Pierre M. Windal
Université de Sherbrooke

Doyle L. Weiss
University of British Columbia

An Iterative GLS Procedure for Estimating the Parameters of Models with Autocorrelated Errors Using Data Aggregated over Time

This paper develops an iterative generalized least-squares (GLS) procedure for estimating the parameters of certain economic relations characterized by first-order autocorrelated disturbances ($y_t = bx_t + \epsilon_t$, $\epsilon_t = p\,\epsilon_{t-1} + u_t$) when the available data have been aggregated over time. The estimation procedure is conditional on knowledge of the level of aggregation (the number of subintervals) making up the aggregate data interval. An example of the estimation procedure is provided using a set of annual sales-advertising data.

Introduction

The purpose of this paper is to develop a procedure to estimate the parameters for a class of distributed lag models by making use of data aggregated over time. The general problem of aggregating economic relations over time has already received considerable attention in the literature. Theil (1954) explained the difficulties of obtaining the correct aggregate relation when lagged variables appeared in the micro relationship, but he did not consider the role of the disturbance term. Mundlak (1961) studied the effects of aggregation over time on the partial adjustment model and developed the relationship between the parameters of the micro relation and those of the mispecified (to accommodate aggregate data) time-aggregated macro model. Mundlak was also unconcerned with the full role of the disturbances in such circumstances. Morigouchi (1970), taking account of disturbances, was able to quantify both the estimation bias and the loss of efficiency resulting from temporal aggregation for certain cases of the independent variable $x(t)$. More recently, Rowe (1976) has demonstrated the effect of temporal aggregation on regression t-ratios and R^2s. All of these studies are similar in that they did not devise an estimation procedure to correct the bias arising from temporal aggregation. In a later sec-

(*Journal of Business*, 1980, vol. 53, no. 4)

0021-9398/80/5304-0004$01.50

tion, we will develop an estimation procedure that eliminates the temporal aggregation bias altogether for distributed lag relations in which the lagged variable arises from first-order autocorrelation among the disturbances.

Such a model has been used in efforts to understand the influence of advertising expenditures on sales (a partial list of these efforts includes Bass and Clark [1972]; Clark and McCann [1973]; Houston and Weiss [1975]; Clarke [1976]; and Weiss, Houston, and Windal [1978]). In such cases it has been called the "autoregressive current effects" model (ACE). In a later section we will provide an example of our procedure using the ACE model and a sales-advertising data series.

Temporal Aggregation Bias in the Autoregressive Current Effects Model

The ACE model can be written as

$$y(t) = b_1 x(t) + \epsilon(t), \tag{1}$$

where

$$\begin{aligned} \epsilon(t) &= b_2 \epsilon(t-1) + u(t), \\ u(t) &\sim N(0, \sigma_u^2), \end{aligned} \tag{2}$$

and $y(t)$ = sales at period t, $x(t)$ = advertising expenditures at period t, and $t = 1, 2, \ldots, n$. By successive substitution, one obtains the distributed lag

$$y(t) = b_1 x(t) + b_2 y(t-1) - b_3 x(t-1) + u(t), \tag{3}$$

where $b_3 = b_1 b_2$.

When the basic relation as shown by (3) is aggregated over r periods, it can be written as[1]

$$Y(t) = b_1 X(t) + b_2 Y\left(t - \frac{1}{r}\right) - b_3 X\left(t - \frac{1}{r}\right) + U(t), \tag{4}$$

where

$$Y(t) = \sum_{i=0}^{r-1} y(t \cdot r - i),$$

$$X(t) = \sum_{i=0}^{r-1} x(t \cdot r - i),$$

$$U(t) = \sum_{i=0}^{r-1} u(t \cdot r - i),$$

$$t = 1, 2, \ldots, \frac{n}{r}.$$

1. Our notation is in the spirit of Morigouchi (1970).

Usually data are not available for both $Y(t)$ and $Y(t - 1/r)$ or $X(t)$ and $X(t - 1/r)$. This causes the researcher to specify the following statistical relation for a longer basic period:

$$Y(t) = c_1X(t) + c_2Y(t - 1) - c_3X(t - 1) + U(t). \tag{5}$$

To illustrate the nature of the specification error in (5) we will clarify the relation between the b's in (4) and the c's in (5) by drawing on Morigouchi's results.

