TEMPORAL AGGREGATION OF AN ECONOMETRIC EQUATION

ERIK RUIST

Stockholm School of Economics

Box 6501, S-11383 Stockholm, Sweden

Abstract: Structural breaks in an economy may make the time period available for estimation of an econometric equation exceedingly short. To use the existing information efficiently, it may be profitable to use high-frequency data, say monthly data, for estimation of a particular equation, even if the rest of the model is expressed in terms of data of lower frequency, say quarterly or half-yearly. In order to be included in the model, this equation has to be transformed to the same data frequency as the rest of the model. If the variables are of different types, or if some of the variables are lagged, exact transformations to equations that produce equivalent predicted values of the dependent variable are not possible. This note gives approximations and estimates for the various terms of the equations. Linear interpolation estimates as well as estimates that are optimal in a certain sense are given for the case of aggregation of monthly variables to semi-annual ones. It turns out that in the exchange rate equation in the KOSMOS model, the approximations do not increase the equation error substantially.

Keywords: Short time series, structural breaks, temporal aggregation.

Acknowledgement: The research reported in this paper was sponsored by the National Institute of Economic Research. The author wants to thank Alek Markowski and Lars-Erik Öller for fruitful suggestions and Jan Alsterlind for important help.

1. Introduction

In an econometric model, some equations may, for various reasons, be estimated using data with a frequency different from that of the main part of the data. These equations then have to be transformed in order to be consistent with the other equations.

This was the case in the financial part of the KOSMOS model for Sweden, built by the National Institute of Economic Research, see Markowski (1996). The real part of the model was estimated with semi-annual data. However, some financial variables, such as the rate of exchange, had been more or less constant or severely constrained for such a long time in the past that only a short period, say a couple of years, was available for estimation. In order to get more observations, monthly data were used to estimate the coefficients of, among others, the exchange rate equation. A similar situation arises in other cases with regime shifts, occurring e.g. when deregulating a market. An extreme example is the introduction of market economy in a formerly communist country.

To use the two groups of equations with different data frequencies in a common model, there is a choice between two strategies. Either the equations using aggregated (semi-annual) data have to be reestimated with disaggregated (monthly) data, which often have to be estimated or guessed, or the equations estimated on disaggregated data have to be transformed to an aggregated form. In the KOSMOS model, the second strategy was used. The present note intends to clarify the problems connected with the transformation of monthly equations into semi-annual ones that can be used in combination with the rest of the model. We will in several cases use the monthly exchange rate equation, estimated for KOSMOS, as an example. The calculations will be made for an aggregation of six terms. The generalization to other aggregates is usually self-evident. Some results for two terms will be mentioned in the Summary. The problem to be taken up in this paper is slightly different from what is ordinarily discussed in the literature on temporal aggregation. The situation most frequently treated is one with an equation or a model specified for fairly short time periods, say months, but estimated with quarterly or annual data. The problems arising in this situation are only partly the same as in the present case, but in the following we shall be able to take advantage of some previous results and experiences.

The starting point for our discussion of time aggregation is that we have specified and estimated an equation on monthly data. We now want to find an equivalent equation in terms of semi-annual data, since these are the only ones that are acceptable in combination with the rest of the model. This situation can be illustrated by a simple equation with only three explanatory variables:

$$y_t = a_0 + a_1 x_t + a_2 z_t + a_3 u_t + x_t \tag{1}$$

where t=1,2,...,N (months)

x_t is a flow variable, e.g. GDP

z_t is a state variable, measured as an average over the month, e.g. a price

u_t is a stock variable, indicating the state on the last day of the month, e.g. an inventory variable

x_t is a random variable, uncorrelated with the previous variables.

To begin with, the dependent variable y_t will be assumed to be of the same type as x_t. The corresponding variables in their semi-annual form are written in capitals

(Y,X,Z,U). The time index for these variables will be T=1,2,...,N/6, where T includes the six months t=(6T-5),(6T-4),...,6T. We define these variables as

$$Y_T = Sy_t \qquad X_T = Sx_t \qquad Z_T = \frac{1}{6}\sum z_t \qquad U_T = u_{6T}$$

(Here and in the following, the summation sign indicates a sum from (6T-5) to 6T, unless otherwise stated.)

The basic aim of the aggregation is to find an equation in Y_T, X_T, Z_T, and U_T that generates the same development of Y_T as would be obtained by summing the y_t derived from (1). A general strategy for finding this equation is to apply the same operator to the right hand side (RHS) of (1) as the one used to transform y_t into Y_T. This strategy has been used by several authors, in particular Zellner and Montmarquette (1971), Brewer (1973), and Weiss (1984). As will be evident in the following, the RHS of the aggregated equation will often contain functions of the original variables that could not be expressed in terms of aggregated variables, and thus have to be estimated. We shall later discuss whether in such cases a different strategy, allowing non-aggregated expressions also in the LHS, may be preferable.

In a static model with a flow variable as the dependent variable, the natural operator to apply is the simple summation, in our case over six terms. For comparison with later, more complicated cases, we may express this summation by a polynomial in B, the backshift operator. We shall here specify two backshift operators. B_t means a shift of one month, so that $B_t x_t = x_{t-1}$. We shall also use B_T, which shifts the variable one half year, so that $B_T X_T = X_{T-1}$. This means $B_T = B_t^6$. It should be noted that B_t causes a shift of only one month, even when operating on X_T. The summation can now be written as

$$X_T = A_1(B_t)x_t = \left(1 + B_t + B_t^2 + B_t^3 + B_t^4 + B_t^5\right) x_t = \frac{1 - B_t^6}{1 - B_t} x_t$$

For u_t , we have $U_T = A_2(B_t)u_t = u_t$. Thus, $A_2(B_t) = 1$.

Applying the summation operator A_1 to both sides of (1) yields

$$A_1(B_t)y_t = A_1(B_t)a_0 + a_1A_1(B_t)x_t + a_2A_1(B_t)z_t + a_3A_1(B_t)u_t + A_1(B_t)x_t \quad (2)$$

We will now investigate each term in (2).

1. For the dependent variable, $Y_T = A_1(B_t)y_t$, and the left hand side of the equation thus becomes Y_T.
2. The constant a_0 is replaced by $6a_0$.
3. For the flow variable x_t the expression $A_1(B_t)x_t$ corresponds exactly to X_T, so this term becomes a_1X_T.
4. Since for the state variable z_t, the corresponding semi-annual Z_T is an average of the monthly variables, this term in (2) becomes $6a_2Z_T$.
5. For u_t measured as the end-of-period values, the aggregation causes difficulties. The sum $A_1(B_t)\, u_t$ is different from $U_T=A_2(B_t)u_t$ and contains the values at the end of each of the six months. These values do not correspond to any observations on the semi-annual variable. We will meet several similar situations in the following. In the spirit of Wei (1978) we try to estimate the sum using available data. We shall discuss this estimation problem in Section 3. For the time being, we shall use a simple linear interpolation between U_T and U_{T-1} in the hope that the series u_t is smooth enough to allow for such a procedure. Since $U_T=u_{6T}$ and $U_{T-1}=u_{6T-6}$, we estimate

$$u_{6T-i} \approx \frac{1}{6}\left[(6-i)U_T + iU_{T-1}\right] = U_T - \frac{i}{6}\left(U_T - U_{T-1}\right)$$

and thus for the sum from (6T-5) to 6T, i.e. i=0,1,...,5:

$$\sum u_t \approx 6U_T - \frac{15}{6}\left(U_T - U_{T-1}\right) = \frac{7}{2}U_T + \frac{5}{2}U_{T-1}$$

Thus, the corresponding term in (2) will be specified by

$$a_3A_1(B_t)u_t \approx 6a_3\left(\frac{7}{12}U_T + \frac{5}{12}U_{T-1}\right)$$

6. The residual term in (2) is just the sum of six residuals in (1). If these are equally and independently distributed, the same is true of the residuals in (2), with a variance that is six times that in (1). If the original residuals are autocorrelated or heteroskedastic, the result is more complicated. The residuals are not further discussed in this paper.

Now, many equations are not as simple as (1). Various complications will be treated in the following. First, in Section 2, we discuss the case of transformed variables, notably logarithms. We continue in Section 3 with lagged exogenous variables. In Section 4, we make the natural extension to first differences. If the dependent variable is a first difference, we have to use a different transformation for converting the equation from monthly to semi-annual data. This situation is discussed in Section 5. We next discuss the general problem of lag specification in Section 6. A completely new situation arises when lagged endogenous variables are present in the RHS of the equation. This is taken up in Section 7. The question of which operator to use for the aggregation is discussed in Section 8. Finally, in Section 9, we use our results for converting the exchange rate equation into semi-annual form, and we make some reflections on the outcome.

2. Transformations of the Exogenous Variable

One very common complication when applying the rules listed above is that some variables are transformed. In econometric modelling it is common that the variable used is not z_t but log z_t. In this case, the summation of six terms yields $S\log z_t$, while the semi-annual counterpart is 6 log Z_T = 6 log 1/6 Sz_t .This is of course equivalent to using (the log of) the arithmetic mean instead of (the log of) the geometric mean. Since for variables with only positive values the former is always larger than the latter one, the equation is somewhat distorted. For variables with values in the range (1,2) - such as the Exchange rate - it is possible to get an idea of the difference between the two expressions by expanding them as Taylor series.

