Discussion Paper 1

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

January 1988

AGGREGATION OF TIME SERIES VARIABLES--A SURVEY

Clive W. J. Granger*

Institute for Empirical Macroeconomics
and University of Minnesota

*On leave from University of California-San Diego.

This material is based on work supported by the National Science Foundation under Grant No. SES-8722451. The Government has certain rights to this material.

Any opinions, findings, conclusions, or recommendations expressed herein are those of the author(s) and not necessarily those of the National Science Foundation, the University of Minnesota, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

1. Introduction

The first three sections of this survey will consider three types of aggregation that occur in the context of the analysis of time series, (a) "small-scale" aggregation, (b) "large-scale aggregation, and (c) temporal aggregation. An example of small-scale aggregation is of X_t is generated by an AR(1) equation, such as

$$X_t = \alpha_1 X_{t-1} + \varepsilon_{1t}$$

and similarly Y_t is AR(1)

$$Y_t = \alpha_2 Y_{t-1} + \varepsilon_{2t}$$

where ε_{1t}, ε_{2t} are each white noises and if X_t, Y_t are independent, so that ε_{1t}, ε_{2t} are independent, then what model does

$$s_t = X_t + Y_t$$

obey? The answer is found usually to be ARMA(2,1). Small scale aggregation involves sums of a few time series variables, which are not necessarily independent.

"Large-scale" aggregations will involve the sums of very many variables, such as total U.S. consumption, which is the sum of the consumptions by the many million individual families that make up the consumers in the economy. One may expect special properties for aggregates over very large numbers of components.

Temporal aggregation occurs when a variable is generated over a month, by some model such as an AR(1) process, but is only observed quarterly for example. The question arises, what model

will the temporally aggregated data obey? The answer will depend on whether the variable is a stock or a flow and is discussed in section 4.

The other sections of the survey consider special topics that are particularly relevant for the time series context, section 5 is on forecasting, section 6 on causation and section 7 on cointegration.

2. Small Scale Aggregation

Many of the results concerning aggregation of univariate series rely on the following simple result (from Granger and Morris [1976]):

Theorem 2.1: If X_{1t}, X_{2t} are MA(q) processes, generated by

$$X_{jt} = \sum_{k=0}^{q} a_{jk}\varepsilon_{j,t-k}$$

where the innovation processes are individually white noise (i.e., $\mathrm{cov}(\varepsilon_{jt},\varepsilon_{js})=0, t\neq s, j\neq k$) then the sum

$$S_t = X_{1t} + X_{2t}$$

is also an MA(q) process, and it will have a representation

$$S_t = \sum_{k=0}^{q} b_k \eta_{t-k}$$

where η_t is a white noise. There seems to be no simple relationship between the b_j and the component coefficients a_{jk}, $k = 1, 2$.

As the result holds for the sum of a pair of moving averages it clearly holds for the sum of any number of MA(q) variables. If ε_{1t}, ε_{2t} are correlated other than contemporaneously, the sum may be MA(q*) for any finite q*. For example, if

$$X_{1t} = \varepsilon_{1t} + b\varepsilon_{1,t-1}$$

and

$$X_{2t} = \varepsilon_{1,t-3} + c\varepsilon_{1,t-4}$$

then each series is MA(1) but the sum is MA(4). In the case of only contemporaneous correlation between innovations, the theorem should strictly state that S_t is MA(q^*), where $q^* < q$. For example, if

$$X_{1t} = \varepsilon_{1t} + b\varepsilon_{1,t-1}$$

$$X_{2t} = \varepsilon_{2t} + c\varepsilon_{2,t-1}$$

and $\mathrm{cov}(\varepsilon_{jt},\varepsilon_{ks}) = 0$ all $t \neq s$ and when $k \neq j$ then $\mathrm{cov}(S_t,S_{t-1}) = \mathrm{cov}(X_{1t},X_{1,t-1}) + \mathrm{cov}(X_{2t},X_{21t-1}) = b\ \mathrm{var}(\varepsilon_1) + c\ \mathrm{var}(\varepsilon_2)$ and clearly this <u>can</u> be zero, making S_t a white noise series. However, of $\mathrm{cov}(S_t,S_{t-1})$ is zero it may be thought of as a coincidental or low-probability event.