Following Morigouchi and substituting across r periods, (3) may be written:

$$\begin{aligned} y(t) = b_1 \sum_{i=0}^{r-1} b_2^i\, x(t - i) - b_3 \sum_{i=0}^{r-1} b_2^i\, x(t - i - 1) \\ + \quad b_2^r\, y(t - r) + \sum_{i=0}^{r-1} b_2^i\, u(t - i). \end{aligned} \tag{6}$$

Then by aggregating (6) over r periods we get a form of (4) which may be conveniently compared with (5), the temporal misspecification. That is, upon aggregate of (6) over r periods we obtain[2]

$$\begin{aligned} Y(t) = b_1X(t) - b_3X(t - 1) + \frac{b_1b_2}{1 - b_2} Z_1(t) \\ + \frac{b_1b_2}{1 - b_2} Z_2(t) + b_2^rY(t - 1) + V_1(t) + V_2(t), \end{aligned} \tag{7}$$

where

$$Z_1(t) = \sum_{i=0}^{r-1} (1 - b_2^i)\, [x(t - i) - b_3x(t - i - 1)],$$

$$Z_2(t) = \sum_{i=0}^{r-2} (b_2^i - b_2^{r-1})\, [x(t - r - i) - b_3x(t - r - i - 1)],$$

$$V_1(t) = \sum_{i=0}^{r-1} \frac{1 - b_2^{i+1}}{1 - b_2} u(t - i),$$

$$V_2(t) = b_2 \sum_{i=0}^{r-2} \frac{b_2^i - b_2^{r-1}}{1 - b_2} u(t - r - i).$$

Now since (5) is "nested" into (7) we can easily see that the specification errors in (5) are the missing Z_1, Z_2, and V_2 terms. By assuming that equation (3) also holds at the aggregate level, one misspecifies the basic relation in that relevant independent variables (Z_1 and Z_2) are ignored. Moreover, the presence of V_2 in (7) implies

2. Actually making only $r - 1$ substitutions. The use of Z_1 and Z_2 in (7) allows the nested relationships between (5) and (7) to be clearly shown. This structuring of the misspecification follows Morigouchi.

TABLE 1 Estimated Value of the Autocorrelation Coefficient (b_2) for Various Aggregation Levels (Simulated Results)

	Aggregation Level		
Simulated Value for b_2	Monthly (No Aggregation)	Semiannual	Annual
.30	.27	.03*	−.02*
.45	.43	−.07*	.04*
.60	.58	.22	.10*
.75	.71	.23	.16*
.90	.89	.68	.48

*The estimate is not significant at the .05 level.

autocorrelated disturbances in the micro model. The combination of autocorrelated disturbances and lagged Y values will cause OLS estimates of the c's in (5) to be inconsistent estimates of the b's in (7).

To illustrate how severe the bias can be as the aggregation level increases, a random series of 1,200 (say monthly) normally distributed observations was generated. The simulated model was

$$y(t) = b_1 x(t) + \epsilon(t), \tag{8}$$

$$\epsilon(t) = b_2 \, \epsilon(t - 1) + u(t), \tag{9}$$

$$u(t) \sim N(0, 10{,}000), \tag{10}$$

$$t = 1, 2, \ldots, 1{,}200.$$

The parameter b_1 was arbitrarily set at 1.2 while the parameter b_2 was allowed to vary between 0.30 and 0.90 by increments of .15. The 1,200 monthly observations were then aggregated into two series of semiannual and annual observations, respectively, and the parameters b_1 and b_2 in (8)–(9) estimated by an application of OLS to equation (3). The resulting parameter estimates for b_2 are summarized in table 1.

