilon_{t+j}$ are uncorrelated. Therefore, given this assumption (and the other standard assumptions that are necessary for consistency), the 2SLS estimator of θ is consistent. It minimizes $v'Z(Z'Z)^{-1}Z'v$.

A problem with the 2SLS estimator in this context is that it ignores the m-dependent property of v_t. Because of the ${}_{t-1}\varepsilon_{t+j}$, v_t will in general be m-dependent with $m = [q] - 1$ if q is not an integer and $m = [q] - 2$ if q is an integer. The 2SLS estimates are consistent, but the standard formula for their covariance matrix in (15) is incorrect and the estimates are not efficient within the class of limited information estimators. Hansen's (1982) method of moments estimator takes account of the m-dependent character of v_t. It is based on minimizing $v'ZM^{-1}Z'v$, where M is some consistent estimate of $\lim T^{-1}\mathrm{E}\{Z'vv'Z\}$. In order to construct an estimate of M one needs an estimate of v_t in (35), such as the 2SLS estimate $\hat{v}_t$.

A general way of computing M is as follows. Let $f_t = \hat{v}_t Z_t$. Let $R_j = T^{-1}\sum_{t=j+1}^{T} f_t f_{t-j}'$, $j = 0, 1, \ldots, m$. M is then $(R_0 + R_1 + R_1' + \cdots + R_m + R_m')$. In many cases computing M in this way does not yield a positive definite matrix, and something else must be done. Hansen (1982), Cumby, Huizinga, and Obstfeld (1983), Andrews (1991), and Andrews and Monahan (1990), among others, discuss the computation of M based on an estimate of the spectral density matrix of $Z_t'v_t$ evaluated at frequency zero. A third approach is to compute M under the following homoskedasticity assumption:

$$\mathrm{E}[v_t v_s | Z_t, Z_{t-1}, \ldots] = \mathrm{E}[v_t v_s] \quad \text{for} \quad t \geq s, \tag{36}$$

which says that the contemporaneous and serial correlation in v do not depend on Z. This assumption is implied by the assumption that $\mathrm{E}[v_t Z_s] = 0$ for $t \geq s$ if normality is also assumed. Under this assumption M can be computed as follows. Let $a_j = T^{-1}\sum_{t=j+1}^{T} \hat{v}_t \hat{v}_{t-j}$ and $B_j = T^{-1}\sum_{t=j+1}^{T} Z_t Z_{t-j}'$, $j = 0, 1, \ldots, m$. M is then $(a_0 B_0 + a_1 B_1 + a_1 B_1' + \cdots + a_m B_m + a_m B_m')$.

The complete estimation procedure in the case of polynomial distributed leads can now be summarized. 1) Estimate β, γ_1, and q in (35) by 2SLS, which minimizes $v'Z(Z'Z)^{-1}Z'v$. This requires searching over values of q or using a gradient method. 2) Given these estimates, compute $\hat{v}_t$ from (35). Then compute M in the one of the above ways. 3) Estimate β, γ_1, and q in (35) by minimizing $v'ZM^{-1}Z'v$. This again requires searching over values of q or using a gradient method. These are the final parameter estimates. The estimated covariance matrix of these estimates is

$$\hat{V} = T\left(\hat{G}'ZM^{-1}Z'\hat{G}\right)^{-1}, \tag{37}$$

where the elements of G are as in (9) except that X_{2t+j} replaces X_{2t-j} and $X_{2t+[q]}$ replaces $X_{2t-[q]}$.

The various extensions discussed in section 2 can also be applied here. The modifications needed for the case of leads rather than lags are slight, and they will not be discussed further.

