

TEMPORAL AGGREGATION AND STRUCTURAL INFERENCE
IN MACROECONOMICS
A COMMENT

JAMES H. STOCK

Kennedy School of Government, Harvard University

and

The Hoover Institution

The pitfalls of drawing econometric inferences using macroeconomic data have been an enduring issue in this conference series. In the inaugural volume, Lucas (1976) argued that the econometric equations underlying a generation of macroeconomic models, which at the time were given "structural" interpretations, actually were not structural at all but simply captured correlations that are likely to change from one policy regime to another. In assessing the next generation of macroeconomic time series models, Leamer (1985) questioned the use of vector autoregressions for drawing causal macroeconomic inferences. In their paper, Christiano and Eichenbaum examine another reason for caution in interpreting macroeconomic results: the bias that can arise from using temporally aggregated data to estimate structural time series models.

Christiano and Eichenbaum's examination of temporal aggregation is timely for three reasons. First, while there has been considerable work developing the theoretical properties of econometric modeling with temporally aggregated data, there have been few empirical attempts to quantify the importance of temporal aggregation bias in practice. Second, more powerful computational techniques than those used in earlier continuous time modeling efforts (e.g., Bergstrom [1976]) are just now becoming widely available. Third, most macroeconomic research uses time series data that represent a temporal aggregate, such as GNP. An assessment of the practical importance of temporal aggregation bias is certainly in order.

The basic difficulty involved with using temporally aggregated data is that temporal aggregation has the effect of changing the serial correlation patterns in the data. Perhaps the simplest case of this is Christiano and Eichenbaum's example of testing whether an exchange rate is a random walk vs. some nonstationary process with additional short-run dependence. Suppose one had monthly data on average daily exchange rates, and chose to test the pure random walk hypothesis by regressing the first difference of this monthly series on its lag and then checking the t-statistic on the lagged first difference. As Working (1960) showed algebraically and as Christiano and Eichenbaum suggest using actual exchange-rate data, even if the null hypothesis were true this test might lead one spuriously to reject the pure random walk hypothesis in favor of one in which the exchange rate exhibited additional short-run dependence. However, this conclusion would be misguided: by aggregating and then taking the first difference of the data, adjacent observations in the monthly first differenced series (say, July minus June and June minus May) would both contain the same daily innovations in exchange rates for the intervening month (June). Thus the transformation induces correlation leading to spurious rejection of the pure random walk hypothesis.

The theoretical effects of temporal aggregation have been known for some time (e.g., Sims [1971] and Geweke [1978]). What makes Christiano and Eichenbaum's paper exciting is its two applications -- both of intrinsic economic interest -- in which aggregation bias apparently matters. In the first, the stylized view of money Granger-causing output seems to be overturned by moving to a finer timing interval. In the second, using finer timing intervals and temporally disaggregated data leads to a much shorter adjustment time for inventories than appears elsewhere in the literature, one much more in line with the observations that inventories typically constitute only a few days worth of production.

The results of both examples are consistent with the theoretical predictions of the effects of temporal aggregation. Thus I would like to address not the interpretation of the effects of temporal aggregation bias on the empirical results, but rather whether the results, so interpreted, shed useful light on the underlying economic issues.

MONEY-OUTPUT CAUSALITY

Christiano and Eichenbaum find that, at the monthly level, the support

for the hypothesis that money Granger-causes output is limited, although at the quarterly level it is not. This conclusion alone is enough to cause concern in interpreting earlier findings about causality such as those of Sims (1980) or Litterman and Weiss (1985). There are, of course, other ways in which these existing causality results might not be robust. I shall explore one: the possibility that the relationship between money and income has changed over the postwar period.

To investigate this possibility, Table 1 contains the same test statistics that are reported in Christiano and Eichenbaum's Table 2, but for ten-year samples arbitrarily ending in election years. Column (1) presents the p-values for the money causality tests using monthly growth rates of industrial production and M1; the quarterly aggregated results are reported in column (2) (using a VAR(12) on the aggregated data) and in column (3) (using a VAR(6)). Three features of these results are relevant. First, using these ten-year samples, the evidence that temporal aggregation reverses the conclusions of the "standard" quarterly results is much weaker than is suggested by Christiano and Eichenbaum's Table 2. While the p-values drop (as predicted from the theory of temporal aggregation, if in fact there is no Granger causality from money to industrial production) in three of the sample periods, in the other three they rise. Second, this qualitative feature is not sensitive to using only six rather than twelve lags of money and income in the regressions. Third, the p-values appear to be very unstable over time. During the 1950s and 1970s, money was a useful predictor of monthly output, given lagged output; this does not, however, seem to have been the case in the 1960s.

