Improving Econometric Forecasts
By Using Subperiod Data

Paul A. Anderson
Thomas M. Supel*

Staff Report #: 21

PACS File #: 2930

April 1977

*Research Fellow and Senior Economist, Federal Reserve Bank of Minneapolis. The views expressed herein are solely those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. The authors would like to thank Lizbie Lin and Nels Johnson for invaluable programming and computation assistance, and John Geweke and our colleagues, especially Jim Duprey, Preston Miller, Art Rolnick, and Thomas Sargent, for helpful comments.

Abstract

The method proposed here includes two innovations which should improve the accuracy of econometric forecasting. First, it replaces the subjective, judgmental adjustments commonly used with a more formal, objective econometric procedure. Second, it includes a methodology for testing the usefulness of subperiod data which forecasters often inspect when choosing intercept adjustments. A sample application to the MIT-Penn-SSRC Model demonstrates that the procedure is both feasible and potentially helpful in the context of a large macroeconometric model.

Improving Econometric Forecasts
By Using Subperiod Data

It is widely recognized that existing large-scale structural econometric models do not forecast well when operated mechanically. For example, Nelson [8], Cooper and Nelson [1], and Cooper [2] present evidence that existing models can be "beaten" by low-order autoregressive forecasts. In spite of this evidence, structural models continue to be used for several reasons: a) reduced forms are inadequate for certain kinds of policy evaluations, b) the available time series are too short to permit direct estimation of the reduced forms of any but the simplest models, and c) economic theory usually implies restrictions on the parameters of structural equations.

Because of their weak mechanical forecasting performance, large models are typically kept "on track" by adjusting the constant terms of the structural equations. It is unlikely that the systematic errors in mechanical forecasts are solely the result of the parallel shifts of schedules that this practice implicitly assumes. However, the method of constant (or "con") adjustments is widely used as a convenient means of reducing forecast error while leaving the policy response characteristics of a model substantially intact.

Determination of the size of the adjustment for the initial forecast quarter, and its projection to future quarters, is usually accomplished by a process which blends mechanical processing of equation residuals with judgment based on outside data and nonquantifiable information. In general, the model forecasts are not reproducible by anyone other than the model's proprietor. Haitovsky, Treyz, and Su [6] discuss this adjustment process in some detail, and Green [5] describes a mechanical procedure in the context of the OBE model.

We hypothesize that subperiod data which become available during the first period of a multiperiod forecast contain useful information about shifts in the structural equations. We have developed a methodology for testing whether a set of subperiod data help predict the "constants" in the structural equations of a given model and for incorporating those data that pass the test into the model forecasts. This paper describes our procedure and presents an application of our method to the problem of using monthly data to improve the forecasts of the quarterly MIT-Penn-SSRC (MPS) model.

We obtain a substantial reduction in the mean-squared forecast errors (MSEs) of real GNP, the implicit deflator, and the unemployment rate by using our method of constant adjustments. For one-quarter-ahead forecasts, our procedure leads to a 65 percent reduction in the mean-squared forecast error (MSE) of the GNP deflator, but the MSE for real GNP turns out to be larger. However, for two-, three-, and four-quarter-ahead forecasts, we obtain sizeable reductions in the MSE for all three primary system variables. The reductions in the MSEs range from 25 to 54 percent for real GNP, from 13 to 36 percent for the GNP deflator, and from 25 to 54 percent for the unemployment rate. A second set of simulations indicate that a great deal of reduction in the forecast error variance can be obtained using time series methods of Granger and Newbold [4], even without resorting to monthly data.

Our evidence indicates that consideration should be given to the estimation of mixed monthly-quarterly models with a higher-order error structure. But until such models are available, our procedure should be of interest to forecasters and to policy analysts who wish to incorporate the most recent monthly data into their projections.

The next section describes the rationale underlying our procedure and the following section describes the empirical results. The data and procedural details are relegated to the appendix.

Improving Econometric Forecasts

The hypothesis that subperiod data may be used to update an econometric forecast is based on both theoretical and practical considerations. Discrete-time econometric models are commonly viewed as approximations to a dynamic system which operates over shorter time intervals than the quarterly or monthly periods imposed, usually, by data availability. Most models reflect the quarterly National Income Accounting interval; more frequently available data (typically financial and employment data) are averaged to a quarterly basis. The conditions under which the structure of the "finer" model may be inferred from the aggregate or "gross" model are very stringent (see Skoog [10] and Geweke [3]). This situation opens up the possibility that the information in the detailed time record of monthly and weekly observations of certain variables may not have been incorporated into forecasts efficiently.

As a practical matter, the authors have found inspection of monthly and weekly data helpful in making constant adjustments to the MPS model for forecasting. A complete integration of subperiod data into a quarterly model would require the formulation of a well-articulated model on, say, weekly intervals and the deduction of the form of the relation between quarterly, monthly, and weekly time series on the basis of that consistent structure. We satisfy ourselves, for the time being, with a direct regression of the quarterly structural equation residuals on the monthly data.1/

1/A similar effort is included by P. A. Tinsley [11]. Tinsley considers a variety of issues. In the section of greatest relevance to our project, he develops a scheme for predicting quarterly residuals by using the residuals of an explicit monthly model.