Let $v_t = z_t - 1$. Then

$$\log z_t = \log(1 + v_t) = v_t - \frac{1}{2} v_t^2 + \frac{1}{3} v_t^3 - \ldots$$

A summation over six terms gives

$$S_1 \equiv \Sigma \log z_t = \Sigma \log(1 + v_t) = 6\left(\frac{1}{6}\Sigma v_t - \frac{1}{12}\Sigma v_t^2 + \frac{1}{18}\Sigma v_t^3 - \ldots\right)$$

while

$$S_2 \equiv 6 \log Z_T = 6 \log\left(1 + \frac{1}{6}\Sigma v_t\right) = 6\left[\frac{1}{6}\Sigma v_t - \frac{1}{2}\left(\frac{1}{6}\Sigma v_t\right)^2 + \frac{1}{3}\left(\frac{1}{6}\Sigma v_t\right)^3 - \ldots\right]$$

This gives

$$\frac{1}{6}(S_2 - S_1) = \frac{1}{2}\left[\frac{1}{6}\Sigma v_t^2 - \frac{1}{36}(\Sigma v_t)^2 + \ldots\right] = \frac{1}{2} s_v^2 + \ldots = \frac{1}{2} s_z^2 + \ldots$$

Thus, the first term in the difference depends on the variance of z_t within each half year.

We have investigated two variables that enter the KOSMOS exchange rate equation in log form, and found that the difference between the two expressions is negligible in comparison with the overall variation of the variables. In fact, the variances of these variables *within each half year* are very small, see Table 1.Thus, no great harm is done by using 6 log Z_T instead of $\Sigma \log z_t$.

Table 1. The difference between log of the average $\left(\frac{1}{6}S_2\right)$ and average of the logs $\left(\frac{1}{6}S_1\right)$ for two variables in KOSMOS.

	Period (half -year)	$\frac{1}{6}S_1$	$\frac{1}{6}S_2$	$\frac{1}{6}(S_2 - S_1)$	$\frac{s^2}{2}$
Exchange	2/93	0.2480	0.2481	0.0001	0.00019
rate	1/94	0.2250	0.2251	0.0001	0.00006
	2/94	0.2294	0.2295	0.0001	0.00022
Price	2/93	0.06864	0.06865	0.00001	0.000008
relation	1/94	0.08234	0.08236	0.00002	0.000018
	2/94	0.10185	0.10186	0.00001	0.000008

Problems similar to that of logarithmic transformations arise when a variable in the model is defined as the ratio of two variables, since the mean of monthly ratios is not the same as the ratio of the means. In this case also, the error of approximation from using the latter expression instead of the former one is smaller, the smaller are the variances of the two variables. We shall meet an example of a ratio variable in Section 9.

3. Lagged Exogenous Variables

Having found that the logarithmic approximation does not seem to be very harmful, at least for variables of the type used in the KOSMOS exchange rate equation, we can move on to the next difficulty: lagged exogenous variables. Again, we formulate a model containing explanatory variables of the three kinds included in (1), but now lagged one period:

$$y_t = b_0+b_1x_{t-1}+b_2z_{t-1}+b_3u_{t-1}+x_t \qquad (3)$$

Summing six consecutive observations (applying the operator $A_1(B_t)$ to all terms**)** as before, we find that now the RHS variables form sums like Sx_t stretching from December to May or from June to November.Thus, the x_t sum is

$$S \equiv A_1(B_t)\,x_{t-1} = \frac{1-B_t^6}{1-B_t}\,B_t x_t \tag{4}$$

No corresponding semi-annual observations exist. As in the case of the end-of-period variable u_t under point 5 of Section 1, one solution is to estimate the lagged sums with the help of available data. In the terms used by Tiao and Wei (1976), we shall find the expected value of the sum, given the sums corresponding to calender half-years.

In order to make an efficient estimate, it is necessary to formulate and estimate a model for x_t. Compared to the situation covered in most of the literature on temporal aggregation, we are here at an advantage, since we have monthly observations of the x_t. As the reason for estimating the model equation on monthly data was that a structural break had made earlier observations irrelevant for the model, there is, however, a very short period available for identifying and estimating a model for the exogenous variables. It is obviously dangerous to rely on an assumption that the estimated model will describe the path of x_t also in the future. We shall therefore have to rely on less efficient but perhaps more robust methods.

We shall use two different approaches. The first is to approximate the series locally by fitting a polynomial to the semi-annual observations. In fact, due to the short series, we have deemed it appropriate to use only straight lines connecting two adjacent observations. This is the linear interpolation that we introduced already in point 5 of Section 1.

The second approach is to assume a simple ARMA model for x_t. We make explicit calculations only for AR(1) models with different autocorrelation coefficients γ. In several cases, it turns out that the result is rather robust for changing γ, and also rather similar to those obtained by linear interpolation.

When using interpolation, we make no assumption about the data-generating process for the variable we are treating. In the literature, the most frequently discussed situation is the following: Given a whole series of quarterly or semi-annual observations, estimate a corresponding monthly series which is as smooth as possible according to some criterion, and that adds up to the given quarterly/semi-annual values. For this approach, see e.g. Boot et al. (1967). For a recent survey of available methods, see Marcellino (1996).

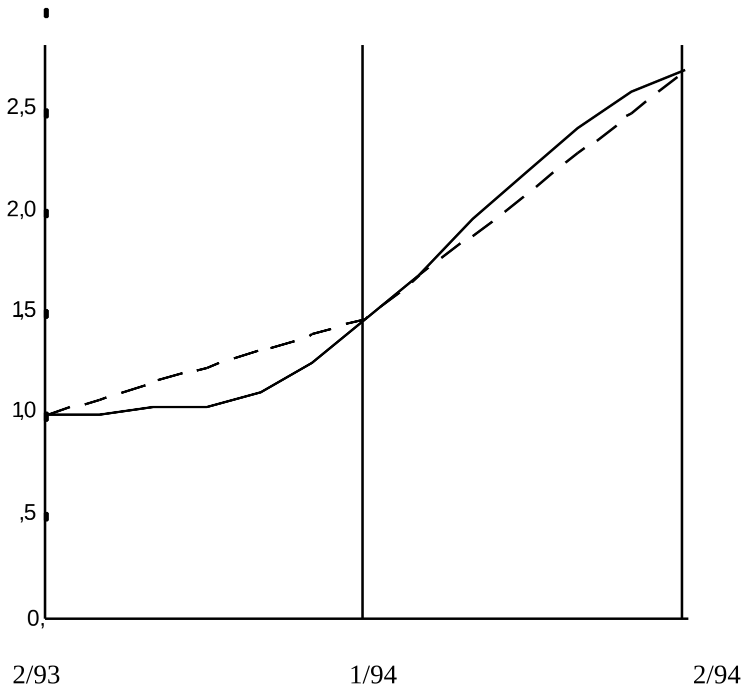
For our problem set-up, the estimation has to be made successively and not for all periods simultaneously. This means that future semi-annual observations could not be used for the estimation. It also seems complicated and probably not very informative to use observations from a too distant past. We shall therefore use only the last two or three semi-annual observations.

We will later compute error variances for the interpolation estimates, given certain assumptions about the data-generating process, but already here we want to point out the difference in information content of a semi-annual observation of an X or Z variable (according to the notation above), and a U variable. In the former case, the monthly value or values we are seeking to estimate are parts of a sum or average that we observe. In the end-of-period variable case, we want to find the values of observations *between* those in the semi-annual series. It is intuitively clear that the errors will usually be larger in this case.

Fig. 1. Interest rate difference between Sweden and Germany.

Moving 6 month average (curve) and linear interpolation (broken line).

. Per cent



The problem we encounter when the RHS of the equation contains lagged exogenous variables is to estimate the sum S in equation (4). Of the six terms in the sum, five are parts of the observable sum X_T, while one is part of X_{T-1}. Thus, a simple estimate of S is obtained by a linear interpolation between X_T and X_{T-1} :

$$\sum x_{t-1} \approx \frac{5}{6} X_T + \frac{1}{6} X_{T-1}$$

Since Z_T is a mean and not a sum, we get for z_{t-1}

$$\sum z_{t-1} \approx 6\left(\frac{5}{6} Z_T + \frac{1}{6} Z_{T-1}\right)$$

For u_t , the situation is somewhat different, since U_T and U_{T-1} are not sums, but single values. Thus, the sum Σu_{t-1} contains six terms, one of which is $U_{T-1}=u_{6T-6}$, while U_T does not enter. If nothing is known about the process that generates u_t, an estimate of the sum is like a shot in the air. However, if the curve is smooth,

Table 2. The accuracy of interpolation estimates of a lagged average for three variables in KOSMOS.

	Period (half-year)	Monthly average M_1	Estimate from interpolation M_2	M_1-M_2	s
Interest rate difference	1/94	1.207	1.353	- 0.146	0.762
	2/94	2.555	2.453	+0.102	
Log (relative prices)	1/94	0.07985	0.08005	- 0.00020	0.0143
	2/94	0.09910	0.09858	+0.00052	
Wealth (M3)	1/94	743 700	744 400	-700	14 600
	2/94	724 200	741 700	-17 500	

we may try a linear interpolation here, too. It turns out to be similar to the one we used in (2):

$$\sum u_{t-1} \approx 6\left(\frac{5}{12}U_T + \frac{7}{12}U_{T-1}\right)$$

The nature of the *x* and *z* approximations can be illustrated by moving sums or moving averages, see Fig.1. We use a variable that enters the exchange rate equation (see eq.(19) in Section 9) with a lag of one period, i.e. one month: The interest rate difference between Sweden and Germany.

The curve shows six months moving averages, and the vertical lines indicate values that correspond to calender half years. These are the only values that could be used in the semi-annual model. The straight lines between these points indicate the values obtained by a linear interpolation. The values corresponding to averages lagged one period are marked by dots, so the differences between these values and the actual averages (on the curves) are easily noted. Table 2 gives a more accurate account for the interest rate difference, for the price relation between Sweden and Europe, and also for the wealth variable, which is an end-of-period variable. To facilitate an evaluation of the approximation error, the standard deviation of each variable over the period July 1993 - Dec. 1994 is also given.

