ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B, 00120 Helsinki 12, Finland, tel. 601322

Keskusteluaiheita
Discussion papers

Timo Teräsvirta

TEMPORAL AGGREGATION IN FINITE DISTRIBUTED LAG MODELS*

No. 70 25.11.1980

* Part of this research was carried out while the author was visiting CORE, Université Catholique de Louvain, Louvain-la-Neuve. He wishes to thank Wolfgang Birkenfeld, Ilkka Mellin, Markku Rahiala, Jean-François Richard and Yrjö Vartia for help and comments at various stages of the work but retains the responsibility for any errors. Financial support from the Yrjö Jahnsson Foundation is gratefully acknowledged.

Abstract. When a finite linear distributed lag model is aggregated over time and only aggregated observations are assumed to be available, not all the lags of the original model are observable. This paper is concerned with the prediction of the temporally aggregated dependent variable using both the aggregated model in which the unobservable lags are missing and the original model. It also considers the estimation of the total response of the dependent variable to a change in the independent variable using the two models. Necessary and sufficient conditions for the predictions/estimates from the aggregated model to be superior to those from the original one are derived and their properties discussed in the light of a simple example.

Keywords: biased estimation; distributed lag; improved estimation; quadratic risk; temporal aggregation.

1. Introduction

Sometimes the unit time interval thought to be appropriate in modelling economic relationships may be too short in the sense that there are no observations available for estimating the parameters of the model. In such a case building the model on temporally aggregated data involves a loss of information and the properties of the model will be affected by the aggregation, for discussion see e.g. Zellner and Montmarquette (1971).

Since the economic relationships are often dynamic, at least if the appropriate time unit is sufficiently short, the models may contain distributed lags. An early discussion of temporal aggregation in distributed lag models can be found in Theil (1954), Chapter 4. Mundlak (1961) and Engle and Liu (1972) have considered geometric distributed lag schemes, while Brewer (1973) has shown how the rational distributed lag structures change due to temporal aggregation. Sims (1971) and Geweke (1978) have studied temporal aggregation in connection with distributed lags starting from a continuous time model.

Tiao and Wei (1976) have addressed the problem of estimating the dynamic relationship between two time series using temporally aggregated models. They have concluded by some examples that the loss of accuracy may be substantial in parameter estimation but less severe in prediction. The unbiased estimation of finite distributed lag models from temporally aggregated data has been studied by Wei (1978) who pointed out specific reasons for the loss of efficiency due to temporal aggregation. A recent paper by Wei and Mehta (1979) contains some results on forecasting the temporally aggregated dependent variable using Monte Carlo techniques.

This paper discusses the circumstances under which temporal aggregation in distributed lag models before estimation does not imply reduced prediction or estimation accuracy as compared to temporal aggregation after estimation. The organisation of the paper is as follows: The distributed lag model is defined in Section 2 while the next section discusses its temporal aggregation. Section 4 contains the theoretical results of the paper which are illustrated by virtue of an example in Section 5. Finally, Section 6 offers a brief comment on some results of Wei and Mehta (1979) in the light of the present findings.

2. The distributed lag model

We assume a finite distributed lag model

$$y_t = \sum_{j=0}^{pk-1} \beta_j x_{t-j} + \varepsilon_t, \varepsilon_t \sim n(0,\sigma^2), \ t=1,\ldots,nk, \ E\varepsilon_t\varepsilon_s = 0, \ t \neq s \quad (1)$$

or, in matrix form

$$y = X\beta + \varepsilon, \ \varepsilon \sim N(0,\sigma^2 I_{nk}) \quad (2)$$

where y and ε are stochastic $nk \times 1$ vectors, X is an $nk \times pk$ observation matrix with full column rank and β is a $pk \times 1$ parameter vector. Further assume that $k > 1$. Temporal aggregation means aggregating k subsequent observations to a new observation. Letting $A = (I_n \otimes 1_k')$, where $1_k = (1,1,\ldots,1)'$ is a $k \times 1$ vector, the temporally aggregated version of (2) can be written as

$$Ay = AX\beta + A\varepsilon \tag{3}$$

where

$$A\varepsilon \sim N(0,\sigma^2 kI_n) \text{ because } AA' = kI_n.$$

Assume now that only aggregated time series are available. Since the columns of X are lags of the first column the matrix of aggregates AX is not wholly observable. In fact, only

$$AX_1 = A(x_1, x_{k+1},\ldots,x_{kj+1},\ldots,x_{k(p-1)+1})$$

where x_j is the j^{th} column of X, is observed, while

$$AX_2 = A(x_2,x_3,\ldots,x_k,x_{k+2},\ldots,x_{pk})$$

remains unobserved. Dividing β into two blocks conformably with X, the aggregated model (3) can be written as

$$Ay = AX_1\beta_1 + \varepsilon^* \tag{4}$$

where

$$\varepsilon^* = AX_2\beta_2 + A\varepsilon.$$

Model (4) is thus a misspecified version of (2).