A Single-Equation Model

The source of the predictive gain we hope to achieve is easily seen by examining a simple model. Consider the following model:

(1) $y_t = bx_t + e_t$

(2) $e_t = re_{t-1} + u_t$

(3) $u_t = dm_t + v_t$

where y is the endogenous variable whose value at time t we wish to forecast, x is an exogenous variable all of whose past, present, and future values are known exactly, e, u, and v are unobservable error terms, and m is an observed subperiod datum which is related to u_t via Equation (3). We assume that m_t and v_t are jointly normal, serially uncorrelated, independent random variables with zero means and variance s_m and s_v, respectively. We shall further assume that m and v are orthogonal to x and to past y's, i.e.,

(4) $E(m_t|y_{t-1},X) = E(v_t|y_{t-1},X) = 0$

where X stands for the sequence $\{x_{t-\infty}, \ldots, x_t, \ldots, x_{t+\infty}\}$. The distribution of u_t may be derived from the distribution of m and v. It is a serially uncorrelated random variable with mean zero. Further, we know

(5) $E(u_t|y_{t-1},X) = dE(m_t|y_{t-1},X) + E(v_t|y_{t-1},X)$

$= 0$

and

(6) $s_u \equiv E(u_t^2|y_{t-1},X_t) = d^2s_m + s_v.$

Equation (1) is a simplified version of a behavioral equation from a model. Equation (2) is the first-order error structure most commonly assumed. The parameters b and r are usually estimated from time series data. For the prediction problem we consider, we shall assume that they are estimated without error. Equation (3) is the relation between u and the observed (perhaps monthly) quantity m which is not used in standard forecasts.

The standard forecast is formed on the basis of the model stated in terms of y and x as

(7) $$y_t = bx_t + r(y_{t-1}-bx_t) + u_t$$

where (7) is derived by combining (1) and (2). Taking expectations in Equation (7), with respect to y and x, and noting Equation (5), yields the forecasting equation

(8) $$E(y_t|y_{t-1},X) = bx_t + r(y_{t-1}-bx_{t-1}).$$

Under the assumptions we have made, the error in forecasts from Equation (8) is identically u_t. Hence, we know that the expected forecast error (conditional on y_{t-1} and X_t) is zero from Equation (5) and the forecast error variance (conditional on the same set) is s_u from Equation (6).

This forecast is an unbiased and minimum variance forecast conditional on the data set consisting of lagged y and the x process. However, the forecast is not unbiased with respect to the entire set of observable data, i.e.,

(9) $$E(u_t|y_{t-1},X,m_t) = dm_t \neq 0, \text{ in general.}$$

An unbiased and minimum variance forecast conditional on all observable data can be generated from a representation of the model in terms of y, x, and m. Consider

(10) $$y_t = bx_t + r(y_{t-1}-bx_{t-1}) + dm_t + v_t$$

where (10) is derived by combining (1), (2), and (3). Taking expectations in Equation (10), with respect to observable data, and noting Equation (4) and the independence of v and m, yields the prediction equation

(11) $$E(y_t|y_{t-1},X,m_t) = bx_t + r(y_{t-1}-bx_{t-1}) + dm_t.$$

The forecast error of this new procedure will be identically v_t. From Equation (4) we know that the forecast is unbiased and from Equation (6) we know that

(12) $$E(v_t^2|y_{t-1},X,m_t) = s_v \leq s_u$$

with equality only if d or s_m is zero. The forecast from Equation (11) is related very simply to the forecast from Equation (8), i.e.,

$$E(y_t|y_{t-1},X,m_t) = E(y_t|y_{t-1},X) + dm_t.$$

The forecasts of Equation (8) could be transformed to optimal forecasts by the addition of dm_t as a "constant adjustment" in each period. Our proposal is basically to test the hypothesis that d = 0 in Equation (3), and to make the appropriate "constant adjustment" when we reject the null hypothesis.

Recognizing that in actual applications, the X process will not be perfectly known, and that b, d, and r will be estimated with error does not affect our conclusion of smaller forecast error variance. Furthermore, if m_t is not orthogonal to x_t, Equation (11) can still be assembled, but in a slightly more complicated way. The analog of (11) becomes

$$E(y_t|y_{t-1},X,m_t) = bx_t + r(y_{t-1}-bx_{t-1}) + d(m_t-cx_t) \tag{12}$$

where c and d have been estimated from the regression equations

$$\begin{aligned} &E(m_t|X) = cx_t \\ &E(u_t|y_{t-1},X,m_t) = d(m_t-cx_t) \end{aligned} \tag{13}$$

using estimates of u_t generated by the estimates of b and r.

Equation (1) is only one of a set of simultaneous equations. The forecasts of the model are generated by solving the entire system and no amount of reduction in the residual variance of individual equations can guarantee that the forecast errors of the whole model will be reduced. That model forecasts are usually improved by these techniques might be inferred, however, from their widespread use.

During the empirical work with the MPS model, we encountered a problem with the specification of the structural equation error processes (Equation (2) in our example). Since it is likely that the problem is shared by many other large macroeconometric models, we feel it deserves a separate treatment here.

Serial Correlation Structure

In the simple example considered above, the error process $\{e_t\}$ was known to be first-order. However, in econometric models we do not have certain prior knowledge of such characteristics. If the $\{e_t\}$ are a higher-order process, e.g., if

(14) $e_t = r_1e_{t-1} + r_2e_{t-2} + u_t,$

and r_2 is nonzero, then the $\hat{u}_t$'s estimated from the first-order structure as

(15) $\hat{u}_t = y_t - bx_t - r(y_{t-1} - bx_{t-1})$

will not be serially uncorrelated. Since the prediction of the "shifts" of Equation (1) is essentially the prediction of the $\hat{u}_t$'s in the future, forecast could be improved potentially by incorporating the information in the past $\hat{u}_t$'s. The marginal contribution of the monthly data in our example would be overstated if we predicted the future of the $\{u_t\}$ process without considering the useful own-lagged values.