These results, taken together with Christiano and Eichenbaum's Table 2, present a puzzle. If money is causally prior in the early and later subsamples, why is it not (at the monthly level) over the entire period? One answer is that monetary technologies -- and Federal Reserve policy -- have been changing over this period, and that empirical analysis based on the entire postwar history "averages over" these changes. Some evidence supporting this view is provided in columns (4) and (5). Column (4) contains the sum of the coefficients on industrial production growth in the regression of monthly industrial production growth on its lags and on lagged money growth, while column (5) contains the corresponding sum for the coefficients on money. While this quantity was negative for income during the 1950s and early 1960s, in the later periods it was positive. Similarly, the magnitude of the sum of the money coefficients has fluctuated substantially. Of course, these calculations alone do not tell

Table 1

Money Industrial Production Causality: Significance Levels for Ten-Year Subsamples

$$\text{VAR: } \begin{bmatrix} \Delta \ln IP_t \\ \Delta \ln MI_t \end{bmatrix} = \phi + A(L) \begin{bmatrix} \Delta \ln IP_{t-1} \\ \Delta \ln MI_{t-1} \end{bmatrix} + u_t$$

	(1)	(2)	(3)	(4)	(5)
Sample	M	Q	Q	M	M
Period	VAR(12)	VAR(12)	VAR(6)	A ₁₁ (1)	A ₁₂ (1)
54-64	.053	.360	.123	-.435	3.896
58-68	.502	.419	.479	-.196	1.423
62-72	.295	.094	.104	.586	.777
66-76	.130	.059	.079	.537	.983
70-80	.024	.097	.060	.510	.832
74-84	.053	.717	.341	.564	1.251
54-84	.784	.062	.187	.424	.292

Notes: "M" and "Q" refer to monthly and quarterly estimation, where the quarterly data were computed by averaging the monthly observations. VAR(12) and VAR(6) denote the number of lags used in VAR's from which the p-values reported in the table were computed.

us whether these shifts are statistically significant. (The estimates of these sums over the entire sample need not be the average of the subsample estimates because of the changing money-income correlations.)

These results suggest a relationship between income and money that has changed over the postwar period, a view consistent with other evidence in the literature. For example, in their investigation of the money-income relationship, Litterman and Weiss (1985) found statistically significant shifts in the coefficients of three- and four-variable VAR's with monthly industrial production, money, an interest rate, and inflation when they split their sample in 1966. Accordingly, it seems that Christiano and Eichenbaum's conclusion that money does not cause income at the monthly level obtains only because the causal patterns evident in decade-level data are obscured by treating the postwar data as if generated by a single, time-invariant mechanism, a view apparently inconsistent with the data.

INVENTORIES AND SALES

Christiano and Eichenbaum's second example is apt, given the broader context of their paper, since it examines a model designed to address Lucas' (1976) criticism by estimating "deep structural" parameters: here, the six describing the preferences, technology, and costs of the representative infinitely-lived consumer and price-taking firm comprising the United States nondurable manufacturing sector. This example is methodologically intriguing, since it points both to one feature of their model that is not robust to temporal aggregation and to another that is. On the one hand, they conclude that the estimated speed of adjustment to a target level of inventories is highly sensitive to the level of temporal aggregation. On the other hand, a conclusion that is robust to temporal aggregation is that the proposed model of inventories is not consistent with United States time series data. To support this second conclusion, they point to the large serial correlation remaining in the residuals from both the continuous and discrete models, indicating that there evidently are important sources of temporal dependence in the data that are not treated in the model.

Further evidence against the model comes from considering the long-run behavior of the series. As an empirical matter, (log real) sales and inventories appear to be cointegrated as defined by Engle and Granger (1985). That is, applying Dickey and Fuller's (1979) $\hat{\tau}_r$ test to sales and inventories of manufacturers and wholesalers indicates that each of these series contains a unit root. This is consistent with the small continuous

time roots ($\alpha \approx \beta \approx .08$) reported in the paper. However, applying either Engle and Granger's (1985) or Stock and Watson's (1986) test for the number of unit roots in multiple time series to inventories and sales yields a rejection (at the 5% level) of the null hypothesis that there are two unit roots in the bivariate system, against the alternative that there is but one. The optimizing model, which assumes stationary processes, thus seems to be misspecified for the nonstationary data at hand. Moreover, there is reason to suspect that inferences based on the misspecified system might be misleading, since estimators involving cointegrated variables can have distributions that are very different than they would have were the variables stationary. Providing this broad array of evidence against this particular inventory model is perhaps the most important contribution of the paper, one made stronger by its robustness to issues of temporal aggregation.

III.