3. Temporally aggregated models

A model builder with only temporally aggregated time series seldom thinks in terms of β and model (4) but rather has in mind the following model

$$y^* = X^*\gamma^* + \varepsilon^{**} \tag{5}$$

where $y^* = Ay$, $X^* = (AX_1, AX_3)$, $X_3 = (x_{kp+1}, x_{(k(p+1)+1}, \ldots, x_{k(p+\ell-1)+1})$, $\gamma^* = \underset{n\to\infty}{\text{plim}}(X^{*\prime}X^*)^{-1}X^{*\prime}y^*$ is a $(p+\ell) \times 1$ parameter vector and the elements of ε^{**} are i.i.d. with expectation zero and uncorrelated with X^*. The matrix AX_3 contains observable temporally aggregated lags longer than those in AX_1. The model builder then wants to estimate γ^* and/or predict y^*.

In this paper, attention will also be paid to a special case of (5),

$$y^* = AX_1\gamma_1^* + \varepsilon_1^* \tag{6}$$

which actually is the same model as (4). Model (6) offers a natural interpretation for γ_1^* in terms of β which is not so straightforward for γ^* in the general case (5), namely $\gamma_1^* = (I_p \otimes 1_k')\beta$. Formally (4) implies $\gamma_1^* = \beta_1$ and $\varepsilon_1^* = \varepsilon^*$, but the parameters are usually not interpreted that way by a model builder who starts from (5).

Before proceeding further, we stress the obvious fact that the i.i.d. assumptions made about ε^{**} or ε_1^* do not hold, because the errors are as a rule autocorrelated and furthermore correlated with X^*, see Griliches (1972), p. 734, for a specific example. In fact, (2) being the true model, these assumptions only hold if $\{x_t\}$ is white noise with expectation zero.

It can be pointed out here that Wei (1978) and Wei and Mehta (1979) retain (3) as the basic model and estimate the unknown AX_2. The resulting model is called the conditional aggregate model

$$y^* = AX_1\beta_1 + \hat{X}_2^A\beta_2 + u \qquad (7)$$

where $\hat{X}_2^A$ is an estimator of AX_2, and u is an error vector independent of $\hat{X}_2^A$ and consisting of ε and the error caused by the estimation of AX_2. The parameter vector β_2 can be estimated consistently and unbiasedly from (7), for details see Wei (1978).

4. Comparisons between models

In order to consider the effects of temporal aggregation on the prediction and estimation of parameters we choose (5) as our starting-point and estimate γ^* using OLS. The least squares estimator of γ^* is

$$g = M_{13\cdot 13}^{-1}X_{13}'A'Ay = \beta_1 + M_{13\cdot 13}^{-1}M_{13\cdot 2}\beta_2 + M_{13\cdot 13}^{-1}X_{13}'A'\varepsilon^*$$

where

$$X_{13} = (X_1, X_3), M_{13 \cdot 13} = X_{13}'A'A\,X_{13} \text{ and}$$

$$M_{13\cdot 2} = X_{13}'A'AX_2 = (M_{21}, M_{23})'$$

with

$$M_{ij} = X_i'A'AX_j, \quad i,j = 1,2,3.$$

The interest will now be focussed on two problems. First, we want to study the circumstances under which the temporally aggregated dependent variable can be predicted better using the misspecified model (5) than the correct original (2). Second, it will be investigated when the total response $1_{pk}'\beta$ is more accurately estimated by $1_{p+\ell}'g$ than by the best linear unbiased estimator $1_{pk}'b$.

4.1. Prediction of independent variable

Since we are interested in predicting y^* it is feasible to consider its conditional expectation $E(y^*|X,\beta,\sigma^2) = AX\beta$ and ask how well it can be estimated from (2) and (5), respectively. Let us define a criterion for superiority of a predictor over another by saying that of two predictors the one with smaller predictive mean square error is superior to the other. Hence we want to know when

$$\begin{aligned} \text{pmse}(AXb) - \text{pmse}(AX_{13}g) &= E(b-\beta)'X'A'AX(b-\beta) \\ &- E(AX_{13}g - AX\beta)'(AX_{13}g - AX\beta) \geq 0. \end{aligned} \tag{8}$$

/

We have

$$\text{pmse}(AXb) = \sigma^2 \text{tr} X'A'AX(X'X)^{-1}$$

whereas deriving $\text{pmse}(AX_{13}g)$ requires slightly more computation. Write

$$AX_{13}g - (AX_1\beta_1 + AX_2\beta_2)$$

$$= (AX_{13}M^{-1}_{13\cdot 13}X'_{13}A'-I)(AX_1\beta_1 + AX_2\beta_2)$$

$$+ AX_{13}M^{-1}_{13\cdot 13}X'_{13}A'\varepsilon^*$$

$$= [AX_1M^{-1}_{11}X'_1A' + (I-AX_1M^{-1}_{11}X'_1A')AX_3D^{-1}_{33\cdot 1}X'_3A'(I-AX_1M^{-1}_{11}X'_1A')-I]$$

$$\times (AX_1\beta_1 + AX_2\beta_2) + AX_{13}M^{-1}_{13\cdot 13}X'_{13}A'\varepsilon^* \qquad (9)$$