We test whether $\hat{u}_t$'s derived from the MPS model are serially correlated by fitting a fourth-order autoregression to the series and testing the null hypothesis that the coefficients in that regression are zero.[2/] For those equations for which we reject the null hypothesis, we include lagged $\hat{u}_t$'s in the subsequent regression on monthly data. We then test the null hypothesis d = 0 in the equation

(16) $\hat{u}_t = A(L)\hat{u}_{t-1} + dm_t + v_t$

where A(L) is a third-order polynomial in the lag operator L defined by $L^n z_t = z_{t-n}$. In cases where we reject the null, Equation (16) is used to predict the shift of Equation (1) in the full-model forecast.

[2/] More powerful tests for whiteness are available. A spectral test based on the cumulated periodogram of equation residuals would permit finer discrimination. However, our purpose is mainly illustrative and so we did not consider more complicated alternative tests.

To this point we have only considered the problem of forming single-period forecasts. However, forecast simulations are made for periods of several quarters. In the context of our single structural equation, forecasts are usually made using Equation (8). For our case of a known X process, a forecast of y_t is really an exact function of a forecast of e_t. We may think of the process as

$$(17) \qquad e_t^* = re_{t-1}$$

$$(18) \qquad y_t^* = bx_t + e_t^*$$

where the *'s indicate forecasts as of time t-1. For the forecast quarters beyond the first, the forecast is formed as

$$(19) \qquad e_{t+k}^* = r^k e_t^*$$

$$(20) \qquad y_{t+k}^* = bx_{t+k} + e_{t+k}^*.$$

The key thrust of our method is to replace e_t^* by a better forecast (based on more information), i.e., to replace (17) by

$$(21) \qquad E(e_t | e_{t-1}, m_t) = re_{t-1} + dm_t$$

and then to forecast using

$$(22) \qquad e_{t+k}^* = r^k E(e_t | e_{t-1}, m_t)$$

and Equation (20).

However, in the case where the $\{e_t\}$ process has more than first-order serial correlation, this scheme is not optimal. There is an inconsistency in forming forecasts of future e_{t+k}'s by simply discounting the initial forecast of e_t when we know that the e_t's do not follow a first-order process. To provide an indication of the possible forecasting gains to be gained by specifying a more complete error structure, we present among our sample simulations one example which uses only lagged quarterly errors and no monthly data. In that example, a fourth-order autoregression

$$u_t = a_1u_{t-1} + a_2u_{t-2} + a_3u_{t-3} + a_4u_{t-4} + v_t \tag{23}$$

has been fit to the $\{\hat{u}_t\}$ process. This autoregression is used to project a sequence of future u_t's using the chain rule of forecasting, i.e.,

$$\begin{aligned} u^*_t &= a_1\hat{u}_{t-1} + a_2\hat{u}_{t-2} + a_3\hat{u}_{t-3} + a_4\hat{u}_{t-4} \\ u^*_{t+1} &= a_1u^*_t + a_2\hat{u}_{t-1} + a_3\hat{u}_{t-2} + a_4\hat{u}_{t-3} \\ u^*_{t+2} &= a_1u^*_{t+1} + a_2u^*_t + a_3\hat{u}_{t-1} + a_4\hat{u}_{t-2} \end{aligned} \tag{24}$$

and so on. The sequence of future e_t's is generated by

$$e^*_{t+k} = re^*_{t+k-1} + u^*_{t+k} \tag{25}$$

and the sequence of y_t's is generated by Equation (20). This procedure is much like the prescription of Granger and Newbold [4].

Empirical Results

The hypothesis tests outlined above were performed on the residuals from selected structural equations of the quarterly MIT-Penn-SSRC (MPS) model for the period 1954.1-1973.4. For illustrative purposes, four different schemes of generating "add-factors" (or in our preceding notation, forecasts of u_t) were used to produce one-, two-, three-, and four-quarter-ahead ex post forecasts for the seven-quarter span 1974.1-1975.3. The mean-square errors (MSEs) for selected variables in those forecasts and in mechanical model forecasts without any adjustments are presented in Tables 1-4.

The experimental period presents a forecasting challenge for any mechanical method, since it includes the period of the OPEC oil embargo and the subsequent downturn sometimes linked to the embargo. Though the test sample is necessarily small, all four add-factor methods demonstrate general reductions in the MSEs of forecasts of real GNP, the GNP deflator, and the unemployment rate, when compared with the unadjusted MPS model. The results of the ex post simulations are summarized in Tables 1-4. The first column of each table contains the MSEs of the pure model forecasts. The second column contains the results of simulations which used lagged residuals to calculate the first-period error and then forecast future errors using the first-order process assumed in the model as described in (21) and (22) above. The fifth column, labeled "Chain Rule Adds" contains the errors generated using the autoregressive structure in the sample-period residuals to forecast all future errors as described in Equations (23)-(25) above. The third and fourth columns of each table contain the errors from forecasting methods which used distributed lags in selected monthly variables as well as lagged residuals to predict future errors. In generating the forecast errors for Column three, labeled "A+B0," only monthly data available at the beginning of the first forecast quarter were used, i.e., no monthly data for the current quarter were included. For Column four, labeled "A+B3," data for all three months of the first forecast quarter were used in the ex post simulation.