It seems that for the first variable, the approximation error is not negligible, amounting to about one sixth of the standard deviation, while for the log of the relative prices, the error is of little consequence. However, for wealth, the end-of-period variable, the error is of the same order as the standard deviation. It seems important to investigate whether an alternative approach could yield a better estimate, and, in any case, to determine the accuracy of the approximations more carefully than by examples from two periods. This will be done in the remainder of this Section.

In order to formulate more accurate estimation methods than linear interpolation, we have to exploit all information that may be available regarding the generation of the exogenous variables. If it is known that a variable is generated by an ARMA process, this information can be used when calculating the expected values of the desired functions. We shall investigate this situation, starting with the case where there is only one exogenous variable x_t, which appears in the model with a lag:

$$y_t = a_1 x_{t-1} + x_t \qquad (5)$$

and x_t is known to be generated by an ARMA process. This case has been extensively treated by Brewer (1973), Tiao and Wei (1976), Weiss (1984), and others.

As noted above, we have to find the expected value of the sum S in (4), given the observations of X_{T-i} for various values of i, possibly including negative as well as positive lags. There is, however, some disagreement among previous authors as to which values of X_T should be considered as given. Brewer (1973) uses only X_{T-i} (i=1,2,...) and not X_T in order not to introduce simultaneity which could complicate the estimation of the aggregated equation. Tiao and Wei (1976) note that X_T is usually the term which has the largest correlation with Y_T, and not taking it into account means a reduced efficiency in estimating the micro relation. They find that neglecting forward terms (i=-1,-2,...) also leads to difficulties in estimation by making the residuals correlated with the X's. Thus Tiao and Wei use in principle all available X's, but in practice the terms with long leads or lags get very small coefficients and can be disregarded.

Tiao and Wei present an example, using the basic model (5) with x_t generated by an AR(1) process with autoregressive coefficient γ. The aggregated equation becomes

$$Y_T = a_1 L(B_T)X_T + Q_T$$

where $L(B_T)$ is a polynomial in B_T, used for estimating S, and Q_T is the sum of the residuals x_t. Applied to a summation of six terms and for γ=0.9, their estimate can be calculated to be:

$$L(B_T)=-0.005B_T^{-3}+0.023B_T^{-2}-0.093B_T^{-1}+0.940+0.167B_T-0.042B_T^{2}+.010B_T^{3}$$

Thus, the coefficients of the negative lags are not negligible, although the contemporaneous term and the first positive lag are most important.

Weiss (1984) is disturbed by the fact that the procedure used by Tiao and Wei destroys the one-way causality that exists in the basic model, and notes that the estimation problem encountered when neglecting the negative lags (correlation between residuals and explanatory variables) can be overcome by the use of instrumental variables. Thus, he takes X_{T-i} (i=0,1,2,...) as the base for computing the conditional expected value of the "skew" sums.

In our problem set-up, we need not take into account the difficulties in the estimation of the aggregated relation, since we have already estimated the monthly relation. Furthermore, a relation containing leads in one or more explanatory variables would be disturbing in the context of the overall model. Thus, we stick to the approach favoured by Weiss, and want to find a function $L(B_T)X_T = k_0X_T + k_1X_{T-1} + k_2X_{T-2} + \ldots$ that is a good estimator of S in (4). The coefficients k_i depend on the form of the ARMA model for x_t. As has already been pointed out, the observation period for the KOSMOS financial variables is very short, so it is not possible to have much confidence in the models that are identified by the use of these obervations. In most cases, AR(1) models give a reasonably good fit with autocorrelation coefficients in the range (0.60, 0.95). The coefficients for AR2 or MA1 generally are not significantly different from 0. Thus, we have deemed it sufficient to make calculations for the AR(1) model. In this case

$$(1-\phi B_t)(x_t-\mu) = \varepsilon_t \tag{6}$$

where μ is the arithmetic mean of x. The interpolation estimate (p. 4) used only X_T and X_{T-1}. It seems now appropriate to include also X_{T-2} in the specification of $L(B_T)$ in order to find out if its coefficient is so small that it can be disregarded. Tiao's and Wei's results indicate that it is not necessary to include any further terms.

To determine $L(B_T)X_T$, we express it in terms of x_t, thus:

$$L(B_T)X_T = k_0 X_T + k_1 X_{T-1} + k_2 X_{T-2} = \left(k_0 + k_1 B_t^6 + k_2 B_t^{12}\right)\frac{1-B_t^6}{1-B_t}x_t$$

This can now be compared with $S(B_t)x_t$, the function of the monthly observations that we want to estimate. In the present problem, this is

$$S(B_t)x_t = \frac{1-B_t^6}{1-B_t}B_t x_t$$

and thus the error of estimation, to be denoted $F(B_t)x_t$, is

$$F(B_t)x_t = S(B_t)x_t - L(B_T)X_T = \left(B_t - k_0 - k_1 B_t^6 - k_2 B_t^{12}\right)\frac{1-B_t^6}{1-B_t}x_t$$

Given (6), we can now determine the k_i so as to minimize the variance of $F(B_t)x_t$. In order to facilitate the comparison with the interpolation estimates, we impose the restriction that $\Sigma k_i = 1$.

Brewer and Weiss use a slightly different procedure, as they want to express the estimator as an ARMA process which is as simple as possible. They apply an operator $T(B_t)$ to both sides of a slightly rearranged equation (6):

$$T(B_t)(1-\gamma B_t)x_t = T(1)(1-\gamma)\mu + T(B_t)\varepsilon_t \qquad (7)$$

Now, if $T(B_t)$ is chosen in such a way that

$$T(B_t)(1-\gamma B_t)x_t = S(B_t)x_t - L(B_T)X_T \equiv F(B_t)x_t \qquad (8)$$

we get $$F(B_t)x_t = T(1)(1-\gamma)\mu + T(B_t)\varepsilon_t$$

If $T(1)$ is made=0, then $T(B_t)\varepsilon_t$ is the error of estimation. Since ε_t is white noise, it is easy to minimize the error variance.

Applying this method to our present problem, equation (8) becomes

$$T(B_t)(1-\gamma B_t)x_t = \frac{1-B_t^6}{1-B_t}\left(B_t - k_0 - k_1 B_t^6 - k_2 B_t^{12}\right) x_t \qquad (9)$$

Let $$T(B_t) = q_0 + q_1 B_t + q_2 B_t^2 + q_3 B_t^3 + \ldots + q_{16} B_t^{16}$$

By equating the coefficients of B_t^i for every i in (9), it is possible to determine the coefficients of $T(B_t)$ as functions of k_0, k_1, and k_2 (and of γ). Since the highest power of B_t in the RHS of (9) is 17 - when the ratio is expressed as a sum - there are 18 such equations. They determine the q_i (i=0,1,...,16). The 18th equation can be used to get a restriction that k_0, k_1, and k_2 have to satisfy:

$$\gamma^{11} - k_0\gamma^{12} - k_1\gamma^6 - k_2 = 0 \qquad (10)$$

Neither Brewer nor Weiss determines the size of the coefficients, but Weiss indicates that, in addition to restriction (10), he would set k_0=1. If instead the same restriction as above, i.e. $\Sigma k_i=1$, is used, the estimate of the "skew sum" could be seen as a weighted average of X_T, X_{T-1}, and X_{T-2}. It also makes the first term of the RHS of (7) equal to 0, since in that case

$$T(1)(1-\gamma) = 6(1-k_0-k_1-k_2) = 0 \qquad (11)$$

according to (9). By (7), (8), and (11)

$$S(B_t)x_t = L(B_T)X_T + T(B_t)e_t \tag{12}$$

Since ε is assumed not to be autocorrelated, the variance of the moving average $T(B_t)e_t$ is minimized if the sum Sq_i^2 is as small as possible, subject to (10).

The difference between the Brewer/Weiss procedure and the one described on pp. 14 -15 is the imposition of restriction (10). It turns out that this restriction in certain cases affects the coefficient estimates in such a way that the error variance is considerably increased. Since for us it is not necessary to express the estimates in terms of ARMA processes, we stick to the estimates derived without the restriction (10). The results for various situations are given in Tables 3-8 in the following. In each table, the coefficients optimal for various values of γ are given, together with the error variance, assuming that the variance of e_t is 1. For comparison, the simpler interpolation estimates are also shown in the tables together with their error variances, given that the x_t follows an AR(1) process. In order to get a dimension-less measure of the goodness of various estimators, we have also calculated an equivalent of a determination coefficient R^2, to be denoted R_*^2:

$$R_*^2 = 1 - \frac{\operatorname{var} F(B_t)}{\operatorname{var} S(B_t)}$$

Table 3. Optimal coefficients in the expression $L(B_T)X_T=(k_0X_T + k_1X_{T\text{-}1} + k_2X_{T\text{-}2})$ for estimation of $S(B_t)=A_1(B_t)x_{t\text{-}1}$ when x_t is AR(1)

γ	k_0	k_1	k_2	Error variance	R_*^2	Interpolation estimate Error variance	Interpolation estimate R_*^2
0.01	0.835	0.165	-0.001	1.668	0.73	1.668	0.73
0.1	0.852	0.155	-0.007	1.696	0.76	1.700	0.76
0.2	0.867	0.147	-0.014	1.754	0.80	1.769	0.80
0.3	0.879	0.141	-0.020	1.842	0.83	1.876	0.83
0.4	0.881	1.139	-0.027	1.966	0.86	2.027	0.86
0.5	0.894	0.140	-0.034	2.125	0.89	2.224	0.88
0.6	0.896	0.145	-0.041	2.314	0.91	2.467	0.91
0.7	0.892	0.155	-0.047	2.516	0.94	2.736	0.93
0.8	0.883	0.167	-0.050	2.696	0.96	2.994	0.96
0.9	0.867	0.181	-0.048	2.809	0.98	3.153	0.98
0.99	0.849	0.189	-0.038	2.805	1.00	3.084	1.00
inter-polation	0.833	0.167	-				

This is given in every table for the optimal as well as for the interpolation estimates. The tables are commented upon as the various cases are presented in the following.