where

$$D_{33\cdot 1} = M_{33}-M_{31}M^{-1}_{11}M_{13}.$$

Then, using (9) and noting that

$$(AX_1M^{-1}_{11}X'_1A'-I)AX_1 = 0$$

we obtain, after a straightforward calculation,

$$\text{pmse}(AX_{13}g) = \beta'_2(D_{22\cdot 1}-D_{23\cdot 1}D^{-1}_{33\cdot 1}D_{32\cdot 1})\beta_2+\sigma^2 k(p+\ell) \qquad (10)$$

where

$$D_{ij\cdot k} = M_{ij}-M_{ik}M^{-1}_{kk}M_{kj}.$$

The first term on the r.h.s. of (10) results from omitting AX_2 in (5) while the second is the variance term. Now, (8) holds if and only if

$$(1/\sigma^2)\beta_2'(D_{22\cdot1}-D_{23\cdot1}D_{33\cdot1}^{-1}D_{32\cdot1})\beta_2 \leq tr(X'X)^{-1}X'A'AX-k(p+\ell). \quad (11)$$

A necessary condition for (11) to hold is

$$tr(X'X)^{-1}X'A'AX - k(p+\ell) \geq 0$$

indicating the point after which adding new lags into the temporally aggregated model never pays off in terms of the predictive mean square error.

If our model is (6), i.e. a direct result of aggregating (2) and omitting AX_2, (11) takes the form

$$(1/\sigma^2)\beta_2'D_{22\cdot1}\beta_2 \leq tr(X'X)^{-1}X'A'AX - kp. \quad (12)$$

On the other hand, forming $pmse(AX_1g_1) - pmse(AX_{13}g)$, where $g_1 = M_{11}^{-1}X_1'A'Ay$, we can conclude that $AX_{13}g$ is superior to AX_1g_1 as a predictor of $AX\beta$ if and only if

$$(1/\sigma^2)\beta_2'D_{23\cdot1}D_{33\cdot1}^{-1}D_{32\cdot1}\beta_2 \leq k\ell$$

which shows exactly when a longer lag than implied by (6) improves the prediction accuracy in predicting the aggregated quantity.

In general, we can conjecture from (11) that temporal aggregation may improve our prediction performance if the model (2) is not a very accurate description of y, i.e. if σ^2 is not small with respect to the absolute values of the components of β_2. It is also obvious that if $\beta_2 \neq 0$, (11) can only hold in small samples, because $AX_{13}g$ is then generally an inconsistent estimator of $AX\beta$. Of course, if $\beta_2=0$, then (11) holds trivially, but this is a rare case in practice. The validity of (11) is also dependent on the covariance structure of the input process $\{x_t\}$, but (11) as such does not reveal anything very specific about the nature of this dependence. We shall return to this point in the example of Section 5.

4.2. Estimation of total response

The accuracy in estimating total response $1'_{pk}\beta$ is measured here using the mean square error. We need

$$\mathrm{mse}(1'_{pk}b) = E(1'_{pk}b-1'_{pk}\beta)^2 = \sigma^2 1'_{pk}(X'X)^{-1}1_{pk}$$

and

$$\mathrm{mse}(1'_{p+\ell}g) = E(1'_{p+\ell}g-1'_{pk}\beta)^2$$

$$= E(1'_{p+\ell}M^{-1}_{13\cdot13}X'_{13}A'Ay-1'_{pk}\beta)^2$$

$$= E(1'_{p+\ell}M^{-1}_{13\cdot13}(M_{21},M_{23})'\beta_2-1'_{p+\ell}M^{-1}_{13\cdot13}X'_{13}A'\varepsilon^*-1'_{p(k-1)}\beta_2)^2$$

$$= (1'_{p+\ell}M^{-1}_{13\cdot13}(M_{21},M_{23})'\beta_2-1'_{p(k-1)}\beta_2)^2+\sigma^2k1'_{p+\ell}M^{-1}_{13\cdot13}1_{p+\ell}.$$

Thus $\mathrm{mse}(1'_{pk}b) - \mathrm{mse}(1'_{p+\ell}g) \geq 0$ if and only if

$$(1/\sigma^2)[1'_{p+\ell}M^{-1}_{13\cdot 13}(M_{21},M_{23})'\beta_2 - 1'_{p(k-1)}\beta_2]^2$$

$$\leq 1'_{pk}(X'X)^{-1}1_{pk} - k1'_{p+\ell}M^{-1}_{13\cdot 13}1_{p+\ell}. \tag{13}$$