The literature on temporal aggregation bias gives the impression that all time series properties of data change once they have been temporally aggregated. This is not so, and I would like to focus on one property of economic interest that remains unaltered by temporal aggregation, cointegration. Suppose that the cointegrated variables, sales and inventories, have a representation in terms of a common random-walk component in continuous time:

$$I(t) = \gamma_1 \mu(t) + \xi_1(t)$$

$$s(t) = \gamma_2 \mu(t) + \xi_2(t)$$

where $\xi_1(t)$ and $\xi_2(t)$ are stationary continuous time processes and $\mu(t)$ is a continuous time random walk with drift: $d\mu(t) = \delta dt + d\eta(t)$, where $\eta(t)$ is a Wiener process and δ is a constant. Since $\mu(t)$ is an integrated process, both $s(t)$ and $I(t)$ are integrated processes. However, $s(t) - (\gamma_2/\gamma_1)I(t)$ is not integrated in continuous time, so $s(t)$ and $I(t)$ in this framework are cointegrated. Clearly this property of cointegration survives if $s(t)$ and $I(t)$ are observed at points in time. But does it survive if, say, $I(t)$ is observed at a point in time, but $s(t)$ is observed as an average over h time periods. It does: letting $s_{t+h} = \int_{\tau=t}^{t+h} s(\tau) d\tau$, since

$$\mu(t+h) = \delta h + \mu(t) + \int_{\tau=t}^{t+h} d\eta(\tau),$$

$$I_{t+h} = \gamma_1 \mu(t) + \gamma_1 \delta h + \gamma_1 \int_{\tau=t}^{t+h} d\eta(\tau) + \xi_1(t+h)$$

$$s_{t+h} = h \gamma_2 \mu(t) + \frac{1}{2} \gamma_2 \delta h^2 + \gamma_2 \int_{\tau=t}^{t+h} (t+h-\tau) d\eta(\tau) + \int_{\tau=t}^{t+h} \xi_2(\tau) d\tau$$

Evidently $s(t) - (h\gamma_2/\gamma_1)I(t)$ cancels out the nonstationary term $\mu(t)$; even with arbitrary temporal aggregation, observed inventories and sales are cointegrated in discrete as well as continuous time. While temporal aggregation bias can be important when examining short-run characteristics of the data, such as adjustment times, it does not affect long-run population properties such as cointegration.

One conclusion from Christiano and Eichenbaum's paper is that the specification and estimation of continuous time models provides a feasible answer to the challenge of temporal aggregation bias, and that continuous time modeling at a minimum provides a useful complement to the usual discrete time models. Their argument that ignoring temporal aggregation issues can lead to severe misinterpretations of empirical results is persuasive. Of course, as their examples make clear, estimation using a finer timing interval or using a continuous time model is not a panacea. However, the specification and estimation of continuous time models does provide an avenue for addressing the problems inherent with using temporally aggregated data.

REFERENCES

- Bergstrom, A.R.
(1976) *Statistical Inference in Continuous Time Models*. Amsterdam: North-Holland.
- Dickey, D.A. and Fuller, W.A.
(1979) Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74: 427-431.
- Engle, R.F. and Granger, C.W.J.
(1985) Dynamic Model Specification with Equilibrium Constraints: Cointegration and Error Correction. University of California, San Diego, Discussion Paper #85-18.
- Geweke, J.B.
(1978) Temporal Aggregation in the Multivariate Regression Model. *Econometrica*, 46: 643-661.
- Leamer, E.E.
(1985) Vector Autoregressions for Causal Inference? *Understanding Monetary Regimes*, (Eds.) K. Brunner and A.H. Meltzer, Carnegie-Rochester Conference Series on Public Policy, 22. Amsterdam: North Holland.
- Litterman, R.B. and Weiss, L.
(1985) Money, Real Interest Rates, and Output: A Reinterpretation of U.S. Postwar Data. *Econometrica*, 53: 129-156.
- Lucas, R.E. Jr.
(1976) Econometric Policy Evaluation: A Critique, *The Phillips Curve and Labor Markets*, (Eds.) K. Brunner and A.H. Meltzer, Carnegie-Rochester Conference Series on Public Policy, 1. Amsterdam: North Holland.
- Sims, C.A.
(1971) Discrete Approximations to Continuous Time Distributed Lags in Econometrics. *Econometrica*, 39: 545-563.

Sims, C.A.

(1980) Macroeconomics and Reality. *Econometrica*, 48: 1-34.

Stock, J.H. and Watson, M.W.

(1986) Testing for Common Trends. Harvard Institute for Economic Research, Discussion Paper #1222.

Working, H.

(1960) Note on the Correlation of First Differences of Averages in a Random Chain. *Econometrica*, 28: 916-918.