Darroch, Jirina, and McDonald [1986] have shown that if the moving averages each have a unit root, so that

$$X_{1t} = (1-B)a_1(B)\varepsilon_{1t}$$

$$X_{2t} = (1-B)a_2(B)\varepsilon_{2t}$$

where $a_j(B)$ $j = 1, 2$ are polynomials in B of order $q - 1$, then S_t will also contain a unit root and be of the same form. Thus, if the components are noninvertible moving averages, of this particular kind, than so will be their sum.

A univariate series X_{1t} will be said to be ARMA(p_1,q_1) if it is generated by

$$a_1(B)X_{1t} = b_1(B)\varepsilon_{1t}$$

where ε_{1t} is white noise, $q_1(B)$ is a polynomial of order p_1, $b_1(B)$ is a polynomial of order q_1 and $a_1(B)$, $b_1(B)$ have no common roots. The immediate generalization of the theorem for moving averages is:

<u>Theorem 2.2</u>

If

$$X_{1t} \sim \text{ARMA}(p_1,q_1)$$

$$X_{2t} \sim \text{ARMA}(p_2,q_2)$$

and $(\varepsilon_{1t},\varepsilon_{2t})$ is a bivariate white noise, then

$$S_t = X_{1t} + X_{2t}$$

is ARMA(m,n) where

$$m \le p_1 + p_2$$

$$n \le \max(p_1+q_2,p_2+q_1)$$

(proved in Granger and Morris [1976]). Thus, for example, if $X_{1t} \sim$ AR(1), $X_{2t} \sim$ AR(1) then generally S_t is ARMA(2.1) but in special cases these orders can be lower. If S_t is given by

$$a_s(B)S_t = b_s(B)\eta_t$$

then $a_s(B)$ will consist of the product of $a_1(B)$ multiplied by all the roots of $a_2(B)$ that are not also in $a_1(B)$. It follows that aggregates may "usually" be expected to have ARMA models rather than the simpler AR models. In particular if X_t is AR(p) but is observed with a white noise measurement error, the observed series will be ARMA(p,p), from this theory.

A particularly important consequence of these results is that if <u>any</u> component series contains a unit root (and thus is I(1)) so that $a_1(B) = (1-B)\alpha_1(B)$, for example, then the sum S_t will also be I(1).

Some generalizations of Theorem (2.2) have been provided by Peiris [1985] for the multivariate case who proves:

<u>Theorem 2.3</u>: If $\underline{X}_{1t}$ is a multivariate series with N components generated by

$$\underline{A}_1(B)\underline{X}_{1t} = \underline{C}_1(B)\underline{\varepsilon}_{1t}.$$

where $\underline{A}_1(B)$ is an $N \times N$ matrix in the lag operator B, each component of this matrix being a polynomial of order p_1 in B, and similarly the matrix $\underline{C}_1(B)$ consists of polynomials of order q_1, then $\underline{X}_{1t} \sim ARMA(p_1,q_1)$. If also $\underline{X}_{2t} \sim ARMA(p_2,q_2)$, and if the stacked vector $(\underline{\varepsilon}_{1t},\underline{\varepsilon}_{2t})$ is an 2N white noise vector (so that components are at most correlated contemporaneously), <u>and</u> if the matrix $\underline{A}_1(B)\underline{A}_2(B)$ is symmetric, then

$$\underline{S}_t = \underline{X}_{1t} + \underline{X}_{2t}$$

is ARMA(m,n), with m, n as given in Theorem (2.2). In particular the sum of two MA(q) vectors will also be MA(q). The symmetry condition is rather a stringent one and it is unclear what occurs when it does not hold.

A caveat about these results and some of those in later sections is that, although they are correct in theoretical situations they are of somewhat limited value is practice when analyzing an actual data series. One may assume that an ARMA(p,q) model is appropriate but the values of p and q have to be identified either using the methods proposed by Box and Jenkins [1970] or model size selection criteria such as AIC. In practice, the true p, q may be very large--as suggested by the aggregation results, for instance--but low values of p and q may provide an adequate approximate model. This suggests that researchers should not completely believe their identified models, and so should not be surprised if the results from aggregation theory do not work perfectly with estimated models.