Dividing $1_{p+\ell}$ into two vectors conformably with $X_{13} = (X_1,X_3)$, inequality (13) can be written as

$$(1/\sigma^2)[1'_pM^{-1}_{11}M_{12}\beta_2-d'_\ell D^{-1}_{33\cdot 1}D_{32\cdot 1}\beta_2-1'_{p(k-1)}\beta_2]^2$$

$$\leq 1'_{pk}(X'X)^{-1}1_{pk}-k[1'_pM^{-1}_{11}1_p+d'_\ell D^{-1}_{33\cdot 1}d_\ell] \tag{14}$$

where

$$d_\ell = M_{31}M^{-1}_{11}1_p-1_\ell.$$

Again, a necessary condition for (13) to hold is that the r.h.s. of the inequality be positive. Likewise, it can be conjectured that σ^2 should not be small with respect to β_2 if (13) is hoped to hold and that the sample should not be too large if $\beta_2 \neq 0$. If our estimator is $1'_pg_1$, the equivalent of (13) becomes, cf. (14),

$$(1/\sigma^2)(1'_pM^{-1}_{11}M_{12}\beta_2-1'_{p(k-1)}\beta_2)^2 \leq 1'_{pk}(X'X)^{-1}1_{pk}-k1'_pM^{-1}_{11}1_p. \tag{15}$$

Finally, adding lags into the aggregated model improves the estimation as compared to $1'_pg_1$ if and only if

$$(1/\sigma^2)[(d_\ell' D_{33\cdot1}^{-1} D_{32\cdot1}\beta_2)^2 - 2d_\ell' D_{33\cdot1}^{-1} D_{32\cdot1}\beta_2$$

$$\times\ (1_p' M_{11}^{-1} M_{12}\beta_2 - 1_{p(k-1)}'\beta_2)] \leq d_\ell' D_{33\cdot1}^{-1} d_\ell$$

which can only hold in small samples if $\beta_2 \neq 0$ since the r.h.s. decreases with sample size.

4.3. Estimating the aggregated impulse response function

As mentioned shortly, model (6) allows for an interpretation of γ_1^* as a vector of aggregated components of β. The components of γ^* can then be regarded as points of the temporally aggregated impulse response function of y. Proceeding as above, we can easily derive a necessary and sufficient condition for this vector to be estimated better using g_1 than Bb, where $B = (I_p, C)$ with $C = (I_p \otimes 1_{k-1}')$, if the superiority criterion is again the mean square error. The condition becomes

$$(1/\sigma^2)\beta_2'(M_{11}^{-1}M_{12} - C)'(M_{11}^{-1}M_{12} - C)\beta_2 \leq tr(X'X)^{-1}B'B - k\,trM_{11}^{-1}.$$

All the above conditions depend on the unknown parameters of the true model and the unobservable X_2. If X_2 were observed, the validity of the conditions could be tested as discussed in Toro-Vizcarrondo and Wallace (1968). This is a hypothetical case, however, and in order to assess the practical significance of conditions (11) and (13) we have chosen to illustrate them with an example demonstrating when we might expect to gain from temporal aggregation in predicting the temporally aggregated independent variable or estimating the total response.

5. Example

5.1. Prediction of dependent variable

Assume that (2) is the following monthly model

$$y_t = \beta_1 x_t + \beta_{21} x_{t-1} + \beta_{22} x_{t-2} + \varepsilon_t, \quad t=1,\ldots,3n \tag{16}$$

while its quarterly counterpart (5) is chosen to be

$$y_t^* = \gamma_1^* x_t^* + \gamma_2^* x_{t-1}^* + \varepsilon_t^*, \quad t=1,\ldots,n \tag{17}$$

so that $k = 3$, $\ell = 1$ and $p = 1$. We also consider the alternative that $\gamma_2^* = 0$ corresponding to (6). Further assume that $x_t \sim AR(1)$, as may often be the case in economic time series. Let $Ex_t = 0$, $cov(x_t) = \sigma_x^2$, and let ρ be the AR-parameter, $|\rho| < 1$. The assumptions concerning ε_t remain as in Section 2, and

$$\varepsilon_t^* = \beta_{21} x_{3t-1} + \beta_{22} x_{3t-2} + \varepsilon_{3t} + \varepsilon_{3t-1} + \varepsilon_{3t-2}, \quad t=1,\ldots,n.$$

In order to obtain illustrative numerical results without extensive computations the moment matrices in (11) and (13) are replaced by their probability limits multiplied by the number of observations. We have

$$\operatorname*{plim}_{n\to\infty} [(1/3n)(X'X)]^{-1} = \sigma_x^{-2}(1-\rho^2)^{-1} \begin{bmatrix} 1 & -\rho & 0 \\ & 1+\rho^2 & -\rho \\ & & 1 \end{bmatrix} \qquad (18)$$

see Theil (1971), p. 252. Furthermore,

$$\operatorname*{plim}_{n\to\infty}(1/3n)X'A'AX = \sigma_x^2 \begin{bmatrix} m_0 & m_1 & m_2 \\ & m_0 & m_1 \\ & & m_0 \end{bmatrix} . \qquad (19)$$

where $m_j = \operatorname{cov}(\sum_{i=0}^{2} x_{t-i}, \sum_{i=0}^{2} x_{t-j-i})$ and, more specifically,

$$m_0 = 3 + 4\rho + 2\rho^2$$
$$m_1 = 2 + 4\rho + 2\rho^2 + \rho^3$$
$$m_2 = 1 + 2\rho + 3\rho^2 + 2\rho^3 + \rho^4.$$