It is interesting to note how the serial correlation simulated in the disaggregate data has been "washed out" by the aggregation process. Even for a value of b_2 as high as 0.75, a researcher, working with annual data, would not find a significant estimate for b_2.[3]

Iterative GLS Estimation in the Presence of Temporal Aggregation

To ease notation, but without loss of generality, consider the following simple model:

$$y(t) = \beta \, x(t) + \epsilon(t), \qquad t = 1, 2, \ldots, n; \tag{11}$$

3. These simulation results also hold for the geometric and partial adjustment form of the distributed lag models (Weiss and Windal 1977).

where $y(t)$ stands for some dependent variable; $x(t)$ stands for some nonstochastic independent variable; $\epsilon(t)$ is a disturbance term characterized by a first-order autocorrelation scheme, that is, $\epsilon(t) = \rho\, \epsilon(t-1) + \eta_t$, $\eta_t \sim N(0, \sigma_\eta^2)$; and β and ρ are parameters to be estimated from the data. As a result of this last assumption, the variance-covariance matrix of the disturbance terms is well known (Johnston 1972) and is

$$\Omega = E(\epsilon\epsilon') = \frac{\sigma_\eta^2}{1-\rho^2} \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-2} \\ \vdots & & & & & \\ \rho^{n-1} & \rho^{n-2} & \dots & & & 1 \end{bmatrix}, \tag{12}$$

where $\epsilon' = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$.

Upon aggregation over r time periods, equation (1) can be written as

$$1'Y(t) = \beta\; 1'X(t) + 1'U(t),\; t = 1, 2, \dots, n/r, \tag{13}$$

where

$$1'_{1xr} = (1, 1, \dots, 1), \tag{14}$$

$$Y(t) = \begin{bmatrix} y(tr) \\ y(tr-1) \\ \vdots \\ y(tr-r+1) \end{bmatrix},\; X(t) = \begin{bmatrix} x(tr) \\ x(tr-1) \\ \vdots \\ x(tr-r+1) \end{bmatrix},\; U(t) = \begin{bmatrix} \epsilon(tr) \\ \epsilon(tr-1) \\ \vdots \\ \epsilon(tr-r+1) \end{bmatrix}.$$

To obtain consistent estimates of the parameters β and ρ from the aggregate data only [i.e., $1'\,Y(t)$ and $1'\,X(t)$], it is necessary to compute the generalized least-squares (GLS) estimator with the proper variance-covariance matrix. The variances are given by $\sigma_U^2 = E\{[1'U(t)][1'U(t)]'\} = 1'E[U(t)U'(t)]1 = 1'\;\Omega\;1$, where Ω is now assumed to be of dimension $r \times r$. Consequently σ_U^2 is the sum of all the elements of Ω. Thus,

$$\sigma_U^2 = \frac{\sigma_\eta^2}{1-\rho^2}\,[r + 2\,(r-1)\,\rho + 2\,(r-2)\,\rho^2 + \dots + 2\,\rho^{r-1}] \tag{15}$$

$$= \frac{a\,\sigma_\eta^2}{1-\rho^2}, \tag{16}$$

where

$$a = \frac{r\,(1+\rho)}{(1-\rho)} - \frac{2\rho\,(1-\rho^r)}{(1-\rho)^2}. \tag{17}$$

Similarly, the covariances can be expressed as:

$$\gamma_k = E\{[1'U(t)][1'U(t+k)]'\} = 1'E[U(t)U'(t+k)]1. \tag{18}$$

To carry on the derivation, one needs to form

$$\Sigma = E[U(t)U'(t+k)]$$

$$= \mathrm{E}\left\{\begin{bmatrix}\epsilon(tr)\\ \epsilon(tr-1)\\ \vdots \\ \epsilon(tr-r+1)\end{bmatrix}[\epsilon\,(tr-kr)\ \ldots\ \epsilon\,(tr-kr-r+1)]\right\}$$

$$= \frac{\sigma_\eta^2}{1-\rho^2}\begin{bmatrix}\rho^{kr} & \rho^{kr+1} & \rho^{kr+2} & \ldots & \rho^{kr+r-1}\\ \rho^{kr-1} & \rho^{kr} & \rho^{kr+1} & \ldots & \rho^{kr+r-2}\\ \vdots & & & & \\ \rho^{kr-r+1} & \rho^{kr-r+2} & \rho^{kr-r+3} & \ldots & \rho^{kr}\end{bmatrix}. \tag{19}$$