3. Large Scale Aggregation

Many of the most important variables in macroeconomics are simple, unweighted sums or aggregates of very large numbers of components. Thus, for example, consumption of nondurable goods in the U.S. is the sum of this quantity over 80 million households and total corporate profits are the sum of profits for 2 1/2 million individual firms. If the components are all AR(1), with different parameters, and are independent, their sum will be ARMA(N,N-1) using the results in the previous section when there are N independent components. When N is the millions, the number

of parameters will be unreasonably large. Clearly, a model with fewer parameters will provide an adequate approximation, perhaps because individual AR(1) models will have similar parameter values and because roots may (almost) cancel for the AR and MA polynomials in B of the aggregate series. A different approach is to assume that the AR(1) parameters are drawn from some distribution, and then some curious results can occur. Consider the case where the jth component is AR(1), generated by

$$X_{jt} = \alpha_j X_{j,t-1} + \varepsilon_{jt}$$

where ε_{jt} is a zero-mean white noise, with variance $(\varepsilon_{jt}) = \sigma^2$ and the vector $(\varepsilon_{jt}, j=1,\ldots,N)$ consists of independent components. The spectrum of the sum

$$S_t = \sum_{j=1}^{N} X_{jt}$$

will be the sum of the individual spectra. If the α_j are assumed to be independently drawn from a distribution $F(\alpha)$, the spectrum of S_t will be approximately

$$(3.1) \qquad \bar{f}(\omega) = \frac{N}{2\pi}\,\sigma^2 \int \frac{1}{|1-\alpha z|^2}\, dF(\alpha)$$

where $z = e^{i\omega}$. (The assumption that all ε_{jt} have the same variance is easily relaxed and is of little consequence.) A reasonable assumed distribution for the α's is the beta distribution of the form

$$dF(\alpha) = \frac{2}{B(p,q)}\,\alpha^{2p-1}(1-\alpha^2)^{q-1} d\alpha$$

$$= 0 \text{ elsewhere}$$

$$0 \le \alpha \le 1,\ p,\ q > 0$$

each α_j lies in the region zero to one and so, each X_{jt} is stationary, with probability one. It is shown in Granger [1980b] that in this case S_t will be I(d), where d = 1-q/2. Thus, $(1-B)^d S_t$ has a stationary MA(∞) representation. Note that d will not be a positive integer as q > 0, and so S_t generally will be fractionally integrated. If $0 < q \le 1$, S_t will have an asymptotically infinite variance but if q > 1, S_t will have a finite variance. The point of this example is that series with unusual long-memory properties can arise from the aggregation of independent components. Some of the simplifying assumptions is this example can be relaxed without changes in the basic result. However, if $0 < \alpha_j < \bar{\alpha} < 1$ for all j, so that there is an upper bound to the α_j values, which is strictly less than one, then the fractional integration result is lost.

A more general result is found if the independence assumption is removed. The model considered is

$$X_{jt} = \alpha_j X_{j,t-1} + \beta_j W_t + \varepsilon_{jt}$$

where again the α_j are from the distribution $F(\alpha)$, the β_j are from some distribution with nonzero mean $\bar{\beta}$, and the common factor W_t is $I(d_\omega)$. It is then found that

$$S_t \sim I(d)$$

where d is the largest of the two terms, $1 - q + d_\omega$, (from the W component) and 1 - q/2 (from the ε component). In particular, if W_t is stationary, so that $d_\omega = 0$, then d = 1-q/2 as before.

It is perhaps interesting that S_t will not be I(1), as so frequently "observed" with macrovariables from aggregation of this form and with the beta distribution as here assumed. It is unclear if I(d) aggregates can occur with other distributions, having the property Prob($\alpha \geq 1$) = 0. However, the result that S_t is I(1-q/2) has to be interpreted with some care, as

$$S_t = \left(\sum_j \frac{\beta_j}{1-\alpha_j B}\right) W_t + \sum_j \frac{1}{(1-\alpha_j B)} \varepsilon_{jt}$$

and the first term can be approximated by

$$N\bar{\beta}\left[\int \frac{1}{1-\alpha B}\, dF(\alpha)\right] W_t. \tag{3.2}$$

It follows that, provided $\bar{\beta} \neq 0$, the first term will have variance of order N^2 whereas the second term has variance of order N, as seen from (3.1). Thus, for large N, the first term will dominate in size but the second term will provide the larger value of d, provided $d_\omega < q/2$. In this analysis, W_t is a "common factor," that occurs in the generating process for (almost) all components and because of this it provides the dominant component of the aggregate.