4. Monte Carlo results

We report results from some Monte Carlo experiments in this section. The experiments are based on three of the equations estimated in Fair (1990). The equations are price equations for fairly specific commodities, with distributed lag or lead values of an aggregate price variable added to pick up aggregate expectational effects on individual price setting behavior. The data are monthly. The equations are estimated using 330, 359, and 323 observations, respectively, and include 21, 18, and 21 explanatory variables, respectively. Eleven of the explanatory variables in each equation are seasonal dummy variables. The polynomial is taken to be linear, and the aggregate price variable is the only variable to which the polynomial lag distribution is applied. Given the lag length q, each equation is linear in parameters.

The Monte Carlo experiments were run as follows. Each equation was first estimated using the historical data. For the first case for each equation all the estimated parameters were used as the true parameters and the error term in the equation was assumed to be normal with mean zero and variance $\hat{\sigma}^2$, where the latter is the estimated variance of the equation. For each repetition a new data set was generated by drawing error terms from this distribution and using these error terms plus the estimated parameters to compute new values of the dependent variable. The equation was then re-estimated using the new data, and the parameter estimates were recorded (including the estimate of q). The number of repetitions per equation was 750. The largest value of q allowed was 132. (The smallest value of q allowed was 1.)

Three estimators were computed for each repetition: 1) the estimator proposed in this paper, 2) the estimator when q is known, and 3) an estimator proposed by Sargan (1980, pp. 117–118). For the case of a polynomial of known degree, the estimator proposed by Sargan reduces to starting at a specified minimum lag length and increasing the lag length by one until the sum of squared residuals increases. In the Monte Carlo experiments we took the minimum lag length to be one. If the sum of squares is a bowl-shaped function of the lag length, then this procedure gives the same estimate as ours except for the integer-value, real-value difference. If, on the other hand, there are local minima, then Sargan's procedure may stop before reaching the global minimum. For many of the cases reported below there are local minima, and so, as will be seen, Sargan's estimates are not always close to

ours.[2] Sargan's procedure does not provide standard errors for the estimated lag length, and it does not adjust the standard errors of the other coefficient estimates to reflect the estimation of q.

For each repetition the estimates of q and λ were recorded. The means and variances of the estimates of q and λ over the 750 repetitions were then calculated. In addition, the variances of the estimates of q and λ were estimated for each repetition using the asymptotic formulae, and the averages of these estimates over the 750 repetitions were calculated.[3] In table 1, $E(\hat{q})$ denotes the mean of the 750 estimates of q, $SE(\hat{q})$ denotes the square root of the variance of the 750 estimates, and ave. asy. $SE(\hat{q})$ denotes the square root of the average of the 750 calculations of the variance computed from the asymptotic formulae. Similar notation holds for λ. In addition, a 95% confidence interval for $\hat{q}$ was computed for each repetition using the asymptotic formulae, and the percentage of repetitions in which the true value of q fell outside of this interval was calculated. This percentage for each case is presented in the table under the heading '5% nom. test'. If the asymptotic formulae were exact, this percentage would be 5.0. A 90% confidence interval was also computed, and the results are recorded in the table under the heading '10% nom. test'. Similar computations were done for λ.

The first case for each equation in table 1 uses the estimated parameters as the truth. For each of the other cases either q or σ is changed and used as the truth, with the other parameters remaining at their estimated values.

The following conclusions can be drawn from table 1. (1) In a number of cases the Sargan estimates of q are much too small, which reveals a local minima problem. Otherwise, as expected, the Sargan estimates are quite close to the estimates using the method of this paper.[4] (2) The estimates of q using the method of this paper in general show only a small bias, which is always upward, and similarly for the estimates of λ. (3) The asymptotic estimates of the SE of $\hat{\lambda}$ are generally close to the Monte Carlo estimates.[5] Both of these estimates are higher than the estimates when q is known,

[2]Because of the local minima problem, we used a grid search on q in the computations for our estimator. If a gradient algorithm had been used, there would have been no guarantee that the global minimum had been found.

