A Short-Run Forecasting Model of Personal Consumption Expenditures

By Dan M. Bechter and Steven P. Zell

The volume and complexity of monthly data on economic activity make their interpretation difficult for both the layman and the professional economist. At the same time, the economist faces the difficulty of making forecasts of the nation's gross national product (GNP) in advance of the quarterly release of these data by the U.S. Department of Commerce. Finding a relationship between the GNP data and the earlier available monthly economic data would help to resolve both problems. This article seeks to establish such a relationship within the context of a formal econometric model.

The model focuses on the personal consumption expenditures (PCE) sector of the GNP accounts, a sector which constitutes almost two-thirds of GNP. In design, the model is a two-part, 12-equation model relating developments in the PCE sector to several monthly series, notably retail sales. One part of the model examines PCE in nominal, or current dollar, terms because the monthly data are only available in that form. The second part of the model develops a price deflator for the personal consumption sector. From forecasts generated separately for nominal PCE and its deflator, forecasts for real PCE, or consumer spending corrected for inflation, were then made. The results of the forecasts indicate that the model successfully tracks real PCE, including most turning points.

STRUCTURE OF THE MODEL

Econometric models may be created for several different purposes. The most common purpose is to try to determine whether a causal relationship exists between some dependent variable, often called the endogenous variable. and one or more independent or exogenous variables. The hypothetical causal relationship is generally suggested by economic theory, but its actual format, including the specific numbers that define the relationship, is statistically estimated by various techniques. Such a causal model is generally referred to as behavioral, and like all models may consist of one or more equations which may or may not be dependent upon one another. An example of a behavioral model in its simplest form is a consumption function like equation 1:

1.
$$C_t = a + b Y_t + e_t$$
.

This equation, a one-equation model, hypothesizes that the amount of consumption in period t, C_t , is a function of disposable personal income, Y_t . Of the remaining terms, e_t is an unknown error term that arises in part

because the equation is a simplification of reality, while a and b are coefficients whose numerical values are estimated by a statistical procedure.

For short-run forecasting of quarterly PCE, an alternative approach to model building is used in this article. Personal consumption expenditures are forecast by using a model with explanatory variables based on monthly data, such as retail sales, that are as closely related to the PCE series as possible. This type of model is not behavioral, in the sense that movements in the monthly series cause movements in the quarterly series under examination. Instead, the monthly data represent very early measures of the same economic phenomena that the PCE series are attempting to measure. In many cases they serve as one of many inputs into the actual calculation of the preliminary PCE data by the Department of Commerce. One such input is retail sales, which consist largely of goods sold to individual consumers. Retail sales are more closely related to the goods portion of PCE than to total PCE, which includes a large services component. Thus, the model was formulated with one equation relating retail sales to the goods component of PCE, and other equations explaining the services component.

Briefly, a two-part, 12-equation model of the personal consumption sector was developed and statistically estimated.² The first eight equations constitute a self-contained model of the personal consumption sector measured in

The equations were specified so that their estimation would reveal the nature of the relationship that existed between the related monthly and quarterly economic series. The model was structured so that, using monthly data, quarterly forecasts of personal consumption expenditures (both nominal and real) could be made shortly after the end of a quarter, but before the GNP data are released. Since total personal consumption expenditures had to be accounted for, the nominal PCE section of the model had to contain at least two equations; one with goods expenditures as the dependent variable and one with services expenditures as the dependent variable.

The total goods equation was specified as follows. Quarterly PCE expenditures on goods was estimated as a function of the previous quarter's goods expenditures and the quarterly change in an adjusted measure of retail sales.³ Data availability was not a problem here. Retail sales data for a month are available approximately 10 days after the end of that month. Thus, a full quarter's retail sales data are available 10 days after the end of the

nominal or current dollar terms. These eight equations are linked by an identity to three equations which estimate the relationship between the monthly consumer and producer price indexes and the quarterly price deflator for personal consumption expenditures. The identity defines real PCE as nominal PCE divided by the deflator. When combined, these 12 equations yield a model which provides forecasts of real, or inflation-adjusted, personal consumption expenditures.

¹ See Dan M. Bechter and Steven P. Zell, "Preliminary Estimates of GNP: 1972-78," Economic Review, Federal Reserve Bank of Kansas City, January 1979, pp. 19-28. See pp. 20-23 for a discussion of the composition of PCE and retail sales.

² The equations of the model, the definitions of the variables, and estimated results are presented in the appendix to this article. In addition, a more detailed discussion of the model's structure and results is presented in a forthcoming Research Working Paper.