The implications of the existence of common factors in aggregation was studied in Granger [1987]. Simple regressions are investigated and the time-series properties are not given particular attention. The potential importance of common factors can be illustrated from the very simple case

$$Y_{jt} = x_{jt} + cz_t$$

where x_{jt} is independent of x_{kt}, $j \neq k$, and x_{jt} is also independent of z_t. Thus the observed valued of a variable for the jth unit, y_{jt}, consists of a common factor cz_t and an independent component x_{jt}. All series will be taken to be stationary and for this example, suppose also that $var(y_j) = 1$, $var(z) = 1$, so that $var(x_j) = 1 - c^2$. Denote

$$S_{xt} = \sum_{j=1}^{N} x_{jt}$$

and similar S_{yt} is the sum of the y's, the

$$S_{yt} = S_{xt} + Ncz_t$$

so that

$$var(S_{yt}) = N(1-c^2) + N^2c^2.$$

Two extreme cases can be considered. (i) z_t is observable but the x_{jt} are not. Thus, at the micro level y_{jt} is explained by just z_t and at the macro level S_{yt} is also explained by z_t, and (ii) the x_{jt} are observed, but z_t is not, so the micro regression explains y_{jt} by x_{jt} and at the macro level S_{yt} is explained by just S_{xt}. At the macro level it will always be assumed that only macro (aggregate) variables are available, not their components. The following table shows R^2 values for the case where $c^2 = 0.001$ and N is one million

	Case 1	Case 2
	z observed	x observed
	x not observed	z not observed
micro R^2	0.001	0.999
macro R^2	0.999	0.001

It is seen that a very high R^2 can be observed at the micro level but very small at the macro level (in case 2), or vice versa. In this example, the common factor is of very minor relevance at the microlevel and so could be found insignificant in an analysis of micro data, yet it dominates the macro relationship, having variance of order N^2 compared to a variance of order N from the S_x component. In a sense, the macrorelationship is simpler than the microrelationships, if essentially irrelevant terms are ignored. This still true when both x's and the common factor z_t are observed.

A variety of different models are considered in Granger [1987] but the main result can be illustrated by the following microrelationship

$$y_{jt} = x_{jt} + \beta_j z_t + \gamma_j w_t + \varepsilon_{jt}$$

where z_t, w_t are common factors and x_{jt}, ε_{jt} are independent components. Thus, at the macro level

$$S_{yt} = S_{xt} + \bar{\beta} z_t + \bar{\gamma} w_y.$$

The first component has a variance of order N and the other two have variances of order N^2 provided $\bar{\beta}$, $\bar{\gamma}$ are not zero. Some simple cases are (i) no common factors, z_t, $w_t \equiv 0$, R^2 at the

macro level can take any value but S_{xt}, S_{yt} and the residual $S_{\varepsilon t}$ are all perfectly normally distributed, due to the effect of central limit theorem, unless ε_{jt} have extraordinary distributions. (ii) both z_t, w_t are present and observed R^2 will be very near to one. (iii) both common factors present, z_t observed, w_t not observed, macro R^2 takes any value, depending on the relative importance of the two common factors, residuals not necessarily Gaussian. (iv) one or both common factors present, neither observed, R^2 will be very nearly zero.

As case (iii) seems to be the most likely to be observed (if all series are stationary rather than I(1)), one can conclude that common factors are present but are not all observed. It is interesting to ask what these common factors are, particularly the unobserved ones.

The results of this section provide an implied criticism of the "typical decision maker" theory used to suggest macroeconomic relationships from a microtheory. A behavioral equation for a typical consumer, say, is derived from basic microtheory, all consumers are considered to be identical and so the macrorelationship is just N times the micro one. It is seen that a badly misspecified macrorelationship can occur.

It is also suggested in Granger [1987] that nonlinear microrelations may become effectively linear relationships between observed aggregates. Consider the simple case where

$$y_{jt} = \alpha_0 + \alpha_1 x_{jt} + \beta(x_{jt}^2 - \sigma_x^2)$$

so that all microrelationships have the same coefficient, $E[x_{jt}] = 0$ and so there are no common factors and $\sigma_x^2 = var(x_{jt})$ is the same for all x_{jt}. In the aggregate

$$S_{yt} = \alpha_0 + \alpha S_{xt} + \beta \, Sum(x_{jt}^2 - \sigma_x^2)$$

but the last term is not observed, in general, and is poorly estimated by the observed quantities S_{xt}, $(S_{xt})^2$ --as a variance is little related to means or means squared. Thus, the final term becomes part of the residual in the equation, which is effectively linear.