[3]For two repetitions (one for the third case for eq. 1 in table 1 and one for the third case for eq. 3), $\hat{q}$ was equal to one and hence the formula for the asymptotic variance of $\hat{q}$ and $\hat{\lambda}$ was inapplicable. These two repetitions were skipped for purposes of computing ave. asy. $SE(\hat{q})$ and ave. asy. $SE(\hat{\lambda})$, although they were not skipped for the other calculations. This skipping was not important. Instead of skipping the repetitions, we also computed the asymptotic formulae using $\hat{q} = 1.2$, and this had a trivial effect on the values in table 1.

[4]In three cases in table 1 Sargan's estimate of q is slightly larger than ours. This can happen because of the integer-value, real-value difference between the estimators. Sargan stops at the nearest integer, and our estimate may be slightly below this integer.

[5]In the following discussion, ave. asy. $SE(\hat{\lambda})$ will be referred to as 'the asymptotic estimate of the SE of $\hat{\lambda}$' and $SE(\hat{\lambda})$ will be referred to as 'the Monte Carlo estimate of the SE of $\hat{\lambda}$'. Similar statements hold for $\hat{q}$.

Table 1

Monte Carlo results.[a]

	$E(\hat{q})$	$E(\hat{\lambda})$	$SE(\hat{q})$	Avc. asy. $SE(\hat{q})$	5% nom. test	10% nom. test	$SE(\hat{\lambda})$	Avc. asy. $SE(\hat{\lambda})$	5% nom. test	10% nom. test
Eq. 1: Inner tubes										
$\sigma = 0.02090$										
Truth	28.57	1.840	—	—	—	—	—	—	—	—
q estimated	29.48	1.954	14.49	11.99	15.6	20.9	0.526	0.504	6.4	11.1
q known	28.57	1.861	0.00	0.00	—	—	0.432	0.436	5.3	10.4
Sargan	19.47	1.632	14.54	—	—	—	0.686	—	—	—
Truth	15.00	1.840	—	—	—	—	—	—	—	—
q estimated	16.70	1.935	8.88	7.04	15.9	21.9	0.419	0.419	5.6	11.1
q known	15.00	1.863	0.00	0.00	—	—	0.385	0.385	4.7	10.0
Sargan	13.31	1.802	7.63	—	—	—	0.530	—	—	—
Truth	3.00	1.840	—	—	—	—	—	—	—	—
q estimated	3.20	1.886	1.07	0.79	12.5	16.7	0.331	0.310	6.1	12.0
q known	3.00	1.848	0.00	0.00	—	—	0.264	0.270	4.1	8.8
Sargan	3.22	1.887	1.04	–	—	—	0.334	—	—	—
$\sigma = 0.01045$										
Truth	28.57	1.840	—	—	—	—	—	—	—	—
q estimated	28.84	1.893	6.24	5.51	9.2	14.9	0.255	0.260	5.3	10.4
q known	28.57	1.867	0.00	0.00	—	—	0.232	0.237	5.5	9.6
Sargan	27.70	1.860	7.57	—	—	—	0.317	—	—	—
Eq. 2: Insect wire screening										
$\sigma = 0.01070$										
Truth	4.26	0.921	—	—	—	—	—	—	—	—
q estimated	4.51	0.942	1.69	1.23	12.8	16.4	0.156	0.150	6.9	12.8
q known	4.26	0.928	0.00	0.00	—	—	0.139	0.137	5.2	11.5
Sargan	4.41	0.938	1.37	—	—	—	0.158	—	—	—
Truth	15.00	0.921	—	—	—	—	—	—	—	—
q estimated	17.19	0.956	9.92	7.18	19.2	23.3	0.206	0.189	7.5	12.9
q known	15.00	0.927	0.00	0.00	—	—	0.177	0.175	5.9	10.8
Sargan	13.78	0.902	7.91	—	—	—	0.232	—	—	—
Truth	30.00	0.921	—	—	—	—	—	—	—	—
q estimated	31.46	0.971	15.57	12.87	16.9	21.9	0.247	0.224	7.7	12.9
q known	30.00	0.929	0.00	0.00	—	—	0.197	0.195	5.1	11.1
Sargan	21.70	0.844	15.28	—	—	—	0.301	—	—	—
$\sigma = 0.00535$										
Truth	4.26	0.921	—	—	—	—	—	—	—	—
q estimated	4.30	0.932	0.58	0.53	7.6	12.1	0.087	0.082	5.6	11.7
q known	4.26	0.927	0.00	0.00	—	—	0.076	0.075	5.5	11.1
Sargan	4.32	0.932	0.64	—	—	—	0.089	—	—	—
Truth	15.00	0.921	—	—	—	—	—	—	—	—
q estimated	15.53	0.936	3.60	3.03	10.9	17.1	0.106	0.101	6.0	12.7
q known	15.00	0.928	0.00	0.00	—	—	0.098	0.096	5.6	10.7
Sargan	15.25	0.933	3.45	—	—	—	0.107	—	—	—
Truth	30.00	0.921	—	—	—	—	—	—	—	—
q estimated	30.65	0.944	6.88	6.18	10.0	14.3	0.119	0.117	5.6	11.2
q known	30.00	0.929	0.00	0.00	—	—	0.108	0.106	5.3	10.7
Sargan	29.77	0.936	7.62	—	—	—	0.130	—	—	—