Using (18) and (19) we obtain the r.h.s. of (11) which equals $\sigma^2(3 + 4\rho + 2\rho^2)$, while the r.h.s. of (12) adds up to $2\sigma^2(3 + 2\rho + \rho^2)$. From (11) we have, approximately,

$$(1/\sigma^2)\operatorname*{plim}_{n\to\infty}\beta_2'(D_{22\cdot 1}-D_{23\cdot 1}D_{33\cdot 1}^{-1}D_{32\cdot 1})\beta_2$$
$$\approx (3n\sigma_x^2/\sigma^2)[\tilde{\alpha}_1^{(1)}(\beta_{21}^2+\beta_{22}^2)+\tilde{\alpha}_2^{(1)}\beta_{21}\beta_{22}] \qquad (20)$$

where

$$\tilde{\alpha}_1^{(1)} = m_0 - [m_0^2-(\rho m_2)^2]^{-1}[m_0(m_1^2+m_2^2)-2\rho m_1 m_2^2] \qquad (21)$$

and

$$\alpha_2^{(1)} = 2\{m_1-[m_0^2-(\rho m_2)^2]^{-1}[2m_0 m_1 m_2-\rho m_2(m_1^2+m_2^2)]\}. \qquad (22)$$

Analogously, the l.h.s. of the superiority condition (12) corresponding to the model (6) has the form

$$(1/\sigma^2)\operatorname*{plim}_{n\to\infty}\beta_2'(M_{22}-M_{21}M_{11}^{-1}M_{12})\beta_2$$

$$\approx (3n\sigma_x^2/\sigma^2)[\tilde{\alpha}_{11}^{(0)}\beta_{21}^2+\tilde{\alpha}_{12}^{(0)}\beta_{22}^2+\tilde{\alpha}_2^{(0)}\beta_{21}\beta_{22}]$$

where

$$\tilde{\alpha}_{1j}^{(0)} = m_0-m_0^{-1}m_j^2, \quad j=1,2$$

and

$$\tilde{\alpha}_2^{(0)} = 2m_1(1-m_0^{-1}m_2).$$

After a standardization of (11) and (12) through dividing by the r.h.s. which in this example is positive in both cases for $-1 < \rho < 1$, (11) and (12) take the approximate form

$$(3n\sigma_x^2/\sigma^2)[\alpha_1^{(1)}(\beta_{21}^2+\beta_{22}^2)+\alpha_2^{(1)}\beta_{21}\beta_{22}] \leq 1 \qquad (23)$$

and

$$(3n\sigma_x^2/\sigma^2)[\alpha_{11}^{(0)}\beta_{21}^2+\alpha_{12}^{(0)}\beta_{22}^2+\alpha_2^{(0)}\beta_{21}\beta_{22}] \leq 1 \qquad (24)$$

where

$$\alpha_j^{(1)} = (3+4\rho+2\rho^2)^{-1}\tilde{\alpha}_j^{(1)}, \quad j=1,2$$

$$\alpha_{1j}^{(0)} = 2^{-1}(3+2\rho+\rho^2)^{-1}\tilde{\alpha}_{1j}, \quad j=1,2$$

and

$$\alpha_2^{(0)} = 2^{-1}(3+2\rho+\rho^2)^{-1}\tilde{\alpha}_2^{(0)}.$$

In order to illustrate conditions (23) and (24), we have plotted $\alpha_1^{(1)}$, $\alpha_2^{(1)}$, $\alpha_{11}^{(0)}+\alpha_{12}^{(0)}$ and $\alpha_2^{(0)}$ in Figure 1. The sum $\alpha_1^{(0)}=\alpha_{11}^{(0)}+\alpha_{12}^{(0)}$ is strictly comparable to $\alpha_1^{(1)}$ only when $\beta_{21}=\beta_{22}$, but the essential features of the situation, when β_{21} and β_{22} have the same sign and are not too far from each other, can be captured more easily if $\alpha_{11}^{(0)}$ and $\alpha_{12}^{(0)}$ are pooled together. For comparison they are plotted separately in Figure 2.

Figure 1 demonstrates the fact that, if β_{21} and β_{22} do have the same sign, gains from temporal aggregation are more likely on average than otherwise if the input variable is heavily positively or very heavily negatively autocorrelated. An intuitive explanation of this is that if the independent variable is heavily autocorrelated then the temporally aggregated lags retained in the model carry almost the same amount of information as the whole set of lags. Hence omitting part of them does not have crucial importance.