4. Temporal Aggregation and Systematic Sampling

The discussion in this section follows Weiss [1984] which contains references to important work in this area. Suppose that the basic time interval for which a series is generated is unity, but that observations occur every k units (k>1), then the series may be said to have been "systematically sampled." For example, some price may be determined monthly but only recorded--or observed--quarterly, so that k = 3. Systematic sampling may be viewed as a type of temporal aggregation appropriate for "stock" variables. For a "flow" variable, a summation will occur over the k units before systematic sampling. An example is automobile production, which can be observed monthly or quarterly, the quarterly figure being the sum of the component monthly production figures.

Consider initially the ARMA(p,d,q) series Y_t generated by

$$a(B)(1-B)^d Y_t = b(B)\varepsilon_t$$

and suppose that

$$a(B) = \prod_{j=1}^{P} (1-\delta_j B)$$

where $|\delta_j| < 1$, all j.

If now the series is systematically sampled every k units, (for some integer k>1) then the new observed series, obeys an ARMA(p,d,r) model where

$$r = [(p+d)+(q-p-d)/k]$$

and [x] represents the integer part of x. Further, the AR polynomial for the sampled series is $a_k(B) = \prod_{j=1}^{P} (1-\delta_j^k B^k)$, where B is the unit lag operator. For example if Y_t is generated by the AR(1) model

$$(4.1) \qquad Y_t = \alpha Y_{t-1} + \varepsilon_t$$

on the unit interval, where $\mathrm{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$, $j \neq 0$ the sampled process on k intervals, Y_t^k will appear to be generated by

$$Y_T^k = \alpha^k Y_{T-1}^k + \varepsilon_T^k$$

where T = kt and

$$\mathrm{cov}(\varepsilon_t^k, \varepsilon_{T-j}^k) = 0, \ j \neq 0.$$

If $|\alpha| < 1$, then as k increases α^k will become small and Y^k will become nearly a white noise. It is generally true that if a stationary series is systematically sampled, it's memory will decline. However, if y_t is a random walk, so that $\alpha = 1$ in (4.1),

the sampled process Y_T^k will also be a random walk. More generally, the ARMA(p,d,q) process becomes approximately an MA(d,d-1) model for k large.

Somewhat similar results occur with temporal aggregation. An ARMA(p,d,q) process becomes ARMA(p,d,r) where

$$r = [(p+d+1)+(q-p-d-1)/k].$$

For example, an AR(1) process, becomes ARMA(1,1) and a random walk becomes an IMA(1,1) process. An early result in this area is due to Working [1960] who showed that for $k \geq 4$, the MA component was

$$\varepsilon_T + 0.25\ \varepsilon_{T-1}$$

to a close approximation.

For both cases it is seen that if the process is generated over small time intervals compared to the observation period, so that k is large, the AR component of the generating mechanism becomes unimportant, the unit root components remain unchanged and the moving average component simplifies but can remain relevant.

Weiss [1984] also considers a temporal aggregation of models with seasonal components and ARMAX models. For detailed results, see that paper. The ARMAX model considered is

$$a(B)(1-B)^d Y_t = c(B)(1-B)^f X_t + b(B)\varepsilon_t$$

$$D(B)(1-B)^f X_t = F(B)e_t$$

with C(0) = 0, so that X_t causes Y_t but not vice versa and there is no instantaneous relationship between X_t and Y_t. After systematic sampling and temporal aggregation, equations for the

maximum lag of the independent variable are provided by Weiss [1984] and are generally rather complicated. As k becomes large the relationship between Y_T^k and X_T^k takes the form

$$(4.1) \qquad (1-B^k)^d Y_T = (1-B^k)^f X_T + a_T$$

where a_T is MA(d-1) for systematic sampling (stock variables) and is MA(d) for temporal aggregation (flow variables). Again, simplification occurs when k is large and it may be particularly noted that temporal aggregation has produced a contemporaneous relationship.