Since $\gamma_k = 1' \, \Sigma \, 1$, the covariances are the sum of the elements of Σ:

$$\begin{aligned}\gamma_k &= \frac{\sigma_\eta^2}{1-\rho^2}\rho^{(k-1)r}\,[r\rho^r + (r-1)\,\rho^{r+1} + \ldots + \rho^{2r-1}\\ &\quad + (r-1)\,\rho^{r-1} + (r-2)\,\rho^{r-2} + \ldots + \rho]\\ &= \frac{\sigma_\eta^2}{1-\rho^2}\cdot b\,\rho^{(k-1)r},\end{aligned} \tag{20}$$

where

$$b = \rho\left(\frac{1-\rho^r}{1-\rho}\right)^2. \tag{21}$$

The variance-covariance matrix of the aggregate disturbance terms can therefore be written as

$$\Lambda = \frac{\sigma_\eta^2}{1-\rho^2}\begin{bmatrix}a & b & b\rho^r & b\rho^{2r} & \ldots & b\rho^{(n-2)r}\\ b & a & b & b\rho^r & \ldots & b\rho^{(n-3)r}\\ \vdots & & & & & \vdots\\ b\rho^{(n-2)r} & \ldots & & & & a\end{bmatrix}. \tag{22}$$

As can be readily checked, Λ reduces to Ω when no aggregation takes place, that is, when $r = 1$. Note that the quantity a represents the ratio of two variances. That is, $a = \sigma_U^2/\sigma_\epsilon^2$, where σ_U^2 = variance of the aggregate disturbances and σ_ϵ^2 = variance of the disaggregate disturbances.

As one would intuitively expect, the variance of the aggregate disturbances increases with the aggregation level: The greater the aggregation level, the larger the value for the ratio a as illustrated by table 2 for various levels of ρ and r.

This increase in the variance of the disturbances brought about by temporal data aggregation is remarkably visible for high values of ρ and r. An aggregation level of $r = 12$ combined with an autocorrelation coefficient of $\rho = .9$ would generate close to a 100-fold increase in the variance of the aggregate disturbances as compared with that of the corresponding disaggregate disturbances.

TABLE 2 **Values for the Ratio a (Eq. [17]) as a Function of the Autocorrelation Coefficient ρ and the Aggregation Level r**

	r		
ρ	3	6	12
.1	2.25	7.09	14.42
.5	5.50	14.06	32.00
.9	8.22	29.66	98.84

Interestingly enough, the autocorrelation between the disturbances of two successive periods decreases sharply as the aggregation level increases. Indeed,

$$\frac{E[\epsilon(t)\ \epsilon(t+1)]}{E[\epsilon(t)^2]} = \rho, \tag{23}$$

whereas

$$\frac{E\{[1'U(t)](1'U(t+1)]\}}{E\{[1'U(t)]^2\}} = \frac{\rho\,(1-\rho^r)^2}{\mathrm{r}\,(1-\rho^2) - 2\rho\,(1-\rho^r)}. \tag{24}$$

For low values of ρ and high aggregation level, the autocorrelation coefficient between the aggregate disturbances of two successive periods is approximately equal to ρ/r. Even for a ρ value as high as .7, the corresponding value for the aggregate disturbances ($r = 12$) is a mere .144 as illustrated in table 3.

As a result, the variance-covariance matrix of the aggregate disturbances tends to the identity matrix for low to moderate values for ρ and high aggregation level. This theoretical result explains the simulation results presented in table 1.

A brief outline of an iterative GLS procedure to estimate the model parameters from the aggregate data alone is now presented.

Estimation Procedure

Given Λ, the variance-covariance matrix of the aggregate disturbances, the generalized least-squares (GLS) estimator can be computed from

$$b = (X'\ \Lambda^{-1}\ X)^{-1}\ X'\ \Lambda^{-1}\ Y, \tag{25}$$

TABLE 3 **Theoretical Value for the Autocorrelation between Aggregate Disturbances for Various Values of ρ and r (Computed from Eq. [14])**

	r		
ρ	3	6	12
.1	.036	.017	.009
.5	.278	.138	.062
.7	.495	.299	.144
.9	.804	.666	.469

where $Y' = [1' \, Y(1), 1' \, Y(2), \ldots, 1' \, Y(n/r)], X' = [1' \, X(1), 1' \, X(2), \ldots, 1' \, X(n/r)]$, and b is the GLS estimate of the parameter β in (1) and (3).