Table 1 (continued)

	$E(\hat{q})$	$E(\hat{\lambda})$	$SE(\hat{q})$	Ave. asy. $SE(\hat{q})$	5% nom. test	10% nom. test	$SE(\hat{\lambda})$	Ave. asy. $SE(\hat{\lambda})$	5% nom. test	10% nom. test
Eq. 3: Ball and roller bearings										
$\sigma = 0.00816$										
Truth	23.57	0.799	—	—	—	—	—	—	—	—
q estimated	25.30	0.836	12.65	9.89	17.1	21.1	0.169	0.162	8.0	13.1
q known	23.57	0.809	0.00	0.00	—	—	0.160	0.153	7.3	13.1
Sargan	20.57	0.775	11.61	—	—	—	0.238	—	—	—
Truth	15.00	0.799	—	—	—	—	—	—	—	—
q estimated	16.78	0.830	9.07	6.70	12.7	18.5	0.160	0.152	7.7	14.0
q known	15.00	0.811	0.00	0.00	—	—	0.154	0.147	7.2	12.9
Sargan	14.76	0.799	7.68	—	—	—	0.198	—	—	—
Truth	3.00	0.799	—	—	—	—	—	—	—	—
q estimated	3.15	0.815	0.83	0.63	11.2	14.5	0.131	0.120	8.1	13.6
q known	3.00	0.800	0.00	0.00	—	—	0.106	0.105	5.2	10.4
Sargan	3.18	0.815	0.87	—	—	—	0.135	—	—	—
Truth	40.00	0.799	—	—	—	—	—	—	—	—
q estimated	42.16	0.851	19.68	16.58	14.7	19.3	0.203	0.192	6.9	11.1
q known	40.00	0.807	0.00	0.00	—	—	0.171	0.164	6.8	12.3
Sargan	29.64	0.715	22.14	—	—	—	0.325	—	—	—

[a]Each case is based on 750 repetitions. The seed is the same for each case within an equation, but it differs across the three equations. $SE(\cdot)$ is the Monte Carlo standard error. Ave. asy. $SE(\cdot)$ is the square root of the average of the 750 variances computed using the asymptotic formulae. 5% (10%) nom. test is the percentage of repetitions in which the true value of q or λ is outside of the 95% (90%) confidence interval, where the confidence interval is computed using the asymptotic formulae. If the asymptotic formulae were exact the percentages would be 5.0 (10.0).