It can also be noted that for the most part condition (23) is more easily satisfied in practice than condition (24), indicating that (5) with $\ell=1$ is a better alternative than (6). If β_{21} and β_{22} have opposite signs, (23) is more likely to hold than if they have the same sign, except for an interval of negative values of ρ, see Figure 1. Whether (23) or (24) really holds, depends also on n, σ_x^2, σ^2 and the size of β_{21} and β_{22} as discussed in the preceding section.

5.2. Estimation of total response

Next, consider the estimation of the total response. We obtain

$$\underset{n\to\infty}{\text{plim}}[1'_{p+\ell}M^{-1}_{13\cdot 13}(M_{21},M_{23})'\beta_2-1'_{p(k-1)}\beta_2]^2$$

$$= [(\frac{m_1+m_2}{m_0+\rho m_2}-1)(\beta_{21}+\beta_{22})]^2 \tag{25}$$

and, assuming (6) to be the temporally aggregated model,

$$\underset{n\to\infty}{\text{plim}}[1'_pM^{-1}_{11}M_{12}\beta_2-1'_{p(k-1)}\beta_2]^2 = [(1-m_0^{-1}m_1)\beta_{21}+(1-m_0^{-1}m_2)\beta_{22}]^2. \tag{26}$$

In large samples the r.h.s. of (13) now equals approximately

$$(3n)^{-1}\sigma_x^{-2}[(3-\rho)(1+\rho)^{-1}-6(m_0+\rho m_2)^{-1}]$$

while a similar approximation of the r.h.s. of (15) becomes

$$(3n)^{-1}\sigma_x^{-2}[(3-\rho)(1+\rho)^{-1}-3m_0^{-1}].$$

After standardizing (25) and (26) so as to make the r.h.s.'s of (13) and (15) equal one these conditions become

$$(3n\sigma_x^2/\sigma^2)\alpha_0^{(1)}(\beta_{21}+\beta_{22})^2 \leq 1, \tag{27}$$

where

$$\alpha_0^{(1)} = (\frac{m_1+m_2}{m_0+\rho m_2}-1)^2[(3-\rho)(1+\rho)^{-1}-6(m_0+\rho m_2)^{-1}]^{-1}$$

and

$$(3n\sigma_x^2/\sigma^2)(\alpha_{01}^{(0)}\beta_{21}+\alpha_{02}^{(0)}\beta_2)^2 \leq 1 \tag{28}$$

respectively, where

$$\alpha_{0j}^{(0)} = (1-m_0^{-1}m_j)[(3-\rho)(1+\rho)^{-1}-3m_0^{-1}]^{-1}, \; j=1,2.$$

Figure 3 depicts the behavious of $\alpha_0^{(1)}$ and, for simplicity, the squared sum $\alpha_0^{(0)} = (\alpha_{01}^{(0)}+\alpha_{02}^{(0)})^2$ between $-1 < \rho < 1$. If β_{21} and β_{22} have the same sign then, unlike previously, (6) seems to be a better aggregated alternative than (5). Our chances of gaining from temporal aggregation while using (6) are greatest when ρ lies in the neighbourhood of zero or one, or minus one. In general, a gain is more likely to occur at positive than negative values of ρ.

If β_{21} and β_{22} have different signs then the chances of estimating the total response more accurately from the aggregated model increase as compared to the opposite case. In particular, if $\beta_{21} = -\beta_{22}$, the sum of coefficients, which then equals β_1, is always estimated better by ordinary least squares from (5) than from the original model (2). Whether (27) and (28) hold in the general case depends again on n, σ_x^2, σ^2 and the size of β_{21} and β_{22}, as discussed above and seen from these inequalities.

6. Comments

It can be conjectured that (11) and (13) are sufficient but no longer necessary conditions for the superiority of $AX_{13}g$ over the corresponding least squares predictor from (7). This is because b is the best linear unbiased estimator of β whereas the least squares estimator of the conditional aggregate model, while still unbiased, is not efficient. Wei and Mehta (1979) considered one-step ahead forecasting of y^* by simulation using a somewhat simpler model than (16) with only one lag. The input was an AR(1) process as in the preceding example, and the number of aggregated observations was 90. They found no marked difference between the forecasts from the conditional aggregate and aggregate models when ρ was varied between -0.7 and 0.7. Nevertheless, at $\rho = 0.7$ the latter model already seemed to have an edge as far as the accuracy of predictions was concerned. Judging from the above example, (5) might well perform better than (7) at higher values of ρ, even if one-step ahead forecasting and not within-sample prediction is concerned.

References

Brewer, K.R.W., 1973, Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. Journal of Econometrics 1, 133-154.

Engle, R.F. and T.C. Liu, 1972, Effects of aggregation over time on dynamic characteristics of an econometric model, in: B.G. Hickman, ed., Econometric models of cyclical behavior, Vol. 2 (Columbia University Press, New York) 673-728.

Geweke, J., 1978, Temporal aggregation in the multiple regression model. Econometrica 46, 643-661.

Griliches, Z., 1972, Discussion of the paper by Engle and Liu, in: B.G. Hickman, ed., Econometric models of cyclical behavior, Vol. 2 (Columbia University Press, New York) 733-737.