5. Forecasting

A univariate stationary series Y_t has optimum (least-squares) one-step forecast $f_{t-1,1}$ based on the information set I_{t-1} given by

$$f_{t-1,1} = E[Y_t | I_{t-1}].$$

In this survey, just linear, one-step forecasts are considered. The forecast error is

$$e_{t-1,1} = Y_t - f_{t-1,1}$$

and a natural measure of (one-step) forecastability is

$$R^2 = 1 - \frac{\mathrm{var}(e)}{\mathrm{var}(Y)}.$$

If Y_t is integrated, then it is assumed that appropriate differencing is applied to produce a stationary series before forecasting is attempted. A convenient notation is to use $f^{(j)}$ for the optimum one-step forecast of Y_{jt}, based on the information set $I^{(j)}$ available at time t - 1. If

$$S_{yt} = \sum_{j=1}^{N} Y_{jt}$$

then the optimum forecast of S_t is

$$f^{(s)} = \sum_{j=1}^{N} f^{(j)}$$

assuming that the union of all the individual information sets $I^{(u)} = \bigcup_j I^{(j)}$ is available. In this, rather extreme case, in a sense nothing is lost by the aggregation, although R^2 for S may take almost any value depending on the extent which the $e^{(j)}$ are intercorrelated. However, the more typical case is that the full information set is not available after aggregation and so some forecasting ability will be lost. For example, it is clear from the result of the previous section that temporal aggregation generally reduces forecastability.

For ordinary small-scale aggregation, Kohn [1982] has obtained necessary and sufficient conditions for no loss of forecasting ability. Suppose that the component series Y_{jt} are written as a vector $\underline{Y}_t$ which is generated by the k-order vector autoregression

$$\underline{Y}_t = \sum_{j=1}^{P} \underline{A}_j \underline{Y}_{t-j} + \underline{\varepsilon}_t$$

where $\underline{\varepsilon}_t$ is a vector white noise and let S_{yt} be the sum of these components, so that

$$S_{yt} = \underline{i}' \underline{X}_t$$

where $\underline{i}$ is a vector of ones. S_{yt} can be forecast either from $I_{t-1}^{(1)}$: $\underline{Y}_{t-j}$, $j \geq 1$ or from $I_{t-1}^{(2)}$: $S_{y,t-j}$, $j \geq 1$. Thus, in the first information set all information is available, in the second

only lagged sums are available. Kohn proves that the two forecasts are identical only if there exist a sequence of constants α_j such that

(5.1) $\quad \underline{i}'\underline{A}_j = \alpha_j \underline{i}'$

and in this case, S_{yt} obeys the AR(p) model

$$S_{yt} = \sum_{j=1}^{P} \alpha_j S_{y,t-j} + \underline{e}_t$$

(5.1) is a very stringent set of conditions which are unlikely to hold exactly, so that usually S_{yt} will be forecast less well if the smaller information set is available compared to the larger one.

However, these results are less clearly relevant when large-scale aggregation occurs with common factors. The results of section 3 show that virtually all of the forecasting ability of the full information set is available if just the common factors are available.

The results give in this section are theoretical. In practice, where models have to be identified and estimated, relative forecasting abilities of different information sets can be less clean cut, as shown by the simulation results of Lutkepohl [1985].

6. Causation

A definition of "causation" that has found wide-spread application says that x_t causes y_{t+1} if y_{t+1} is better forecast from the information set $I_t^{(1)}$: y_{t-j}, $\underline{w}_{t-j}$, x_{t-j}, $j \geq 0$ than from the information set $I_t^{(2)}$: y_{t-j}, w_{t-j}, $j \geq 0$. Thus, x_{t-j} contains

information that is helpful in forecasting y_{t+1} and which is not in $I_t^{(2)}$. Strictly, x_t is a "prima facie" cause of y_{t+1} in mean with respect to the information set $I_t^{(2)}$. The definition is discussed in Granger and Hatanaka [1964], Granger [1980a] and elsewhere. As the definition is based on forecastability, aggregation can be disturbing as information sets are deformed.

With ordinary small-scale aggregation the following rules generally hold (i) if at the disaggregate level $x_{j,t}$ does not cause $y_{j,t+1}$ <u>and</u> $y_{j,t}$ does not cause $x_{j,t+1}$ for most j then $S_{x,t}$ will not cause $S_{y,t+1}$. (ii) If $x_{j,t}$ causes $y_{j,t+1}$ but $y_{j,t}$ does not cause $x_{j,t+1}$ for most j, then S_{xt} will generally cause $S_{y,t+1}$ and $S_{y,t}$ may appear to cause $S_{x,t+1}$. Thus correct causality is still found but a spurious feedback may occur because of aggregation. However, aggregation may weaken the correct causality, as found for forecastability in section 5. (iii) If there is feedback at the disaggregate level it will theoretically occur at the aggregate level, in general.