In addition to MSEs of the three aggregate variables--GNP, the deflator, and the unemployment rate, the tables list the MSEs for ten endogenous variables divided into two groups labelled A and B. These are the dependent variables from the ten structural equations whose constants were adjusted in this experiment. Group A contains variables from equations in which lagged monthly data helped explain current residuals, but current monthly data did not help significantly; Group B contains variables from equations in which both current and lagged monthly data helped explain the residuals. For each equation, an attempt was made to explain the residual by the use of, at most, two monthly series thought to be closely related to "shifts" in the schedule that equation represented. The monthly data were chosen from among the commonly available monthly series on production, employment, prices, and sales. (The details of the selection strategy and forecasting procedures are included in the appendix.)

While no precise statistical tests were performed, the data do seem to indicate a substantial improvement in forecasting accuracy from the various adjustment methods. The "Lags Only" column indicates that a large share of the forecast improvement is attributable to the

Table 1

Mean Square Errors for
1 - Quarter Ahead Predictions

	Model	Lags Only Adds	A+BO Adds	A+B3 Adds	Chain Rule Adds
Real GNP	116.7	189.0	138.3	201.3	189.8
GNP Deflator	10.7	3.8	3.8	3.9	3.8
Unemployment Rate	0.34	0.30	0.22	0.22	0.30
Group A					
CON	13.7	13.3	8.6	8.3	13.3
KI	42.7	48.8	50.4	47.9	48.8
LMHT	3.6	1.3	0.9	1.1	1.3
LH	0.00004	0.00010	0.00008	0.00007	0.00010
QPXB	0.00055	0.00017	0.00018	0.00017	0.00017
Group B					
ECD	43.5	38.4	34.1	29.8	38.4
YDV$	3.88	3.25	2.67	3.01	3.25
RTB	2.19	4.49	3.68	3.94	4.49
RDP	0.68	0.44	0.37	0.17	0.44
LF+LA	0.52	0.21	0.19	0.03	0.21

*See notes at end of Table 4.

Table 2

Mean Square Errors for
2 - Quarter Ahead Predictions

	Model	Lags Only Adds	A+BO Adds	A+B3 Adds	Chain Rule Adds
Real GNP	678.7	419.2	300.9	412.1	388.2
GNP Deflator	38.1	24.1	24.3	24.8	17.3
Unemployment Rate	0.77	0.59	0.56	0.58	0.53
Group A					
CON	47.6	23.7	21.8	17.7	24.2
KI	89.1	47.2	36.4	47.2	51.2
LMHT	6.2	5.5	7.0	5.5	5.2
LH	0.00022	0.00016	0.00012	0.00016	0.00015
QPXB	0.00161	0.00104	0.00105	0.00106	0.00066
Group B					
ECD	81.2	66.1	53.9	66.1	54.4
YDV$	10.37	12.23	10.28	12.62	14.30
RTB	5.53	7.61	4.45	5.96	9.88
RDP	1.68	1.53	1.02	1.10	2.44
LF+LA	1.16	0.85	0.82	0.67	0.86

*See notes at end of Table 4.

Table 3

Mean Square Errors for
3 - Quarter Ahead Predictions

	Model	Lags Only Adds	A+BO Adds	A+B3 Adds	Chain Rule Adds
Real GNP	2423.5	1584.1	1340.0	1476.4	1729.1
GNP Deflator	77.6	60.8	61.6	61.9	45.7
Unemployment Rate	2.27	1.16	1.01	1.09	1.18
Group A					
CON	139.4	76.7	63.8	65.8	78.1
KI	158.7	127.1	108.6	123.5	136.4
LMHT	7.4	5.2	6.3	5.4	4.6
LH	0.00071	0.00046	0.00041	0.00046	0.00046
QPXB	0.00288	0.00234	0.00233	0.00235	0.00171
Group B					
ECD	163.1	127.8	114.0	126.1	114.6
YDV$	12.34	18.65	16.39	19.59	25.38
RTB	11.42	8.42	5.11	7.88	17.05
RDP	2.62	2.18	1.83	1.94	3.19
LF+LA	1.32	1.28	1.48	1.24	1.25

*See notes at end of Table 4.

Table 4

Mean Square Error for
4 - Quarter Ahead Predictions

	Model	Lags Only Adds	A+BO Adds	A+B3 Adds	Chain Rule Adds
Real GNP	5662.7	4221.8	3853.0	3795.9	4311.5
GNP Deflator	124.7	107.6	108.2	109.0	77.9
Unemployment Rate	6.58	4.06	3.62	3.44	4.33
Group A					
CON	358.3	244.6	208.4	216.3	222.4
KI	211.0	184.1	180.4	167.8	232.7
LMHT	21.4	10.7	9.0	8.7	11.8
LH	0.00163	0.00127	0.00122	0.00122	0.00126
QPXB	0.00410	0.00373	0.00367	0.00372	0.00265
Group B					
ECD	301.0	247.7	228.9	232.3	216.4
YDV$	12.26	16.83	14.92	16.27	25.88
RTB	32.84	29.27	24.46	26.63	53.27
RDP	2.86	2.58	2.43	2.36	2.73
LF+LA	0.89	1.13	1.35	1.23	1.06

Notes for Tables 1-4:

i) Group A variables denote the variables whose prediction equations contain no contemporaneous quarter-monthly data.

ii) Group B variables denote the variables whose prediction equations include contemporaneous quarter-monthly data.

iii) The "Model" column contains the MSEs for the control run of the model without any additive adjustments.

iv) The "Lags Only" column contains the MSEs with the first quarter add predicted from the equations containing only own lagged errors and no monthly data, and future quarter adds are phased out as in Equation 3.