Mundlak, Y., 1961, Aggregation over time in distributed lag models. International Economic Review 2, 154-163.

Sims, C.A., 1971, Discrete approximations to continuous time distributed lags in econometrics. Econometrica 39, 545-563.

Theil, H., 1954, Linear aggregation of economic relationships (North-Holland, Amsterdam).

Theil, H., 1971, Principles of econometrics (John Wiley, New York).

Tiao, G.C. and W.S. Wei, 1976, Effect of temporal aggregation on the dynamic relationship of two time series variables. Biometrika 63, 513-523.

Toro-Vizcarrondo, C. and T.D. Wallace, 1968, A test of the mean square error criterion for restrictions in linear regression. Journal of the American Statistical Association 63, 558-571.

Wei, W.S., 1978, The effect of temporal aggregation on parameter estimation in distributed lag model. Journal of Econometrics 8, 237-246.

Wei, W.S. and J.S. Mehta, 1979, Temporal aggregation and information loss in a distributed lag model, in: O.D. Anderson, ed., Analysing time series. Proceedings of the international conference held on Guernsey, Channel Islands, October 1979 (North-Holland, Amsterdam) 391-400.

Zellner, A. and C. Montmarquette, 1971, A study of some aspects of temporal aggregation problems in econometric analyses. Review of Economics and Statistics 53, 335-342.

Figure 1. The graphs of coefficients $\alpha_1^{(0)} = \alpha_{11}^{(0)} + \alpha_{12}^{(0)}$, $\alpha_2^{(0)}$, $\alpha_1^{(1)}$ and $\alpha_2^{(2)}$ in (23) and (24)