In general, of course, neither the aggregation level r nor the value of the autocorrelation coefficient ρ are known, and so the GLS estimates cannot be computed directly. They may, however, be approximated by various methods such as the iterative procedure outlined below once r is specified.[4] This estimation procedure consists of the following steps:

Step 1: Compute a preliminary estimate of β, say b^1, from an application of OLS to equation (13).

Step 2: From the corresponding residuals $u_t^1 = 1' \, Y(t) - b^1 1' \, X(t)$, compute a consistent estimate of the first-order autocorrelation coefficient between the residuals, say $R(b^1)$. Such a consistent estimate is given by Theil (1971) as

$$R(b^1) = \left(\frac{1}{T-1} \Big/ \frac{1}{T-k}\right) \left\{ \left[\sum_{t=1}^{T-1} u_t^1 \, u'(t+1) \right] \Big/ \left[\sum_{t=1}^{T} (u_t^1)^2 \right] \right\},$$

where T is the number of aggregate observations and k is the number of independent variables.

Step 3: Solve the polynomial in ρ obtained by setting the theoretical autocorrelation coefficient of the aggregate disturbances given by (14) equal to the estimated value $R(b^1)$. In case of multiple roots,[5] choose that root which minimizes the residuals, say ρ^1.

Step 4: Compute Λ^1 from ρ^1 and r and obtain a new estimate for β, say b^2, by substituting Λ^1 in equation (25).

Step 5: Back to step 2 with the new estimate b^2 until successive estimates for Λ differ by arbitrarily small amounts.

This estimation procedure will be illustrated below with the now classic Lydia Pinkham annual data.

An Empirical Example

The data consist of 54 annual observations covering the years 1907–60 of domestic sales and advertising expenditures of the Lydia E. Pinkham Medicine Company.[6] For purposes of illustration, the ACE model is assumed to be a correct specification of the sales-advertising relationship, that is,

$$z_t = \beta_0 + \beta_1 \, a_t + u_t, \qquad t = 1, 2, \ldots, 54, \tag{26}$$

4. The aggregation level parameter r has to be specified prior to the estimation procedure. For some product classes, such as that of consumer goods, average interpurchase time may help determine the appropriate data collection interval.

5. There are $(2r + 1)$ roots, not all of them real or less than unity in absolute value. However, there may be more than one real foot falling within the unit interval. No such multiple roots were encountered in practice.

6. For detailed description and extensive analyses of this particular data base, see Palda (1964), Clarke and McCann (1973), and Weiss et al. (1978).

TABLE 4 Parameter Estimates of the Current Effects Model as a Function of the Aggregation Level: Lydia Pinkham Annual Data

Parameter Estimates (GLS)	Aggregation Level r: 1	2	3	4	6	12
$\hat{\beta}_0$	1,084.34	1,173.09	1,197.88	1,207.82	1,215.43	1,220.65
$\hat{\beta}_1$	.642	.561	.537	.527	.519	.514
$\hat{\rho}$	.907	.953	.967	.975	.983	.991

NOTE.—Since the data are annual data, a value of $r = 1$ indicates that the postulated relation (eq. [26]) is assumed to hold at the annual level. In contrast, a value of $r = 12$ would be appropriate if one suspects that the relation between sales and advertising holds but at the monthly level. In that case, our estimation procedure takes into account the discrepancy between what is required (monthly data) and what is available (annual data) by using the appropriate variance-covariance matrix.

where z_t and a_t denote the annual sales and advertising expenditures for year t, respectively, and the disturbances u_t are assumed to follow the variance-covariance matrix expressed in (22). Since we are not sure which time period (monthly, quarterly, etc.) is appropriate for the base relationship in the Lydia Pinkham case, the parameters of the ACE model were estimated under several hypotheses regarding the "true-model" time period. Estimates for β_0, β_1, and ρ are listed for six different true-model time periods by aggregation level. For example, if the true-model time period is hypothesized to be 1 year, the aggregation level (r) is 1. If the true-model time period is hypothesized to be 1 month, then for annual data, $r = 12$. Thus, the estimates in table 4 are conditional estimates, the condition being the assumed true-model time period.