v) The "A+B0" column contains the MSEs with the first quarter add predicted from the equations containing monthly data, but assuming that zero months of the contemporaneous quarter are available, and future adds are phased out as in Equation 3.

vi) The "A+B3" column is like the "A+B0" column, but assumes that all three months of the contemporaneous quarter are available. These first quarter adds are generated by the equations shown in Table I.

vii) The "Lags Only Autoregressive" column contains the MSEs with adds predicted from the "Lags Only" equations in an autoregressive way. The future adds are actually predicted by the equations via the chain rule rather than using the estimated autocorrelation coefficient to phase out the 1-quarter-ahead estimated add.

lagged structural equation errors themselves. This also seems to be evidence that the autocorrelation structure of the model is misspecified. For forecast horizons beyond one-quarter, the lags only procedure produces reductions in the forecast MSE ranging from 14 percent (deflator four-quarters ahead) to 49 percent (unemployment rate three-quarters ahead).

The "A+BO" column shows that monthly data can indeed be helpful in improving the accuracy of quarterly model forecasts. While the percentage reduction in the MSE of the deflator is essentially unchanged, the forecast MSE of real GNP is reduced by a sizeable magnitude from the own-lags experiments. The largest improvement beyond the one-quarter-ahead forecast comes in the two-quarter-ahead forecast where the lags only procedure produces a 38 percent reduction, while the lags together with monthly data produces a 56 percent reduction in the real GNP forecast MSE. The weakest gain in real GNP due to the addition of monthly data is seven percentage points in the four-quarter-ahead forecast.

Interestingly, using data for all three months of the contemporaneous quarter (the A+B3 adds) produces generally larger forecast MSEs than when none of the current-quarter monthly data is used (the A+BO adds). The four-quarter-ahead forecast might be considered an exception, but even there the difference is quite small.

We are puzzled by the performance of the B3 adds. The results seem to imply that forecasters using an econometric model like the FMP model need not spend a lot of time closely monitoring current monthly data. But this conclusion must be modified by at least two major qualifications. The first is that our study relates only to the specific variables described above. Perhaps extending the list of model variables to encompass a greater degree of correlation across structural equations would give greater significance to current-quarter monthly data.

The second qualification involves the fact that our experiments are conditional on knowing the actual exogenous variables of the model. Thus it may be possible that the most important role for monthly data in a forecasting context is to establish values for the exogenous variables. However, we have serious reservations about the practicality of using current-quarter monthly data to help predict the exogenous variables of models like the MPS model. Our casual review of the 136 exogenous variables indicates that not many are likely to be related to monthly data in a simple way. In the absence of a detailed monthly model, autocorrelations are likely to be the best predictors of exogenous variables.

The "Chain Rule" experiment, which uses the "new" serial correlation structure of the disturbances to form all future constant adjustments, improves the price forecasts of the "Lags Only" method but demonstrates little or no improvement in the forecasts of the other variables. This is not an unimportant contribution, however, since prices are typically among the variables with the largest prediction errors. Further experiments which incorporate monthly data as well as higher-order error processes seem to hold considerable promise for increasing forecasting accuracy.

Conclusion

We have presented a method for improving econometric forecasts using subperiod data. We have tested whether a certain body of monthly data helps in forecasting the structural equation residuals from a version of the MPS model. In our example, the monthly data helped and we were able to realize a substantial reduction in system prediction error by using them. We see no reason why similar forecast improvement could not be realized with other econometric models.

Our experience with the MPS model suggests another recommendation that we cautiously generalize. Our tests indicate that in many structural equations there is a higher degree of serial correlation of errors than was assumed in estimation. We found prediction was improved greatly by simply projecting current residuals on past residuals. Therefore, it appears that the MPS model should be reestimated using higher-order error processes. It is likely that such a step would decrease the need for frequent constant adjustments. In the context of this revised model, more powerful tests of the contribution of monthly (and weekly) data could be performed.

References

1. Cooper, Phillip J. and Charles R. Nelson. "The Ex Ante Prediction Performance of the St. Louis and FRB-MIT-PENN Econometric Models and Some Results on Composite Predictors," Journal of Money Credit and Banking, VII, No. 1 (February 1975), pp. 2-32.

2. Cooper, Ronald. "The Predictive Performance of Quarterly Econometric Models of the United States," in Bert G. Hickman ed. Econometric Models of Cyclical Behavior, II, NBER Studies in Income and Wealth, No. 36, 1972, pp. 813-925.

3. Geweke, John. "Temporal Aggregation in the Multivariate Regression Model," Discussion Paper No. 74-50, Center for Economic Research, University of Minnesota, (November 1974).

4. Granger, C. W. J. and P. Newbold. "The Time Series Approach to Econometric Model Building," paper presented to a time-series conference sponsored by the Federal Reserve Bank of Minneapolis (November 13-14, 1975).

5. Green, George R. in association with Maurice Liebenberg and Albert A. Hirsch. "Short- and Long-Term Simulations with the OBE Econometric Model," in Bert G. Hickman, ed. Econometric Models of Cyclical Behavior, I, NBER Studies in Income and Wealth, No. 36, 1972, pp. 25-122.

6. Haitovsky, Yoel, George Treyz, and Vincent Su. Forecasts with Quarterly Econometric Models, NBER Studies in Business Cycles, No. 23, 1974.

7. Kareken, John H., Thomas Muench, and Neil Wallace. "Optimal Open Market Strategy," American Economic Review, LXIII, (March 1973), pp. 156-122.