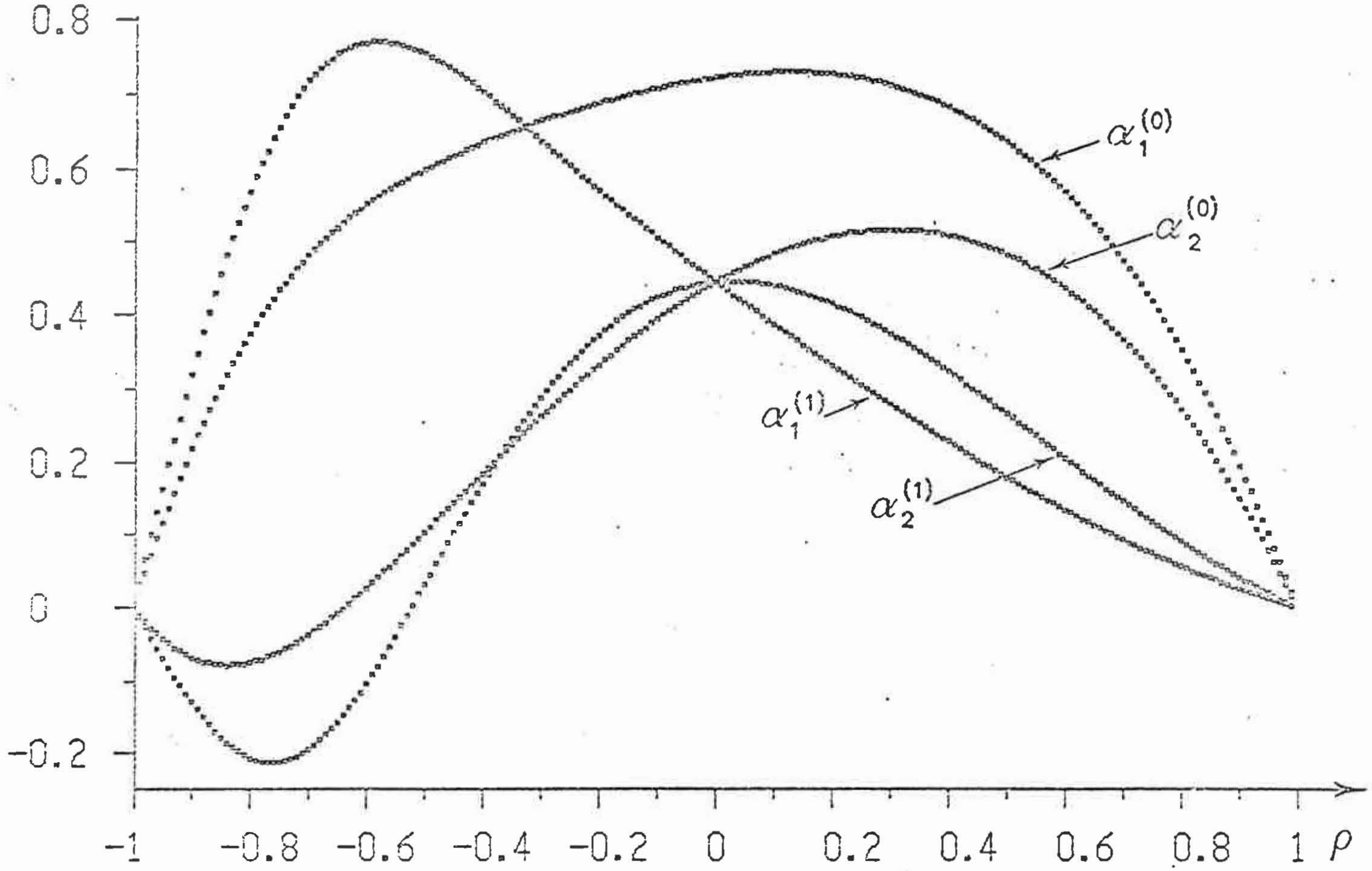

Figure 2. The graphs of coefficients $\alpha_{11}^{(0)}$ and $\alpha_{12}^{(0)}$ in (24)

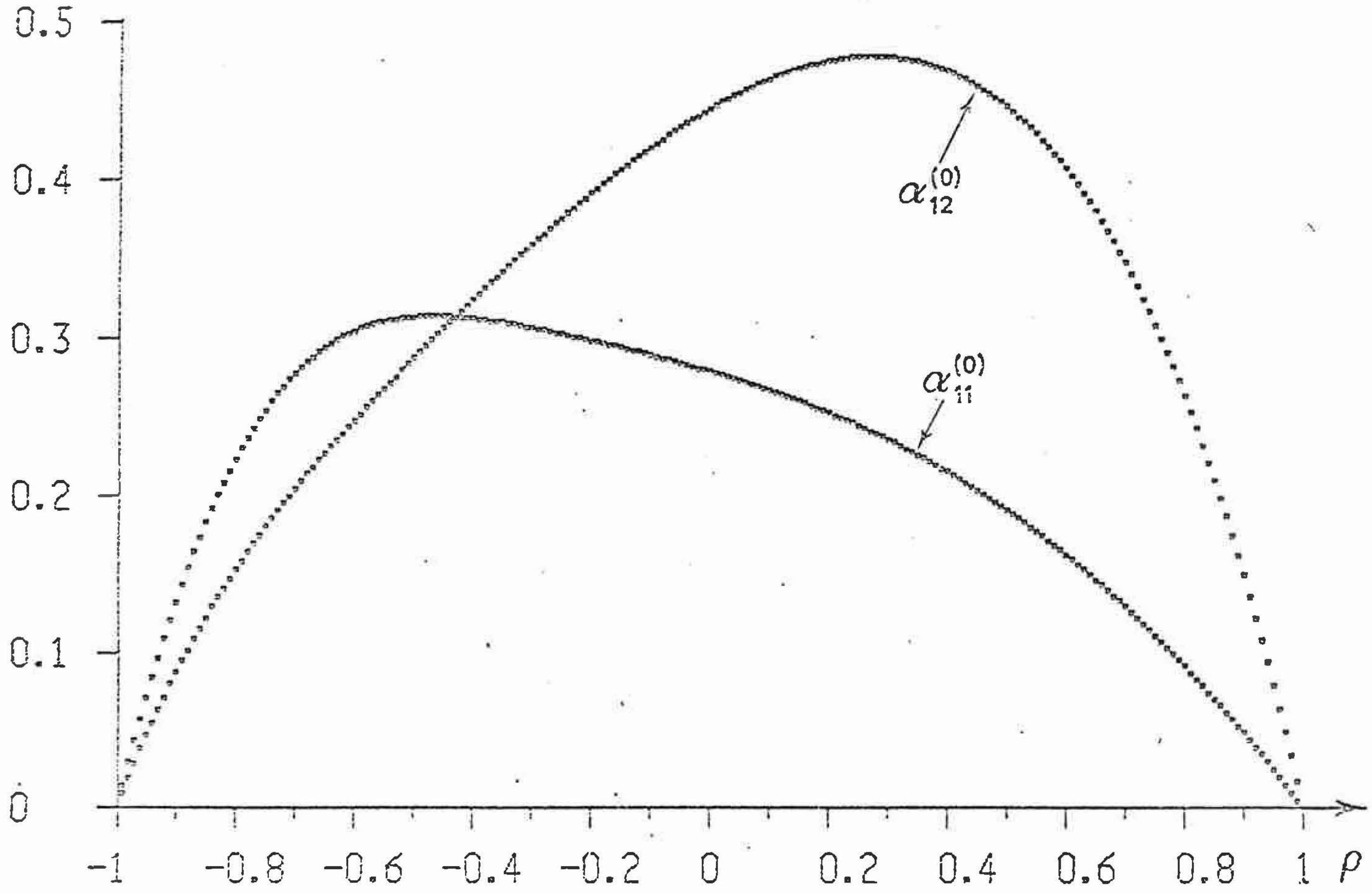

Figure 3. The graphs of coefficients $\alpha_0^{(0)} = (\alpha_{01}^{(0)} + \alpha_{02}^{(0)})^2$ and $\alpha_0^{(1)}$ in (27) and (28)