For the present problem with a lagged exogenous variable of the flow variable type, generated by an AR(1) process, Table 3 shows that whatever the value of γ, the coefficient for $X_{T\text{-}2}$ is rather small and can in most cases be ignored. The coefficients approach the interpolation values when γ tends to 0, i.e. when x_t becomes white noise. Even for other γ values, the error variance of the interpolation estimate is at most 12 per cent higher than the optimal one. In view of the fact that the data generating process could be assessed only with great uncertainty, it seems legitimate to use interpolation for the estimate in this case. The R_*^2 measure is reasonably high for low γ values, and tends for both estimates

Table 4. Optimal coefficients in the expression $L(B_T)U_T=(k_0U_T + k_1U_{T-1} + k_2U_{T-2})$ for estimation of $S(B_t)u_t=A_1(B_t)u_{t-1}$ when u_t is AR(1)

γ	k_0	k_1	k_2	Error variance	R_*^2	Interpolation estimate	
						Error variance	R_*^2
0.01	1.670	2.670	1.660	13.348	-	17.481	-
0.1	1.704	2.704	1.593	13.584	-	17.427	-
0.2	1.750	2.750	1.500	14.064	-	17.580	-
0.3	1.809	2.808	1.382	14.762	-	17.910	-
0.4	1.887	2.882	1.230	15.658	-	18.357	-
0.5	1.990	2.974	1.036	16.708	0.11	18.844	-
0.6	2.120	3.082	0.798	17.804	0.32	19.270	0.27
0.7	2.266	3.205	0.530	18.731	0.53	19.512	0.51
0.8	2.400	3.330	0.270	19.162	0.72	19.420	0.71
0.9	2.482	3.442	0.076	18.791	0.88	18.817	0.88
0.99	2.500	3.499	0.001	17.670	0.99	17.670	0.99
inter-polation	2.5	3.5	-				

towards 1 as γ increases. Thus, for this case, estimating the $S(B_t)x_t$ expression by the function $L(B_T)X_T$ increases the error by only a small amount.

These results are applicable to exogenous variables that are either flow variables or state variables, measured as period averages. In the latter case all coefficients have to be multiplied by 6. For end-of-period variables, the calculations are slightly different, due to the fact that even the semi-annual variables are measured at the end of the period. Since we are estimating a sum from individual values, the coefficients have to sum to 6. Using this condition, and minimizing the error variance, we get the coefficients shown in Table 4.

In this case, the optimal coefficients get closer to the interpolation ones when γ increases towards 1. For lower values, the k_2 coefficient is not negligible. For very small γ, the error variance is 30 per cent higher for the interpolation estimate than for the optimal one, but both are high compared to the variance of the expression to be estimated. For γ values lower than about 0.5, the error variance is even larger than the variance of $S(B_t)u_t$. Calculating R_*^2 as above would result in negative values. This shows that R_*^2 is not directly comparable with a coefficient of determination in an ordinary regression, the reason being that the error is not uncorrelated with the estimate. The result shows, however, that the information contained in the observable U_{T-i} is insufficient to estimate the function $A_1(B_t)u_t$ when the autocorrelation of u_t is low. We shall reach similar results in the following Sections, when other $S(B_t)u_t$ are analyzed. In the concluding Section, we shall return to the problems encountered when using end-of-period variables.

It is interesting to note, that in the case of an unlagged u_t exogenous variable that we met in point 5 of Section 1, but left for later consideration, the results are the same as in Table 4, except that the k_0 coefficient is for all γ decreased by 1 unit, which is instead added to k_1. The error variances and the conclusions are the same.

4. First Differences of Exogenous Variables

A case which is closely connected with that of lagged variables is when first differences of exogenous variables are entered into the equation. This case calls for some additional comments. We shall discuss two different approaches, and then take up the more general question of the lag length in the model. To begin with, we shall consider the case of a period-average variable z_t.

Our first approach is to let $A_1(B_t)$ operate on each of the terms in $(z_t\text{-}z_{t-1})$ separately, yielding

$$A_1(B_t)z_t - A_1(B_t)z_{t-1} = 6Z_T - S \tag{13}$$

where S is the "skew sum" used in the previous Section (formula (4)).

Thus, in (13), the first term is a directly observable magnitude, and the second one could be treated in the same way as when it appeared alone as a lagged variable (Section 3). Since it turned out that, for z_t following an AR(1) process, the results obtained from a linear interpolation were close to the optimal estimate, we use this simpler form. Thus,

$$\sum \left(z_t - z_{t-1} \right) \approx 6Z_T - 6\left(\frac{5}{6} Z_T + \frac{1}{6} Z_{T-1} \right) = Z_T - Z_{T-1}$$

For the flow variable x_t, the equivalent expression becomes $\frac{1}{6}\left(X_T - X_{T-1} \right)$, again an approximation of the same kind as that for z_t.

The end-of-period variable u_t turns out to be the simplest one in the case of differences, as

$$S(u_t\text{-}u_{t-1}) = u_{6T} - u_{6T-6} \ \text{º} \ U_T - U_{T-1}$$

which is an exact expression, not an approximation. This explains the relation between the coefficients in the estimates of Su_t, and Su_{t-1}, noted above (p.19). Taking the difference between these estimates, the result is $(U_T\text{-}U_{T-1})$, as it should.

For the x_t and z_t cases, the errors of approximation are the same as for the lagged variables (although with changed sign), since the first term of the difference is reproduced by the aggregated variable without error. The relative errors, of course, are different from those of the lagged variable, and may sometimes be felt more disturbing.

It is interesting to note that a different approach to the estimation of differenced variables leads to identical expressions, but with a different lag. Taking again the case of z_t, the six month sum includes five terms that are eliminated by five corresponding terms in the sum of z_{t-1}, so that

$$S(z_t\text{-}z_{t-1}) = z_{6T} - z_{6T-6}$$

Neither of the terms of the RHS of this expression can be observed if we have only semi-annual data, but we may estimate both in a way similar to what was done above. However, the estimation is probably made more accurate if we lag the monthly variables by 2½ months so that both terms refer to a period in the middle of a half-year. Then a natural approximation for each of them is the monthly average of the semi-annual variable, thus:

$$B_t^{5/2}(S(z_t - z_{t-1})) \approx (Z_T - Z_{T-1}) \quad \text{or} \quad S(z_t - z_{t-1}) \approx B_t^{-5/2}(Z_T - Z_{T-1})$$

The lag in the expression may seem disturbing, but at the end of the calculations all terms of the equation could be shifted backwards or forwards so that a simple expression is obtained for most of the terms[1].

It may, however, be useful to give some thoughts to the fact that we obtained the same approximation for the sum of first differences, whether lagged by 2½ months or not lagged at all. In fact, it can be shown that the same result is obtained by the linear approximation for all lags between 0 and 5 months. This can be understood by looking at Figure 1. All one-month differences between two consecutive sums (or, as in the figure, averages) are naturally estimated by a

portion of the straight line connecting Z_T and Z_{T-1}, and the one-month differences along this line are of course the same everywhere.

This equality suggests that the problem may be turned around and formulated as a question of which sum of one-month differences is best approximated by the semi-annual difference?

The answer to this question depends upon how the z_t series is generated. If it is a pure random (white noise) series or a random walk, the standard deviation of the

1 This approach was suggested by Alek Markowski.

estimate is the same for all lags between 0 and 5. If, however, z_t is a series with positive autocorrelation, it seems that the standard deviation of the error is smaller for central differences than for the extreme ones.

This suggests that it could be profitable to lag all terms of the monthly equation by a suitable amount before summing its terms to semi-annual values. The lag should be chosen so as to make the necessary approximations as good as possible. For first differences, the optimal lag may be 2½ months, but that destroys the exact equivalence of the unlagged terms, so the choice is not obvious.

We shall return to this question in Section 6, together with a general discussion of the specification of lag length in the model. First it is, however, useful to discuss what happens when the dependent variable is itself a difference.

5. The Dependent Variable as a First Difference

So far, we have discussed what happens to the RHS of the equation, if we apply the operator $A_1(B_t)$ to all terms in order to transform the monthly dependent variable y_t to the semi-annual variable Y_T. If, as in the exchange rate equation, the dependent variable is a monthly difference, we have to find another operator that can transform (y_t-y_{t-1}) into (Y_T-Y_{T-1}).

It was shown by Zellner and Montmarquette (1971) that (Y_T-Y_{T-1}) is in fact a weighted sum of several one-month differences. We can find the operator in the following way. The LHS of the monthly equation is $(1-B_t)y_t$, and we want an operator C to convert this to $(1-B_T)Y_T$. Since $B_T = B_t^6$ and $Y_T = \frac{1-B_t^6}{1-B_t} y_t$ we want

$$C(B_t)(1-B_t)y_t = (1-B_T)Y_T = \left(1-B_t^6\right)\frac{1-B_t^6}{1-B_t}y_t$$

Thus, the necessary operator is

$$C(B_t) = \left(\frac{1-B_t^6}{1-B_t}\right)^2 = (A_1(B_t))^2 = \left(1+B_t+B_t^2+B_t^3+B_t^4+B_t^5\right)^2$$

Developing this, we get

$$C(B_t)=1+2B_t+3B_t^2+4B_t^3+5B_t^4+6B_t^5+5B_t^6+4B_t^7+3B_t^8+2B_t^9+B_t^{10}$$

It is clear that if we apply the operator $C(B_t)$ to the whole equation, first difference terms of x_t and z_t in the RHS will cause no problems, since they are converted to observable semi-annual differences, while lagged or unlagged x_t, z_t, or u_t variables as well as first differences of u_t become more complicated.