Discussion

The empirical results of our example are consistent with relevant empirical results reported elsewhere for the ACE model. When the aggregation level is one for the annual data our estimate for ρ is almost identical to that reported for the same model and data by Weiss et al. (1978). In addition, the findings reported in table 4 are consistent with Clarke's (1976) findings that temporal misspecification in the ACE model leads to overestimates of the parameter for current advertising effects (β_1) and underestimates of the autocorrelation parameter (ρ). In addition, the estimates for ρ (table 4) meet the requirement for directional consistency noted by Clarke (1976) when he wrote, ". . . regardless of what the actual lag structure is, consistent estimation with different data intervals would require that the estimates of ρ (our notation) be larger for shorter data intervals than for longer data intervals."[7]

7. Several of the empirical findings surveyed by Clarke, however, did not show this required directional consistency.

The point to be emphasized in this paper, however, is that annual data do contain usable and recoverable information about the advertising relationship and should not be rejected out of hand. For the model described by (11) and (12) (of which the ACE model is a special case), consistent estimates can be recovered from aggregate data provided one is able to specify an appropriate value for r, the aggregation level. Rarely, however, is a researcher fortunate enough to possess clear-cut evidence of what is the appropriate data interval for the problem at hand. Theoretical studies dealing with this issue are hard to come by but would greatly aid in the assessment of studies bearing upon the cumulative effects of advertising expenditures.

Conclusion

An iterative generalized least-squares procedure was developed to estimate the parameters of the so-called current effects model with data aggregated over time. The variance-covariance matrix of the aggregate disturbances was explicitly derived from that of the disturbances of the disaggregate model. Acceptance or rejection of the model was seen to depend crucially on the data collection interval being utilized. The results obtained in this study raise interesting questions concerning some of the results (and their interpretation) reported in studies of cumulative advertising effects measurement.

References

Bass, Frank M., and Clarke, Darral G. 1972. Testing distributed lag models of advertising effect. *Journal of Marketing Research* 9 (August): 298–308.

Clarke, Darral G. 1976. Econometric measurement of the duration of advertising effects on sales. *Journal of Marketing Research* 13 (November): 345–57.

Clarke, Darral G., and McCann, John M. 1973. Measuring the cumulative effects of advertising: a reappraisal. In T. V. Greer (ed.), *Combined Proceedings*. Chicago: American Marketing.

Houston, Franklin S., and Weiss, Doyle L. 1975. Cumulative advertising effects: the role of serial correlation. *Decision Sciences* 6 (July): 471–88.

Johnston, J. 1972. *Econometric Methods*. 2d ed. New York: McGraw-Hill.

Morigouchi, C. 1970. Aggregation over time in macroeconomic relations. *International Economics Review* 11 (October): 427–40.

Mundlak, Y. 1961. Aggregation over time in distributed lag models. *International Economics Review* 2 (May): 154–63.

Palda, Kristian. 1964. *The Measurement of Cumulative Advertising Effects*. Englewood Cliffs, N.J.: Prentice-Hall.

Rowe, Robert D. 1976. The effects of aggregation over time on t-ratios and R^2's." *International Economic Review* 17 (October): 751–57.

Theil, H. 1954. *Linear Aggregation of Economic Relations*. Amsterdam: North-Holland.

Theil, H. 1971. *Principles of Econometrics*. New York: Wiley.

Weiss, Doyle L.; Houston, Franklin S.; and Windal, Pierre M. 1978. The periodic pain of Lydia E. Pinkham. *Journal of Business* 51 (January): 91–101.

Weiss, Doyle L., and Windal, Pierre M. 1977. The effects of specification error and temporal data aggregation on distributed lag models of advertising sales effectiveness—a Monte Carlo approach. Paper presented at the Ninth Annual Conference of the American Institute of Decision Sciences, Chicago.