8. Nelson, Charles R. "The Prediction Performance of the FRB-MIT-PENN Model of the U.S. Economy," American Economic Review, LXII, (December 1972), pp. 902-917.

9. Rolnick, J. Arthur. "Book Review" of J. Phillip Cooper's "Development of the Monetary Sector: Prediction and Policy Analysis in the FRB-MIT-PENN Model," Journal of Political Economy, LXXXIII, No. 2 (April 1975), pp. 438-441.

10. Skoog, Gary. "Systematically Missing Data in Econometric Models: Some Identification Considerations," Unpublished Research Paper (1975).

11. Tinsley, P. A. "On Proximate Exploitation of Intermediate Information in Macroeconomic Forecasting," Unpublished Research Paper, Federal Reserve Board (1974).

Appendix

Selection Strategy for Equations and Variables

Because of the large number of behavioral equations in the FMP model, and because our purpose is descriptive, we will not attempt to deal with all of them. Our primary criterion for selection is that relevant monthly data exist and that at least one month of the quarter is published before the publication of the National Income Accounts data. This criterion is augmented by observation of the residuals of each equation to see which residuals appear large or systematically different from zero. Exhibit A shows the variables from the MPS model examined in this study along with the 1974.1 value of each variable, which indicates the units of measurement. Since the experiments were run before the 1976 revisions in the National Income accounts, the real variables are on the 1958 base. Exhibit B shows the monthly variables examined in this study.

Our hypothesis may be stated as a linear regression of the generic form

$$u_t = f(u_{t-1}, m_t)$$

where u_t is the observed residual, u_{t-1} represents lags of the residuals, and m_t represents all relevant monthly data including distributed lags. For us, a satisfactory equation is one with statistically significant coefficients. The standard R^2 criterion is only a curiosity because a significant constant term is as important as a significant slope coefficient. Note that this form of regression permits an indirect test that the residuals obey the assumed serial correlation structure of the model.

For each residual there are four regressions representing the number of months of data that are available when the forecast is made. Because of the enormous number of tests suggested by this basic hypothesis, we, a priori, set down the following strategy to establish the data period and the final form of the regression hypothesis:

A. Chow test: Because of the accumulating evidence that few macroeconometric structural equations are able to pass a test for structural change, we were interested in subjecting our residual equations to a Chow test. In order to have both some post-sample model residuals in the residuals sample that was used to estimate our hypothesis, and also to have a few post-sample data points to test prediction accuracy, we made 1973.4 the common endpoint of all the primary data sets. This meant that we had post-sample observations for all variables except the labor compensation variable (PL). But we were also constrained by degrees of freedom in all the labor market variables because of our desire to use certain monthly data from the household survey which began only in 1963. Thus, we did not do Chow tests for the labor market variables.

When an equation failed the Chow test, it was estimated over the period 1964.1-1973.4.

The initial hypothesis for all equations was that the residual was a linear function of trend, four own lags, and two monthly variables, each of which had two-quarter (6 months) lags. As sketched above, our view of the structural equation is that all months of the current quarter are important; therefore, the Chow tests were performed only on the regressions which contained all three months' data of the contemporaneous quarter. Thus, the Chow test equations contained 24 variables.

B. Lag structure: Once the sample period was determined via the Chow test or data limitations, the following tests were used to determine which lags were included in the final form of the residual prediction equation. For these tests, the equation with zero months of the contemporaneous quarter was used on the assumption that if lags are not useful without contemporaneous monthly data, they would be even less useful when current monthly data is available. There is also the practical consideration that the irregular flow of monthly data means that the equation containing three months of contemporaneous data will rarely be used, while the zero months equation will almost always be used (by those who make forecasts at least once a month).

Our initial hypothesis about the form of the lag structure may be represented by an equation of the form

$$u_t = g\ (\text{constant},\ t,\ L,\ M^1_{t-1},\ M^1_{t-2},\ M^2_{t-1},\ M^2_{t-2})$$

where L represents u lagged from one through four quarters, and M^i_{t-j} represents all three months of monthly variable i in the t-j quarter. There are fourteen subsets of this set of seven explanatory variables which can be formed by observing the constraints that (1) all subsets contain the constant and t terms and (2) M^i_{t-2} cannot be in the subset without M^i_{t-1}. By performing F-tests between appropriate combinations of these fourteen regressions, we established the lag structure for the final form of the residual prediction equation.

As an example of the computed F-statistics, the following table shows the results for the test of u_t = g (constant, t, L) versus

u_t = h (constant, t):

Variable	Degrees of Freedom	Computed F
KI	4,34	10.75
LH	4,32	141.20
QPXB	4,58	5.40
YDV$	4,58	4.61
RDP	4,34	4.39
CON	4,58	0.60
LMHT	4,32	0.52
ECD	4,34	0.72
RTB	4,34	0.53
LF+LA	4,34	0.61
PL	4,32	0.27

C. Contemporaneous monthlies: In order to test for the significance of current-quarter monthly data, we used the three-month version of all regressions. Taking the lag structure as determined above as given, there are four possible subsets of explanatory variables that can be formed with the lag structure as a group and the two potential monthly variables. By performing F-tests between the relevant combinations of these four regressions, we established the contemporaneous monthly variables which remained in the final form of the residual prediction equation.

Exhibit C shows the three-month version of the final form of the residual prediction equations produced by this testing strategy. In these equations, the notation for the MPS model variables is used to represent the corresponding residual rather than the variable itself.