We have for the flow variable x_t:

$$C(B_t)x_t = A_1(B_t)X_T$$

This is a moving six month sum that corresponds to a semi-annual value for only one out of six terms. As before, we may estimate the intervening values. Linear interpolation gives the estimate of $C(B_t)x_t$ as

Table 5. Optimal coefficients in the expression $L(B_T)X_T= (k_0X_T + k_1X_{T-1} + k_2X_{T-2})$ for estimation of $S(B_t)x_t= C(B_t)x_t$ when x_t is AR(1)

γ	k_0	k_1	k_2	Error variance	R_*^2	Interpolation estimate	
						Error variance	R_*^2
0.01	3.503	2.503	-0.006	35.480	0.76	35.480	0.76
0.1	3.527	2.528	-0.056	40.297	0.77	40.330	0.77
0.2	3.554	2.559	-0.113	46.830	0.79	46.997	0.79
0.3	3.581	2.593	-0.174	54.801	0.81	55.288	0.81
0.4	3.607	2.632	-0.239	64.322	0.83	65.487	0.83
0.5	3.630	2.678	-0.308	75.260	0.86	77.776	0.85
0.6	3.642	2.732	-0.374	87.014	0.88	92.049	0.88
0.7	3.634	2.796	-0.429	98.285	0.92	107.560	0.91
0.8	3.594	2.865	-0.459	107.106	0.95	122.337	0.94
0.9	3.514	2.933	-0.448	111.369	0.98	132.238	0.97
0.99	3.419	2.968	-0.388	109.980	1.00	130.671	1.00

inter-polation	3.5	2.5

$$6\left(\frac{7}{12}X_T + \frac{5}{12}X_{T-1}\right)$$

An optimal estimate for the case when x_t is AR(1) could be made along the same lines as in Section 3, and results in the coefficients shown in Table 5. They are equal to the interpolated estimates for low values of γ, but deviate from them more than in earlier cases, when γ is growing. The variance of the interpolation estimates are less than 20 per cent higher than the optimal ones, but both are considerably higher than in the case when the dependent variable is y_t .The transformation of period-average variables is of course similar to that of the flow variables. The only difference is that the factor 6 is replaced by 36. End-of-period stock variables need, however, be treated separately. Since $C(B_t)$ includes terms up to B_t^{10}, it is necessary in the linear interpolation to bring in a term with

Table 6. Optimal coefficients in the expression $L(B_T)U_T = (k_0U_T + k_1U_{T-1} + k_2U_{T-2})$ for estimation of $S(B_t)u_t = C(B_t)u_t$ when u_t is AR(1)

γ	k_0	k_1	k_2	Error variance	R_*^2	Interpolation estimate	
						Error variance	R_*^2
0.01	10.980	15.060	9.960	419.39	-	534.46	-
0.1	10.790	15.642	9.568	416.24	-	517.40	-
0.2	10.562	15.375	9.063	416.84	-	503.03	-
0.3	10.324	17.205	8.470	420.28	-	491.55	-
0.4	10.092	18.130	7.778	424.85	-	481.03	-
0.5	9.890	19.129	6.982	428.73	0.17	469.73	0.10
0.6	9.737	20.169	6.094	429.99	0.43	456.34	0.40
0.7	9.636	21.202	5.162	426.65	0.64	440.22	0.63
0.8	9.549	22.164	4.287	416.67	0.81	421.21	0.80
0.9	9.431	22.952	3.618	398.16	0.92	398.66	0.92
0.99	9.335	23.320	3.337	372.26	0.99	372.26	0.99
inter-polation	9.333	23.333	3.333				

$U_{T-2} = u_{t-12}$, giving

$$C(B_t)\,u_t \approx \frac{28}{3}U_T + \frac{70}{3}U_{T-1} + \frac{10}{3}U_{T-2}$$

The coefficients sum to 36. Efficient estimation for u_t = AR(1) gives similar coefficients for high values of γ, see Table 6. For lower autocorrelations, the weights move toward U_T and U_{T-2} and the differences from the interpolated estimates become rather large.

Since first differences of x_t and z_t are exactly changed to (X_T-X_{T-1}) and $6(Z_T-Z_{T-1})$, we can find the optimal coefficients for the estimation of $C(B_t)x_{t-1}$ and $C(B_t)z_{t-1}$ by using the identity

$$x_{t-1} = x_t - (x_t - x_{t-1})$$

which of course gives

$$C(B_t)x_{t-1} = C(B_t)x_t - C(B_t)(x_t - x_{t-1}) = C_t(B_t)x_t - (X_T - X_{T-1})$$

Table 7. Optimal coefficients in the expression $L(B_T)U_T = (k_0U_T + k_1U_{T-1} + k_2U_{T-2})$ for estimation of $S(B_t) = C(B_t)u_{t-1}$ when u_t is AR(1)

γ	k_0	k_1	k_2	Error variance	R_*^2	Interpolation estimate	
						Error variance	R_*^2
0.01	9.970	16.060	9.970	409.23	-	511.90	-
0.1	9.679	16.642	9.679	404.45	-	494.08	-
0.2	9.312	17.375	9.312	402.78	-	478.42	-
0.3	8.897	18.205	8.897	403.35	-	465.19	-
0.4	8.435	19.130	8.435	404.38	-	452.47	-
0.5	7.936	20.129	7.936	403.99	0.22	438.61	0.16
0.6	7.415	21.169	7.415	400.53	0.47	422.56	0.44
0.7	6.899	22.202	6.899	392.73	0.67	404.05	0.66
0.8	6.418	23.164	6.418	379.74	0.82	383.58	0.82
0.9	6.024	23.952	6.024	360.80	0.93	361.24	0.93
0.99	5.836	24.328	5.836	336.92	0.99	336.92	0.99
inter-polation	5.833	24.333	5.833				

Thus, from Table 5, we get for γ=0.7:

$$\text{est } C(B_t)x_{t-1} = 3.634\, X_T + 2.796\, X_{T-1} - 0.429\, X_{T-2} - X_T + X_{T-1} =$$

$$= 2.634\, X_T + 3.796\, X_{T-1} - 0.429\, X_{T-.2}$$

As in most cases, end-of-period variables are more complicated. Table 7 gives the coefficients for estimation of $C(B_t)u_{t-1}$. Like those for $C(B_t)u_t$ they depend rather heavily on the value of γ. For all γ, we have $k_0=k_2$, and the coefficients tend to those of the interpolation estimate when $\gamma\to 1$. Again, the error variances are high, and for γ<0.5, the calculated R_*^2 is negative.

For end-of-period variables u_t , the estimate of $C(B_t)(u_t-u_{t-1})$ that is obtained by the difference between those shown in Tables 6 and 7 for $C(B_t)u_t$ and $C(B_t)u_{t-1}$ is restricted to $k_0(U_T - U_{T-1}) + k_1(U_{T-1} - U_{T-2})$ but has instead no restriction on the sum (k_0+k_1). Table 8 shows the coefficients arrived at when allowing also a

Table 8. Optimal coefficients in the expression $L(B_T)U_T = k_0(U_T - U_{T-1}) + k_1(U_{T-1} - U_{T-2}) + k_2(U_{T-2} - U_{T-3})$ for estimation of $S(B_t)u_t = C(B_t)(u_t - u_{t-1})$ when u_t is AR(1)

γ	k_0	k_1	k_2	Error variance	R_*^2	Interpolation estimate	
						Error variance	R_*^2
0.01	2.504	2.002	1.494	15.122	-	22.562	-
0.1	2.544	2.022	1.433	16.407	-	23.326	-
0.2	2.600	2.050	1.350	18.282	-	24.611	-
0.3	2.675	2.085	1.244	20.699	-	26.363	-
0.4	2.764	2.130	1.106	23.717	0.07	28.562	-
0.5	2.884	2.186	0.930	27.308	0.16	31.112	0.04
0.6	3.034	2.253	0.712	31.230	0.26	33.787	0.20
0.7	3.204	2.330	0.467	34.874	0.37	36.171	0.35
0.8	3.362	2.408	0.231	37.250	0.51	37.633	0.50
0.9	3.467	2.472	0.060	37.388	0.64	37.420	0.64
0.99	3.500	2.500	0.001	35.340	0.75	35.340	0.75
inter-polation	3.5	2.5	-				

term $k_2(U_{T-2} - U_{T-3})$ but imposing the restriction $Sk_i = 6$, which is the sum of the interpolation coefficients. Again, the coefficients tend to those of interpolation when $\gamma \rightarrow 1$, but both estimates give rather poor results, even for high autocorrelations.

Since the first difference of an end-of-period variable is in fact a flow variable, it would seem that the results for this case should be the same as those for $C(B_t)x_t$, shown in Table 5. That this is not the case depends upon the fact that in Table 8, we assumed u_t to be an AR(1) process, while Table 5 uses the assumption that $x_t=u_t-u_{t-1}$ is AR(1). Thus, the tables are not directly comparable. An AR(1) process with positive autocorrelation for u_t corresponds to a negative autocorrelation for x_t, and x_t is not AR(1). However, the large differences between the results show that the choice of assumption regarding the data generating process is not without consequences. On the other hand, the interpolation coefficients are the same, regardless of which approach is chosen.

6. The Specification of Lags

It has been shown in Sections 4 and 5 that lagged variables and first differences cause trouble when we want to aggregate an equation to observations over periods that are longer than that in which the equation was specified. It is, however, interesting to note, that the aggregated expression (X_T-X_{T-1}) which we have used as an estimate of $S(x_t-x_{t-1})$, possibly with some lag, has an exact equivalent in the disaggregated world, not only as a rather complicated sum $C(B_t)(x_t-x_{t-1})$, but also as $\Sigma(x_t-x_{t-6})$. This means that we would not have run into any trouble in aggregation, if we had specified a lag in the monthly equation of six months instead of one.