which shows that one underestimates the uncertainty of $\hat{\lambda}$ if q is taken incorrectly to be known. (4) The asymptotic estimates of the SE of $\hat{q}$ are always lower than the Monte Carlo estimates. This is true across different values of q and λ, and so there appears to be a general downward bias to the asymptotic estimates. The bias as a percent of the Monte Carlo standard error – $SE(\hat{q})$ – ranges from 9% to 28%. The average of the percent biases in table 1 is 18%. This bias can also be seen in the results for the 5% and 10% nominal tests. The percentage of repetitions in which the true value of q lies outside the 95% confidence interval ranges from 7.6% to 19.2% in the table. The range for the 90% confidence interval is 12.1% to 23.3%. Given that the left tail of the distribution of $\hat{q}$ is truncated at one, it must be that the right tail of the actual distribution is fatter than that of the asymptotic distribution.

We also estimated for each repetition the equation using a quadratic polynomial. We then used first the Akaike and second the Schwarz criterion to choose between the linear and quadratic polynomials. (Remember that the

data are generated using the linear polynomial.) Both criteria always resulted in the linear polynomial being chosen. This result is encouraging in that it says that if the truth is linear and one selects between linear and quadratic, the linear is likely to be chosen.

To conclude, the overall results are fairly favorable to the method proposed in this paper. The biases are fairly small except for the estimates of the SE of $\hat{q}$, which are too small by about 18%.

5. Conclusion

Since it is quite rare that lag and lead lengths are known with certainty, the ability to estimate them and adjust the standard errors of all the coefficient estimates to account for their estimation should prove useful in practice. In addition, although biased downward somewhat, the estimated standard errors on the lag length estimates provided here should help one in deciding how much confidence to place on the overall estimated lag or lead distributions.

References

Akaike, H., 1974, A new look at the statistical identification model, IEEE Transactions on Automatic Control 19, 716–723.

Almon, Shirley, 1965, The distributed lag between capital appropriations and expenditures, Econometrica 33, 178–196.

Andrews, Donald W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817–858.

Andrews, Donald W.K. and Ray C. Fair, 1988, Inference in econometric models with structural change, Review of Economic Studies 55, 615–640.

Andrews, Donald W.K. and J. Christopher Monahan, 1990, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, Cowles Foundation discussion paper no. 942 (Yale University, New Haven, CT).

Cumby, Robert E., John Huizinga, and Maurice Obstfeld, 1983, Two-step two-stage least squares estimation in models with rational expectations, Journal of Econometrics 21, 333–355.

Fair, Ray, C., 1990, Inflationary expectations and price setting behavior, Mimeo.

Frost, Peter A., 1975, Some properties of the Almon lag technique when one searches for degree of polynomial and lag, Journal of the American Statistical Association 70, 606–612.

Gallant, A.R., 1987, Nonlinear statistical models (Wiley, New York, NY).

Hansen, Lars, 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029–1054.

Harper, Charles P., 1977, Testing for the existence of a lagged relationship within Almon's method, Review of Economics and Statistics 59, 204–210.

Hendry, David F., Adrian Pagan, and J. Dennis Sargan, 1984, Dynamic specification, Ch. 18 in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam).

Mallows, Colin L., 1973, Some comments on C_p, Technometrics 15, 661–675.

Pagano, Marcello and Michael J. Hartley, 1981, On fitting distributed lag models subject to polynomial restrictions, Journal of Econometrics 16, 171–198.

Sargan, J.D., 1980, The consumer price equation in the post war British economy: An exercise in equation specification testing, Review of Economic Studies 47, 113–135.
Schmidt, Peter, 1973, On the difference between the conditional and unconditional asymptotic distributions of estimates in distributed lag models with integer-valued parameters, Econometrica 41, 165–169.
Schmidt, Peter and Roger N. Waud, 1973, The Almon lag technique and the monetary versus fiscal policy debate, Journal of the American Statistical Association 68, 11–19.
Schwarz, Gideon, 1978, Estimating the dimension of a model, Annals of Statistics 6, 461–464.
Trevedi, Pravin and Adrian Pagan, 1979, Polynomial distributed lags: A unified treatment, Economic Studies Quarterly 30, 37–49.