Since the sample period for all equations ends with the fourth quarter of 1973, the actual sample period is easily inferred from the number of observations. There is no equation shown for variable PL because none of the tests produced a statistically significant variable--just as one would expect from a properly specified structural equation. Similarly, the monthly variables listed in Exhibit B which do not appear in Exhibit C failed all of our significance tests.

Because of the many combinations of forecast comparisons that could be generated with the ten forecast equations, we were forced to combine the variables in an ad hoc way. Group A variables do not contain contemporaneous quarter-monthly data, and therefore, the residual predictions derived from these equations will not change during a given quarter. Group B variables contain contemporaneous quarter-monthly data, and therefore, the predictions derived from these equations will depend on the month of the quarter in which the forecast is made. Again, to restrict the number of simulations, we constructed only Group B adds with zero month's data (call these BO) and three month's data (B3). Using the adds generated by these equations and setting the path according to the autocorrelation scheme described above in equation (22), we generated new forecast paths with the MPS model. The exception to this rule is the "Chain Rule" adds where the future "u's" are predicted directly from the own-lags equations using the chain rule procedure of repeated substitution for difference equations. These results are given in Tables 1-4.

Exhibit A

MPS Quarterly Model Variables

CON: real consumption of goods and services according to the flow of services concept (1974.1 = 538.4).

ECD: real consumer expenditures on durable goods (1974.1 = 105.2).

KI: real stock of nonfarm business inventories, multiplied by 4, at end of period (1974.1 = 781.7).

LF+LA: labor force including armed forces (1974.1 = 92.7).

LH: total hours at annual rate, per employed person in the nonfarm private domestic business and household sector (1974.1 = 1.933).

LMHT: hours of all persons in the private domestic nonfarm business sector including proprietors and unpaid family workers (1974.1 = 142.6).

PL: employee compensation rate in the nonfarm private domestic business sector (1974.1 = 516.6).

QPXB: log of price deflator for nonfarm business product net of federal indirect business taxes less log of PL (1974.1 = -1.245).

RDP: dividend-price ratio for common stocks (1974.1 = 3.7).

RTB: treasury bill rate (1974.1 = 7.6).

YDV$: corporate dividends (1974.1 = 31.6).

Exhibit B

Monthly Variables

AH:	average weekly hours in total private nonfarm sector.
AHE:	average hourly earnings of production workers on private nonagricultural payrolls.
CPI:	consumer price index (in log form in QPXB equation).
DP:	dividend-price ratio for (Standard and Poor's) common stocks.
DRS:	retail sales of durable goods deflated by the CPI.
H:	index of aggregate weekly hours of production workers on private nonagricultural payrolls.
IP:	industrial production index.
LF:	labor force including armed forces.
MD:	demand deposit component of the money supply.
MFI:	inventories of manufacturing sector deflated by industrial component of WPI.
PI:	personal income (deflated in ECD equation by the CPI).
RS:	retail sales deflated by the CPI.
WPI:	wholesale price index (in log form in QPXB equation).
YDV:	dividend component of PI.

Exhibit C

Estimated Prediction Equations

$$QPXB_t = \underset{(-4.4)}{-0.193} - \underset{(-0.7)}{0.00004t} - \underset{(-0.5)}{0.066QPXB_{t-1}} + \underset{(0.5)}{0.062QPXB_{t-2}} - \underset{(-2.7)}{0.277QPXB_{t-3}}$$

$$- \underset{(-0.7)}{0.093QPXB_{t-4}} - \underset{(-2.0)}{0.204CPI_{3t-3}} + \underset{(2.1)}{0.316CPI_{3t-4}} - \underset{(-1.5)}{0.142CPI_{3t-5}}$$

$$- \underset{(-0.7)}{0.094WPI_{3t-3}} + \underset{(1.1)}{0.205WPI_{3t-4}} - \underset{(-0.4)}{0.039WPI_{3t-5}}$$

$\bar{R}^2 = 0.53$ D-W = 2.2 SE = 0.00147 Obs. = 64

$$CON_t = \underset{(2.2)}{9.168} + \underset{(1.6)}{0.071t} - \underset{(-1.1)}{0.617RS_{3t-3}} + \underset{(1.6)}{1.170RS_{3t-4}} - \underset{(-1.9)}{1.035RS_{3t-5}}$$

$\bar{R}^2 = 0.11$ D-W = 2.3 SE = 1.399 Obs. = 64

$$KI_t = \underset{(-1.8)}{-5.175} + \underset{(2.0)}{0.130t} + \underset{(4.1)}{0.757KI_{t-1}} + \underset{(0.3)}{0.072KI_{t-2}} + \underset{(0.2)}{0.049KI_{t-3}}$$

$$- \underset{(-1.8)}{0.380KI_{t-4}}$$

$\bar{R}^2 = 0.58$ D-W = 1.9 SE = 3.268 Obs. = 40

*See notes at end of table.