This observation naturally leads to the question: Which is the correct lag in the specification of the equation? The appearance of lags in an econometric equation may stem from various sources. We will discuss two of them here: an error correction mechanism and the estimate of a net flow as the first difference of a stock variable. Both are relevant for the exchange rate equation.

As for the error correction mechanism, it may in a simple form look like this:

$$(y_t-y_{t-1}) = b\,t(x_t-x_{t-1}) + a(bx_{t-1}-y_{t-1})+e_t \qquad (14)$$

The interpretation is that the expected change in y depends upon the change in (the equilibrium state) x and the deviation from (equilibrium) bx during the previous period.

The meaning of "the previous period" is rather vague, and economic theory seldom gives any indication of the appropriate length of the period. So, in practice it is almost always set equal to the length of the observation period. In this case, the equation was specified with monthly data, so the lag was set to one month. It was pointed out above, that from a data point of view a more practical lag would be six months. In the present case, this seems to be a very long time for an exchange rate to react to an out-of-balance value, but the reasonableness of such a lag should be discussed when the model is specified.

The other source of lagged values or differences was the estimation of a net flow from the difference of a stock variable. In the exchange rate equation, this is in fact the source of all difference terms. If only the difference of the stock variable enters the equation, no problem arises in the aggregation, since the difference of an end-of-period stock variable has its direct counterpart in the semi-annual data (see p.20). However, in this case, the whole equation determining the stock variable was differenced in order to be substituted into the exchange rate equation. Thus, the one-period differences of several monthly-average variables also enter the equation. The consequences of changing these to six-month differences have to be considered.

However, this doubt about the appropriateness of setting the one period lag equal to one month naturally leads to a suggestion that the equation should be specified in terms of the semi-annual variables, but taking into consideration the necessity of estimation on monthly data. To evaluate this suggestion, we have to think in more or less the same terms as in Section 1, only changing aggregation into disaggregation.

As before, X_T and Z_T terms cause no trouble. It is the lagged effects and the end-of-period variables that are more problematic cases. If a variable is thought to exert its influence on the dependent variable over a period of time or only after some delay, it is common practice in econometric modelling to specify a distributed lag. With only semi-annual observations, it is probably seldom necessary to go beyond the previous half-year. But how do we express $k_0X_T+k_1X_{T-1}$ in terms of monthly variables?

If the delay is considered to be fairly short, only a month or two, most of the effect comes within the half year, and it may not be necessary to include X_{T-1} at all. For estimation on monthly data, one or two lags may be included, and the sum of their coefficients can be allotted to X_T. With longer lags, it may be necessary to use the type of estimation we presented in Sections 3 and 4. For end-of-period variables, similar considerations could be made. They occur often as U_{T-1}, i.e. beginning-of-period values, to which agents are supposed to react. How long a reaction lag is realistic? Is it necessary to interpolate between U_T and U_{T-1} to get a precise timing?

These questions and suggestions are admittedly somewhat provocative, and can perhaps be used as excuses for rough and easy modeling. We believe, however, that it is appropriate to have the semi-annual as well as the monthly perspective in mind when creating a model that should be used in the way we are discussing in this paper. In the following we will, however, return to the situation where we have specified and estimated a monthly model and want to transform it to semi-annual form.

7. Lagged Endogenous Variables

The previous Sections dealt with problems encountered when the RHS of the equation contains lagged exogenous variables. Lagged endogenous variables have, however, to be treated in a different way. In a monthly equation, the appearance of y_{t-1} as an explanatory variable does not create too much of a problem for estimation. But when aggregating to semi-annual data, the observation with a lag of one month is in fact included in the value of the dependent variable and thus appears on both sides of the equation.

Several authors have dealt with this problem in the following form. Given an ARMA model for y_t in monthly data, what would an equivalent process in terms of aggregated data be like? By equivalent process we here mean one that generates the same development of the dependent variable as the monthly model, when the resulting monthly data are summed to half-years. Some of the authors have also treated the same problem with one or more exogenous variables present (an ARMAX process). It turns out that the AR part of the model is not influenced by the presence of an exogenous variable, while the MA part may be considerably more complicated.

It was noted already by Amemiya and Wu (1972) that the *m*-month aggregate equivalent to a monthly process

$$A(B_t)y_t = F(B_t)e_t \quad \text{is} \quad A^*(B_T)Y_T = F^*(B_T)e_T^*$$

where e_t and e_T^* are white noise variables. $A^*(B_T)$ is of the same order as $A(B_t)$ and has roots that are the m^{th} power of the roots of $A(B_t)$. In the case of an AR(1) process, this results in a particularly simple transformation. Thus, if

$$y_t = j\, y_{t-1} + e_t \quad \text{then} \quad Y_T = j^{\,m} Y_{T-1} + F^*(B_T)e_T^*$$

Thus, in contrast to the case with lagged exogenous variables, no estimate of a "skew" sum of months is needed, but instead the coefficient is changed from the monthly equation.

In our case with m=6, the change can be large. Thus, for φ=0.5, the coefficient in the semi-annual equation would be $\varphi^6 = 0.016$. It may be questioned if it is worth-while to keep the term at all.

For the exogenous variables, we must observe that in order to transform the dependent variable in a proper way, we have to use an operator that is different from A_1 and C, used in previous Sections. The proper operator for an AR(1) model is easy to find if we move the y_{t-1} term to the LHS. In order to change $(y_t - j\, y_{t-1})$ into $(Y_T - j^6 Y_{T-1})$, we have to multiply by

$$\frac{1-j^6 B_t^6}{1-jB_t} * \frac{1-B_t^6}{1-B_t}$$

where the latter factor represents the summation over six months, and we take y_t to be a flow variable. For the simple model

$$y_t = j\, y_{t-1} + bx_t + e_t \qquad (15)$$

we get, when Y_{T-1} is moved back to the RHS:

$$Y_T = j^6 Y_{T-1} + b\frac{1-j^6 B_t^6}{1-jB_t} X_T + \frac{1-j^6 B_t^6}{1-jB_t} E_T$$

where E_T is the sum of the corresponding monthly residuals.

It is seen that the coefficient of Y_{T-1} is φ^6 as in the case without exogenous variables. The X_T term is, however, somewhat complicated, and contains many "skew" sums that are not observed in the semi-annual case. As in previous cases, we may estimate them. If this is done again in the simplest way by linear interpolation, we get

$$\frac{b}{6}\left[\left(6+5j+4j^2+3j^3+2j^4+j^5\right)X_T + \left(j+2j^2+3j^3+4j^4+5j^5\right)X_{T-1}\right] \qquad (16)$$

which differs from b in a rather complex way. As an example, for φ=0.7 and b=0.5, we get the estimate of the semi-annual equation as

$$Y_T = 0.118Y_{T-1} + 1.095X_T + 0.376X_{T-1}$$

For this case, we have not calculated optimal coefficients for the situation when x_t is AR(1), and we cannot judge how robust the interpolation estimates are.

First differences of the exogenous variables are somewhat easier to handle than the variable values themselves, due to the fact that the linear interpolation gives the same result for all lags between 0 and 5, thus

$$B_t^i \sum (1 - B_t) x_t \approx \frac{(1 - B_T) X_T}{6} \qquad \text{for } i = 0,1,2,\ldots,5$$

This implies that a term with the first difference of a variable in the RHS of equation (15), say $c(x_t - x_{t-1})$, will be transformed to $c^*(X_T - X_{T-1})$ in the semi-annual equation with

$$c^* = \frac{1 - j^6}{1 - j} \frac{c}{6}$$

assuming that x_t is of the same kind of variables as y_t.

Finally, there remains the case with a lagged exogenous variable with coefficient *d* in the RHS of equation (15). The aggregation of such a term will result in an expression similar to (16), but with slightly different weights:

$$\frac{d}{6}\left[\left(5 + 4j + 3j^2 + 2j^3 + j^4\right)X_T + \left(1 + 2j + 3j^2 + 4j^3 + 5j^4 + 6j^5\right)X_{T-1}\right]$$

Both expressions can be somewhat simplified by using summation formulae for geometric series. For the resulting expressions, as well as for results regarding end-of-period stock variables, see Table 9, to be introduced in Section 9.

In an equation describing an error correction mechanism, as was exemplified by model (14), the exogenous variable enters with the contemporaneous as well as a lagged value. By rearranging the terms, the model can be written

$$y_t = (1-a)y_{t-1} + btx_t + (a-t)bx_{t-1} + \varepsilon_t \qquad (17)$$

Treating this equation in the same way as (15), we get additional terms corresponding to the x_{t-1} term in (17). These terms also contain sums of x which have no equivalent in X and thus have to be estimated. If we stick to linear interpolation, the final result is an equation of the same form as (14):

$$Y_T\text{-}Y_{T-1} = b\ t^*(X_T\text{-}X_{T-1}) + a^*(b\ X_{T-1}\text{-}Y_{T-1}) + E_T^*$$

where

$$a^* = 1 - (1-a)^6 \quad \text{and}$$

$$t^* = 1 - \frac{1-t}{6}\frac{a^*}{a} \qquad (18)$$

and E_T* is an error term. It is here assumed that x and y are variables of the same kind, either flow or period-average variables.

Other exogenous variables in the equation are affected by the aggregation in the same way as in equation (15), with φ replaced by (1-α).

8. Using Alternative Operators

When determining an operator that aggregates the monthly equation into a half-yearly one, we have so far chosen one that transforms the LHS of the equation exactly to the equivalent form of half-yearly variables. Then, some terms on the RHS have had to be estimated, since they contained variable values that do not correspond to the half-yearly values that are assumed to be available.

An alternative strategy would be to use any operator to both sides of the equation, and then estimate the resulting expressions. The operator could then be chosen in such a way that the transformations and the estimates are as simple as possible.