To illustrate the statements in (ii) consider the simple situation where there are two micro-units and for each a pair of variables, X, Y are measured, these being related by

$$Y_{it} = a_i \varepsilon^X_{i,t-1} + \varepsilon^Y_{it}, \quad i = 1, 2$$

$$X_{i,t} = \varepsilon^X_{it} + b_i \varepsilon^X_{i,t-1}, \quad i = 1, 2$$

where $\underline{\varepsilon}^X$, $\underline{\varepsilon}^Y$ are independent (2×1) vector white noise processes. If $S_y = y_1 + y_2$, and using a similar notation for other variables, then

$$S_{yt} = \underline{a}'\underline{\varepsilon}^X_{t-1} + S_{\varepsilon Y,t}.$$

$S_{xt} = S_{\varepsilon x,t} + \underline{b}'\underline{\varepsilon}^X_{t-1}$ where $\underline{a}' = (a_1,a_2)$, $\underline{\varepsilon}^X = \left(\frac{\varepsilon^X_1}{\varepsilon^X_2}\right)$, $\underline{b}' = (b_1,b_2)$. Causality will be found at the macrolevel if

$$E[S_{y,t+1}S_{xt}] \neq 0$$

which will be true if

$$a_1(\sigma_1^2+p\sigma_1\sigma_2) + a_2(\sigma_2^2+p\sigma_1\sigma_2) \neq 0 \tag{6.1}$$

where

$$\sigma_1^2 = \mathrm{var}(\varepsilon^X_1),\ \sigma_2^2 = \mathrm{var}(\varepsilon^X_2)$$

$$\mathrm{cov}(\varepsilon^X_1,\varepsilon^X_2) = p\sigma_1\sigma_2.$$

Clearly (6.1) will usually occur (but need not), so that the causation that occurs at the microlevel will be found at the macrolevel. However, it will also be generally true that $E[S_{x,t+1},S_{y1t}] \neq 0$ as it is a perfect linear combination of $S_{yt} - S_{\varepsilon x,t} = \underline{a}'\underline{\varepsilon}^X_{t-1}$ and $S_{\varepsilon x,t-1}$ which is a component of $S_{x,t-1}$. Thus although there is no causation from Y_{jt} to $X_{j,t+1}$ at the microlevel, it does occur at the macrolevel.

It is hardly surprising that temporal aggregation can be disruptive of causal relations as past and future values get mixed up, part of one aggregate will occur both before and after part of another aggregate. If at the unit interval there is one-way causation between a pair of series, after temporal aggregation a feedback or two way causation may be found. As k, the number of time units being aggregated over becomes large, stationary series

appear to be just contemporaneously related and so actual causality can be lost.

7. Cointegration

A pair of I(1) series X_t, Y_t are said to be cointegration if there exist a linear combination

$$z_t = X_t - AY_t$$

which is stationary (or I(0)). This will occur if X_t, Y_t both possess a common I(1) factor, with all other components being I(0), so that

$$X_t = AW_t + \tilde{X}_t$$

$$Y_t = W_t + \tilde{Y}_t$$

where $W_t \sim I(1)$, $\tilde{X}$, $\tilde{Y}$ both I(0). If the series are cointegrated there will always exist an error-correction mechanism of the form

$$\Delta X_t = -\rho_1 z_{t-1} + \text{lagged } \Delta X_t, \Delta Y_t + \text{residual}$$

$$\Delta Y_t = -\rho_2 z_{t-1} + \text{lagged } \Delta X_t, \Delta Y_t + \text{residual}$$

where at least one of ρ_1, ρ_2 is nonzero.

As integrated series remain integrated under temporal aggregation, it is clear that cointegration remains true for series which are so aggregated. However, the form of the error-correction models may be altered.

The effects of small-scale cross-sectional aggregation has been studied by Gonzalo [1988]. For convenience, denote a pair of series X_{jt}, Y_{jt} for "state" j, which are aggregated into

$$S_{xt} = \sum_j X_{jt}$$

and similarly S_{yt}.