$$LH_t = \underset{(5.0)}{0.019} - \underset{(-5.2)}{0.00045^t} - \underset{(-3.8)}{0.580LH_{t-1}} - \underset{(-2.0)}{0.304LH_{t-2}} - \underset{(-2.1)}{0.327LH_{t-3}}$$

$$+ \underset{(0.3)}{0.040LH_{t-4}}$$

$\bar{R}^2 = 0.65$ DW = 2.2 SE = 0.00200

$$LMHT_t = \underset{(-2.1)}{-0.860} + \underset{(2.9)}{0.023t}$$

$\bar{R}^2 = 0.16$ DW = 1.6 SE = 0.536 Obs. = 38

$$RDP_t = \underset{(-1.8)}{-0.799} + \underset{(1.3)}{0.00034t} + \underset{(0.9)}{0.210RDP_{t-1}} - \underset{(-0.9)}{0.137RDP_{t-2}} - \underset{(-1.3)}{0.179RDP_{t-3}}$$

$$+ \underset{(0.1)}{0.041RDP_{t-4}} + \underset{(0.7)}{0.119DP_{3t}} + \underset{(1.7)}{0.349DP_{3t-1}} + \underset{(0.4)}{0.146DP_{3t-2}}$$

$$- \underset{(-0.1)}{0.020DP_{3t-3}} - \underset{(-1.8)}{0.450DP_{3t-4}} - \underset{(-0.4)}{0.142DP_{3t-5}} + \underset{(0.5)}{0.223DP_{3t-6}}$$

$$- \underset{(-0.5)}{0.177DP_{3t-7}} + \underset{(1.8)}{0.288DP_{3t-8}} - \underset{(-0.6)}{0.038CPI_{3t-3}} + \underset{(1.6)}{0.138CPI_{3t-4}}$$

$$- \underset{(-0.9)}{0.136CPI_{3t-5}} + \underset{(0.6)}{0.113CPI_{3t-6}} - \underset{(-0.1)}{0.006CPI_{3t-7}} - \underset{(-0.6)}{0.074CPI_{3t-8}}$$

$\bar{R}^2 = 0.82$ DW = 2.2 SE = 0.071 Obs. = 40

$$\begin{aligned}
YDV\$_t = & -0.500 - 0.016t + 0.475YDV\$_{t-1} + 0.202YDV\$_{t-2} \\
& \quad (-1.3) \quad (-1.2) \quad (4.6) \quad (2.4) \\
& + 0.094YDV\$_{t-3} + 0.072YDV\$_{t-4} + 0.319YDV_{3t} + 0.165YDV_{3t-1} \\
& \quad (1.1) \quad (0.8) \quad (5.3) \quad (0.7) \\
& + 0.212YDV_{3t-2} - 0.437YDV_{3t-3} - 0.282YDV_{3t-4} + 0.084YDV_{3t-5} \\
& \quad (0.8) \quad (-6.0) \quad (-1.1) \quad (0.4) \\
& - 0.014PI_{3t} - 0.005PI_{3t-1} + 0.018PI_{3t-2} \\
& \quad (-1.4) \quad (-0.3) \quad (1.4)
\end{aligned}$$

$\bar{R}^2 = 0.77$ D-W = 1.4 SE = 0.206 Obs. = 64

$$\begin{aligned}
ECD_t = & -9.042 + 0.005t + 0.001DRS_{3t} + 0.004DRS_{3t-1} - 0.003DRS_{3t-2} \\
& \quad (-0.9) \quad (0.1) \quad (1.0) \quad (2.6) \quad (-2.0) \\
& - 0.00014PI_{3t} - 0.00005PI_{3t-1} + 0.00005PI_{3t-2} - 0.00007PI_{3t-3} \\
& \quad (-1.7) \quad (-0.4) \quad (0.4) \quad (-0.9) \\
& + 0.00039PI_{3t-4} - 0.00021PI_{3t-5} - 0.00004PI_{3t-6} + 0.00006PI_{3t-7} \\
& \quad (2.9) \quad (-1.9) \quad (-0.5) \quad (0.3) \\
& + 0.000003PI_{3t-8} \\
& \quad (0.1)
\end{aligned}$$

$\bar{R}^2 = 0.46$ D-W = 2.9 SE = 1.357 Obs. = 40

$$
\begin{aligned}
LF{+}LA_t = {} & \underset{(-1.8)}{-9.409} - \underset{(-1.2)}{0.041t} + \underset{(1.1)}{0.00009LF_{3t}} + \underset{(3.8)}{0.00039LF_{3t-1}} + \underset{(3.1)}{0.00024LF_{3t-2}} \\
& - \underset{(-1.9)}{0.00022LF_{3t-3}} - \underset{(-2.0)}{0.00021LF_{3t-4}} - \underset{(-2.2)}{0.00014LF_{3t-5}} + \underset{(0.6)}{0.020IP_{3t}} \\
& - \underset{(-1.7)}{0.058IP_{3t-1}} + \underset{(0.3)}{0.011IP_{3t-2}} + \underset{(0.7)}{0.039IP_{3t-3}} - \underset{(-0.9)}{0.035IP_{3t-4}} \\
& + \underset{(0.3)}{0.006IP_{3t-5}}
\end{aligned}
$$

$\bar{R}^2 = 0.85$ D-W = 2.1 SE = 0.082 Obs. = 40

$$
RTB_t = \underset{(-0.3)}{-0.425} - \underset{(-0.5)}{0.023t} + \underset{(3.7)}{0.610MD_{3t}} - \underset{(-2.3)}{0.653MD_{3t-1}} + \underset{(0.3)}{0.050MD_{3t-2}}
$$

$\bar{R}^2 = 0.29$ D-W = 1.6 SE = 0.552 Obs. = 40

Notes:

i) The subscript "t" denotes the t^{th} quarter so that for monthly data the time subscript 3t denotes the third month of the t^{th} quarter, and so on. Thus, quarterly data are referenced by a subscript "t," and monthly data are referenced by a subscript "3t."

ii) The t-statistics are in parentheses, $\bar{R}^2$ is the coefficient of multiple correlation adjusted for degress of freedom, D-W is the Durbin-Watson statistic, SE is the standard error of the regression whose magnitude may be evaluated relative to the levels given in Exhibit A and Obs. is the number of observations in the regression.