As an example, take a simple equation:

$$y_t\text{-}y_{t-1} = ax_t + e_t$$

The equation was aggregated by the operator $C(B_t)$, and the optimal coefficients for estimating $C(B_t)x_t$ are shown in Table 5. The interpolation expression is $(3.5X_T+2.5X_{T-1})$. However, we could try the operator $A_1(B_t)$ instead:

$$A_1(B_t)(y_t-y_{t-1}) = aA_1(B_t)x_t+A_1(B_t)e_t$$

We found on p. 20 that for a flow variable y_t, the interpolation estimate of a difference is $\frac{1}{6}\left(Y_T - Y_{T-1}\right)$, while $A_1\left(B_t\right)x_t = X_T$ exactly. Thus, neglecting the residual, we get $\frac{1}{6}\left(Y_T - Y_{T-1}\right) \approx aX_T$ or $Y_T - Y_{T-1} \approx 6aX_T$. Similar operations could be made with any expressions, although the result may not be so simple as in the present case.

If the result of the aggregation is judged by the increase in variance of the equation´s estimate of the LHS expression, then an operator that does not exactly transform the LHS must necessarily yield estimates that are inferior to the optimal ones that we have given in the tables 3-8, since these are derived as least squares estimates. That the difference can be substantial even in comparison with the previously derived interpolation estimates is shown by the example given above. Again neglecting the residual, assuming a=1, and x_t to be generated by an AR(1) process with variance =1, we get the following variances for the estimates of (Y_T-Y_{T-1}):

	Estimator	
	$(3.5X_T+2.5X_{T-1})$	X_T
γ	Variance	
0.1	40.33	128.25
0.5	77.78	279.85
0.9	132.24	783.56

The variance is increased between 3 and 6 times by using the non-optimal operator. Thus, we stick to the strategy of finding an operator that transforms the LHS exactly. We then have to estimate some terms of the RHS as has been shown in the previous Sections.

9. Summary and Conclusions

We summarize the results in Table 9 (p. 42). The columns represent various forms of the dependent variable, and the rows show the estimates of the multipliers that should be used for each form of the RHS variables. Exact values are written between →arrows← , linear interpolation results are given by formulas, and for optimal estimates when the variable is AR(1), the number of the relevant table is given.

It is understood that the dependent variable is a flow variable. If it is a period-average variable, all entries should be divided by 6. The case of an end-of-period variable as dependent is not covered in the present paper.
As a complement to the results for aggregation of six terms, Table 10 (p. 43) shows interpolation coefficients for the aggregation of two terms, e.g. from quarterly to semi-annual observations. It is similar in form to Table 9, and it covers the case when the dependent variable is a flow variable. If it is a period-average variable, all entries should be divided by 2.

When calculating optimal coefficients for estimation and the corresponding error variances, we have noted a marked difference in the results for flow and average variables on one hand, and end-of-period variables on the other. While the estimates for the former group generally performed fairly well, the end-of-period variables showed large error variances, particularly if their autocorrelation was low.

It seems clear that basing the estimates on every sixth observation gives very uncertain results. Could anything be done to improve the estimates?

As has been noted above, some variables that measure the state of some phenomenon are aggregated by taking averages of their values over the period. When this is the case, these variables can be treated in the same way as flow variables, resulting in better estimates than for a corresponding end-of-period variable. Thus, if in the model a state variable is measured by its average value over the period, the aggregation over time causes smaller errors than if it is measured by its end-of-period value.

From the point of view of economic theory, it seems in many cases to be quite acceptable to measure a stock variable by its period average. In an equation describing the reaction of some actors to the value of a stock variable, say wealth, the traditional method is to specify that it is the value in the beginning of the period that is relevant. This may be a realistic assumption for a relatively short period, but for a longer period, some kind of average seems natural. Thus, in the model building phase, it should be kept in mind that, if measuring a stock variable by its average instead of its ultimate value is acceptable from an economic point of view, it is definitely preferable from a statistical point of view, when the equation is to be aggregated over time.

It should be pointed out that this recommendation does not necessarily extend to the situation when the relation has to be estimated in its aggregated form and tests for Granger-causality are performed. In this case, Kirchgässner and Wolters (1992) have shown that spurious causality is less likely to be found if the dependent and the exogenous variable are aggregated in the same way, whether averaged or end-of-period. This is, however, a situation that is different from ours, where the relation has already been estimated on disaggregated data.

We will end this concluding Section by a practical application. The results of the paper are used for converting the monthly equation for the exchange rate into a semi-annual equation. The equation estimated on monthly data has the following form:

$$lvx_t - lvx_{t-1} = 0.49863(lrp_t - lrp_{t-1}) - 0.36243\,(lvx_{t-1} - lrp_{t-1}) - 18.7(rdiff_t - rdiff_{t-1}) - \\ -12.8915\, rdiff_{t-1} + 0.58920\, cap_t / m3sek_{t-1} + 0.07587 \qquad (19)$$

where lower case letters have again been used to designate monthly values of the variables:

lvx = log exchange rate (in SEK per foreign currency unit)

lrp = log(Swedish prices/foreign prices)

rdiff = interest rate difference Sweden - Germany

cap = change in net foreign liabilities in the private sector

m3sek = money stock

The first three of these variables are period-average variables (z type), *cap* is a flow variable (x type), while *m3sek* is an end-of-period stock variable (u type). The corresponding semi-annual variables will be designated by upper-case letters.

The equation is of the error correction type, with *lrp* as the norm for *lvx*. Thus, we have to use the third column of Table 9, remembering that the dependent variable is of the z type, so we have to divide all entries by 6. We should also note the modifications given in Section 7 for the error-correction case. In terms of the notation given there, we have α=0.36243, and thus φ=0.63757. Furthermore, τ=0.49863 and β=1. From (18), we can calculate

$$\alpha^{*}=0.9328 \text{ and } \tau^{*}=0.7849$$

For the other variables, we use the formulae in Table 9. Thus the coefficient for

$(1-B_T)RDIFF$	is -18.7*0.4290 = -8.0218
$B_T(RDIFF)$	is -12.8915*(1.5756+0.9983B_T)=-20.311-12.8692B_T
the constant	is 0.07587*2.5738=0.1953

The remaining term, $cap/B_t\ m3sek$, is more difficult. It is a ratio between two variables, which are not even of the same type. Since the mean over six months of the ratios is not the same as the ratio of the means, we have a problem of a kind similar to that of logarithms, discussed in Section 2. However, the variation in the denominator variable, *m3sek*, is small. Its standard deviation is only 2 per cent of its mean, so it is not very important which value of *m3sek* that enters the denominator. This means that the difference between $CAP_T/M3SEK_{T-1}$ and $S(cap_t/m3sek_{t-1})$ is not very large, and we disregard it. Thus, we can treat this ratio as a flow variable without much loss of information.

With this approximation, we get the coefficient for $CAP_T/M3SEK_{T-1}$ as

$$0.58920(0.3341 + 0.0949\ B_T) = 0.1969+0.0559B_T$$

To evaluate the approximations, we may calculate the estimates of LVX_T from semi-annual data to see how closely they trace the values obtained from monthly data. In order not to distort the comparison by the residuals of the monthly estimates, we make the calculations as if the monthly equation gave the true values of *lvx*, and thus all residuals were = 0. We still need a starting value for *lvx*, and we chose the observed value for June 1993. Using this, we can calculate *lvx* according to the monthly equation for all months of the second half of 1993. The average of these values is used as a starting point for the semi-annual equation. Unfortunately, this leaves only two observations for the comparison, the first and the second half of 1994. We get

	1/94	2/94
from monthly data	0.2382	0.2235

from semi-annual data	0.2340	0.2267
Difference	0.0042	- 0.0032

The difference could be compared with the standard error of the estimate. In the monthly equation this was 0.0163 for the estimate of $(1\text{-}B_t)lvx$. This translates to an error of 0.0262 for the semi-annual difference, computed from the monthly data. If for a one-step-ahead forecast the estimated difference is added to the observed value of the previous period, the standard error of the forecast will be the same as for the difference. Thus, an additional error of about 0.004 caused by using semi-annual data does not seem to be too serious.

To base a judgement of the reliability of the semi-annual approximations on only two observations seems perhaps not very assuring. Thus, we have also made a

Monte Carlo study of the difference between the monthly and the semi-annual estimates.

The variables *rdiff* and *lrp* were simulated using AR(1) models with autocorrelation coefficients close to those observed for these variables. For cap/B_tm3sek, no ARIMA model was found to fit, so we used an i.i.d. random model with average equal to that observed for the period July 1993 - Dec. 1994 (For details, see the Appendix).

Cutting off a few periods in the beginning of the series, we then calculated *LVX* according to the monthly as well as the semi-annual equation. The mean of the difference between these estimates was very close to 0, and its standard deviation around 0.0022. The two observations on actual data do not challenge this result. We can thus conclude that in this case, the estimates and approximations used for the semi-annual equation increase the residual variance of the *LVX* estimate only very slightly.