To show that a variety of results can occur it should be noted that (i) there can be cointegration at the aggregate level but not at the disaggregated states, and (ii) can be cointegration at the states but not at the aggregate level. To illustrate (1) suppose there are just two states (j=1,2), that S_{xt}, S_{yt} are cointegrated and that $X_{1t} = S_{xt}$, $X_{2t} = 0$ all t, $Y_{1t} = 0$ all t and $Y_{2t} = S_{yt}$. More generally, X_{2t} and Y_{1t} can both be I(0). To illustrate (ii), suppose that all X_{1t}, Y_{it}, i = 1, 2 are I(1), $z_{1t} = X_{1t} - A_1Y_{1t}$, $z_{2t} = X_{2t} - A_2Y_{2t}$ are both I(0) with $A_1 \neq A_2$, then

$$S_{xt} - \alpha S_{yt} = z_{1t} + z_{2t} + (A_1-\alpha)Y_{1t} + (A_2-\alpha)Y_{2t}$$

so that S_{xt}, S_{yt} are cointegrated only if both $A_1 = \alpha$ and $A_2 = \alpha$ and this occur only if $A_1 = A_2$, which is excluded by assumption. This result would not hold if Y_{1t}, Y_{2t} are cointegrated.

Gonzalo [1988] has sufficient reasons for cointegration at one level of cointegration to imply cointegration at another (higher or lower) level. For example, suppose that all X_{it}, Y_{it} are I(1) with Wold representations

$$(1-B)X_{it} = C_{xi}(B)\varepsilon_{it}$$

$$(1-B)Y_{it} = C_{yi}(B)\eta_{it}$$

stacking all the residuals produces a 2N × 1 vector

$$r_t' = (\varepsilon_{1t}, \eta_{1t}, \varepsilon_{2t}, \eta_{2t}, \ldots, \varepsilon_{Nt}, \eta_{Nt})$$

with covariance matrix $E[r_t r_t'] = \sum$.

If S_{xt}, S_{yt} are cointegrated it is shown that a sufficient condition for all X_{1t}, Y_{1t} to be cointegrated is that $\sum$ be of full rank. This result assume no cointegration of components across states. If, for example, Y_{it}, Y_{jt} are cointegrated for some i, j there are more possibilities.

A simple case has been considered by Lippi [1987] for which Y_{it}, Y_{jt} are cointegrated for all i, j, and X_{it}, Y_{it} are cointegrated for all i it follows that X_{it}, X_{jt} are cointegrated for all i ,j. Thus, there is a single common I(1) factor causing all X_{it}, Y_{it} components to be I(1). In this case S_{xt}, S_{yt} will also be I(1) because of this common factor and so will be cointegrated.

8. Conclusion

It is found that aggregation, both cross section and temporal, is inclined to simplify relationships, with some properties being quite robust (integration, cointegration for example) but others to be less-robust (causality, nonlinearity). For large scale aggregation, the distribution properties of equation residuals can be a useful indicator of the precise or not of common factors.

References

Darroch, J., M. Jirina, and J. McDonald, 1986, "The Sum of Finite Moving Average Processes," Journal of Time Series Analysis 7, 21-5.

Gonzalo, J., 1988, "Cointegration and Aggregation," working paper, Economics Department, University of California, San Diego.

Granger, C. W. J., 1980a, "Testing for Causation: A Personal Viewpoint," Journal of Economic Dynamics and Control 2, 329-52.

Granger, C. W. J., 1987, "Implications of Aggregation with Common Factors," Econometric Theory 3, 208-22.

Granger, C. W. J., 1980b, "Long-Memory Relationships and the Aggregation of Dynamic Models," Journal of Econometrics 14, 227-38.

Granger, C. W. J., and M. Hatanaka, 1964, Spectral Analysis of Economic Time Series, Princeton University Press.

Granger, C. W. J., and M. J. Morris, 1976, "Time Series Modeling and Interpretation," Journal of Royal Statistics Society A, vol. 139, 246-57.

Kohn, R., 1982, "When is an Aggregate of a Time Series Efficiency Forecast by its Past?," Journal of Econometrics 18, 337-49.

Lippi, M., 1987, "A Note on Aggregation of Log-Linear Error-Correction Models," working paper, Dipartimento di Economia Politicia, Universita di Modera, Italy.

Lutkepohl, H., 1985, "Comparison of Predictors for Temporally and Contemporaneously Aggregated Multivariate Time Series," working paper, Department of Economics, University of California, San Diego.

Peiris, M. S., 1987, "On the Study of Some Functions of Multivariate ARMA Processes," Stochastic Processes and Their Applications

Weiss, A., 1984, "Systematic Sampling and Temporal Aggregation in Time Series Models," Journal of Econometrics 26, 271-81.

Working, H., 1960, "Note on the Correlation of First Differences of Averages in a Random Chain," Econometrica 28, 916-18.