Table 9. Summary of coefficient multipliers in different situations

Dependent variable	y_t	$(y_t\text{-}y_{t-1})$	$(y_t\text{-}j\,y_{t-1})$
Operator	$A_1 = \frac{1-B_t^6}{1-B_t}$	$C = A_1^2$	$A_1 \frac{1-j^6 B_t^6}{1-j B_t}$
<u>RHS variables</u>			
Constant	$\rightarrow 6 \leftarrow$	$\rightarrow 36 \leftarrow$	$\rightarrow 6\frac{1-j^6}{1-j} \leftarrow$
x_t	$\rightarrow 1 \leftarrow$	Tab. 5 $3.5X_T + 2.5X_{T-1}$	$\frac{(6-7j+j^7)X_T+(j-6j^6+5j^7)X_{T-1}}{6(1-j)^2}$
z_t	$\rightarrow 6 \leftarrow$	$6 \times$ Tab. 5 $21Z_T + 15Z_{T-1}$	$6 \times$ (expression above)
u_t	$3.5U_T + 2.5U_{T-1}$	Tab. 6 $\frac{28}{3}U_T+\frac{70}{3}U_{T-1}+$ $+\frac{10}{3}U_{T-2}$	$\frac{1}{6}[\ (21+15j+10j^2+6j^3+3j^4+j^5)U_T+$ $+(15+21j+25j^2+27j^3+27j^4+25j^5)U_{T-1}+$ $+(j^2+3j^3+6j^4+10j^5)U_{T-2}\ ]$
x_{t-1}	Tab. 3 $\frac{5}{6}X_T+\frac{1}{6}X_{T-1}$	$2.5X_T + 3.5X_{T-1}$	$\frac{(5-6j+j^6)X_T+(1-7j^6+6j^7)X_{T-1}}{6(1-j)^2}$
z_{t-1}	$6 \times$ Tab. 3 $5Z_T + Z_{T-1}$	$15Z_T + 21Z_{T-1}$	$6 \times$ (expression above)
u_{t-1}	Tab. 4 $2.5U_T + 3.5U_{T-1}$	Tab. 7 $\frac{35}{6}U_T+\frac{146}{6}U_{T-1}+$ $+\frac{35}{6}U_{T-2}$	$\frac{1}{6}[\ (15+10j+6j^2+3j^3+j^4)U_T+$ $+(21+25j+27j^2+27j^3+25j^4+21j^5)U_{T-1}+$ $+(j+3j^2+6j^3+10j^4+15j^5)U_{T-2}\ ]$
$(x_t\text{-}x_{t-1})$	$\frac{1}{6}$	$\rightarrow 1 \leftarrow$	$\frac{1}{6}\frac{1-j^6}{1-j}$
$(z_t\text{-}z_{t-1})$	1	$\rightarrow 6 \leftarrow$	$\frac{1-j^6}{1-j}$
$(u_t\text{-}u_{t-1})$	$\rightarrow 1 \leftarrow$	Tab.8 $3.5(U_T\text{-}U_{T-1})+$ $+2.5(U_{T-1}\text{-}U_{T-2})$	$\frac{1}{6(1-j)^2}[\ 6-7j+j^7)(U_T-U_{T-1})+$ $+(j-6j^6+5j^7)(U_{T-1}-U_{T-2})\ ]$

The values are valid under the assumption that y_t is a flow variable. If it is a period-average variable, divide all entries by 6. If y is an end-of-period stock variable, the table does not apply.

→Figures← between arrows are exact equivalents. The table references give optimal estimates when the RHS variable is AR(1), the formulae are linear interpolation results.

Table 10. Interpolation coefficients for aggregation of 2 observations

Dependent variable	y_t	$(y_t - y_{t-1})$	$(y_t - j\,y_{t-1})$
Operator	$A_1 = \dfrac{1-B_t^2}{1-B_t}$	$C = A_1^2$	$A_1 \dfrac{1-j^2 B_t^2}{1-j B_t}$
RHS variables			
Constant	$\rightarrow 2 \leftarrow$	$\rightarrow 4 \leftarrow$	$\rightarrow 2\,(1+j) \leftarrow$
x_t	$\rightarrow 1 \leftarrow$	$1.5X_T + 0.5X_{T-1}$	$\left(1+\dfrac{j}{2}\right)X_T + \dfrac{j}{2}X_{T-1}$
z_t	$\rightarrow 2 \leftarrow$	$3Z_T + Z_{T-1}$	$(2+j)Z_T + j\;Z_{T-1}$
u_t	$1.5U_T + 0.5U_{T-1}$	$2U_T + 2U_{T-1}$	$\dfrac{3+j}{2}U_T + \dfrac{1+3j}{2}U_{T-1}$
x_{t-1}	$0.5X_T + 0.5X_{T-1}$	$0.5X_T + 1.5X_{T-1}$	$\dfrac{1}{2}X_T + \dfrac{1+2j}{2}X_{T-1}$
z_{t-1}	$Z_T + Z_{T-1}$	$Z_T + 3Z_{T-1}$	$Z_T + (1+2j)Z_{T-1}$
u_{t-1}	$0.5U_T + 1.5U_{T-1}$	$0.5U_T + 3U_{T-1} + 0.5U_{T-2}$	$\dfrac{1}{2}U_T + \dfrac{(3+3j)}{2}U_{T-1} + \dfrac{j}{2}U_{T-2}$
$(x_t - x_{t-1})$	0.5	$\rightarrow 1 \leftarrow$	$\dfrac{1}{2}(1+j)$
$(z_t - z_{t-1})$	1	$\rightarrow 2 \leftarrow$	$1+j$
$(u_t - u_{t-1})$	$\rightarrow 1 \leftarrow$	$1.5(U_T - U_{T-1}) + 0.5(U_{T-1} - U_{T-2})$	$\dfrac{2+j}{2}(U_T - U_{T-1}) + \dfrac{j}{2}(U_{T-1} - U_{T-2})$

The values are valid under the assumption that y_t is a flow variable. If it is a period-average variable, divide all entries by 2. If it is an end-of-period stock variable, the table does not apply.

$\rightarrow$Figures$\leftarrow$ between arrows are exact equivalents. The formulae give interpolation estimates.

Appendix

The following models were used for the exogenous variables in the Monte Carlo study mentioned on pp. 40-41:

$$rdiff_t = 0.96\ rdiff_{t-1} + 0.068 + a_t$$
$$lrp_t\ \ = 0.92\ lrp_{t-1} + 0.006 + b_t$$
$$cap_t/m3sek_{t-1} = -0.007 + c_t$$

where a_t, b_t, and c_t are i.i.d. logistically distributed with variances

var a_t = 0.074

var b_t = var c_t = 0.000053

For each of these variables, 6*N* "monthly" observations were generated. From these, *N* "semi-annual" observations were calculated. From an initial value of *lvx*, equation (19) - without residual - was used with the observations of the exogenous variables to generate 6*N* "monthly" *lvx* observations, which were averaged to *N* "semi-annual" observations. These were compared to the values obtained by using the "half-yearly" exogenous variable values with the coefficients given on pp. 38-39 to get approximate *LVX* values, again from an initial value.

N was set to 10 000. The average of the difference was 0.000028, and the standard deviation within the sample was 0.00216.

References

Amemiya, T & Wu, R.Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* 67, 628-632.

Boot, J.C.G. et al. (1967). Further methods of derivation of quarterly figures from annual data. *Applied Statistics 16, 65-75.*

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

Kirchgässner, G. & Wolters, J. (1992). Implications of temporal aggregation on the relation between two time series. *Statistical Papers* 33, 1-19.

Marcellino, M. (1996). Temporal aggregation, missing observations, outliers, and forecasting: A unifying non-model based procedure. *European Meeting of the Econometric Society, Istanbul.*

Markowski, A. (1996). The financial block in the econometric model KOSMOS. *National Institute of Economic Research Working Paper* 53, forthcoming.

Tiao, G.C. & Wei, W.S. (1976). Effect of temporal aggregation on the dynamic relationships of two time series variables. *Biometrika* 63, 513-523.

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

Weiss, A.A. (1984). Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics* 26, 271-281.

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

Tidsaggregering av ekonometriska ekvationer

Sammanfattning

I samband med uppbyggnaden av en finansiell sektor i Konjunkturinstitutets ekonometriska modell KOSMOS uppstod problemet att vissa samband utsatts för strukturella förändringar under senare år. Endast ett fåtal år kunde därför användas för skattning av dessa ekvationer. För att få fler observationer som underlag för skattningen användes då månadsdata, medan den reala delen av modellen är uttryckt i halvårsdata. Månadsekvationerna måste sedan transformeras för att kunna användas tillsammans med de övriga.

Denna transformation visar sig vara inte helt trivial annat än i vissa speciella fall. Man måste skilja mellan tre slags variabler, som uppför sig på olika sätt vid aggregering:

1. flödesvariabler, t ex BNP (betecknas med x)
2. lägesvariabler, uttryckta som genomsnittsvärden över perioden, t ex priser (z)
3. lägesvariabler, uttryckta som värde vid periodens slut, t ex kapitalstock eller lagervolym (u).

Det förutsätts att den beroende variabeln i ekvationen är av en av de två första typerna. Då kan samtidiga värden av exogena variabler av dessa två typer lätt aggregeras. Däremot uppstår problem dels när en variabel är av den tredje typen, dels när en variabel förekommer med ett tidsförskjutet värde. Den aggregerade ekvationens högra led kommer då att innehålla funktioner av de exogena variablerna som inte kan uttryckas med hjälp av halvårsvärdena, vilka är de enda som förutsätts vara tillgängliga inom ramen för modellen.

Man får då tillgripa approximationer eller skattningar av de önskade funktionerna. Sådana skattningar diskuteras i denna skrift. I texten redovisas dels enkla interpolationer, dels minsta-kvadrat-skattningar under antagande av att den exogena variabeln genereras av en AR(1)-process. Resultaten blir olika om vänstra ledet av ekvationen är ett enskilt värde av den beroende variabeln eller om den är en differens. Andra resultat erhålles också när ett tidsförskjutet värde av den beroende variabeln uppträder i högra ledet. En sammanfattning ges i tab 9, sid 42, för en aggregering av sex observationer, och i tab 10, sid 43, för två observationer.

I tabellerna 3-8 ges också varianserna för skattningsfelen, då den exogena variabeln genereras av en AR(1)-process, och man aggregerar sex observationer. Det visar sig att lägesvariabler av typ 3 ger väsentligt högre felvarianser än de övriga variabeltyperna. I avsnitt 9 diskuteras därför om sådana variabler kan undvikas vid modellbygge och ersättas av typ2-variabler.

Slutligen tillämpas resultaten på ekvationen för växelkursen i KOSMOS. Det visar sig att i detta fall approximationerna inte nämnvärt ökar ekvationens residualvarians.