³ Retail sales were adjusted to reflect only consumer purchases of goods. From the total monthly level of retail sales are subtracted all expenditures on building supplies and hardware (treated in the GNP accounts as investment) and 15 per cent of expenditures on automobiles and supplies. The latter subtraction is made to eliminate automobiles sold for business use. See Bechter and Zell.

quarter, while GNP data are published about three weeks after the quarter's close. Thus, besides providing a means for interpreting movements in retail sales in terms of movements in quarterly consumer spending on goods, this equation also satisfies the goal of permitting future forecasts of the goods component of PCE in advance of future GNP publication. To continue to meet this goal of timely forecasts in subsequent equations, the date of retail sales availability was chosen as the closing date for the use of any other monthly series needed in the model. For other required series, if third-month data for a quarter were not available by this time, only two months of published data were used and the third month was estimated.

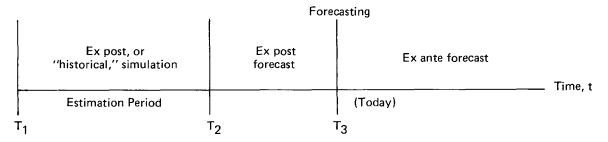
Specifying the equation for personal expenditures on services was considerably more complicated, because of the great difficulty in finding regularly published monthly data to serve as explanatory variables for quarterly PCE on services. Although the best explanatory variable for current quarter expenditures on services was found to be services expenditures in the previous quarter, such a simple trend specification alone means that forecasts from the services equation would be extremely insensitive to economic and other developments since the end of the previous quarter. To help capture the influence of such developments, the difference between current quarter disposable income (after-tax income) and consumer expenditures on goods was added as a second explanatory variable to the services equation. This gave rise to the need to forecast current quarter disposable income, defined as personal income (before tax) multiplied by 1 minus the tax rate. Because of the time constraint on the availability of monthly data noted earlier, however, the estimation of personal income for the current quarter requires prior estimation of its wage and nonwage components for the third month of the quarter. The estimated quarterly

personal income, along with the previous quarter's tax rate, then produces an estimate of current quarter disposable income. Finally, this became—in recursive fashion—an input to the forecast of consumer expenditures on services. The entire nominal PCE section of the model, which estimates both goods and services consumption, consists of five estimated equations and three identities.

To complete the model, a set of four equations was added to the nominal PCE section to forecast the PCE price index. Three of these equations were necessary so that monthly values of the consumer and producer price indexes could be used to forecast the PCE price deflator. Again, this represented a recursive model structure. Because the consumer price index (CPI) for the third month of a quarter is published after the GNP data, the first of these equations is used to estimate the third month of consumer prices. This was accomplished using as independent variables data on earlier months of both consumer prices and producers' prices for consumer finished goods. A second equation, an identity, averages the forecast value for the third month of consumer prices with the two prior months of actual CPI data to yield a quarterly consumer price index. The third equation then uses this quarterly consumer price index to forecast a price deflator for PCE. Finally, real PCE output is generated by another identity that links the two sectors of the model by dividing the simulation output for nominal PCE by that for the deflator.

⁴ A model with a simulation structure of this sort is known as **recursive**. That is, while each equation is estimated separately, the order in which the equations are simulated is important because the forecast results of one equation are used as inputs for the solution of subsequent equations in the model.

Chart 1



SOURCE: Robert S. Pindyck and Daniel L. Rubinfeld, Econometric Models and Economic Forecasts (New York: McGraw-Hill, 1976), p. 313.

In summary, a 12-equation model was specified with five identities and seven estimated equations. Three of the seven estimated equations constitute the central core of the model, with one having as its dependent variable consumer expenditures (PCE) on goods, another consumer expenditures on services, and a third the price deflator for total consumer expenditures. A fourth equation, with disposable personal income as the dependent variable, serves as an input into the services equation. The three remaining estimated equations were necessary because of the problems with monthly data availability. Once estimated and combined with the five required identities, these seven equations provide a recursive structure which, when solved, yields the simulation results that are described in the following sections.

SIMULATION WITH THE MODEL

Once a model is estimated, it may be used to conduct a variety of simulation experiments.⁵ As Chart 1 shows, there are basically three periods over which simulation comparisons may be made. Simulation over the period in which the model was estimated may be referred to as **ex post simulation**, or "historical" simulation. Since over this period, the actual values of the

endogenous (dependent) variables are available, a comparison may be made between these actual values and the simulation results to test whether the model is a valid representation of reality. The second type of simulation is ex post forecasting, in which the model is simulated beyond the estimation period, but no further than the last date for which data are available on all variables. The availability of actual values for the endogenous variables in this ex post forecasting period also permits a comparison with forecast values to test the forecasting accuracy of the model. Finally, ex ante forecasting consists of forecasting beyond "today" into the future. The closest fit of simulated to actual values can be expected to be for the estimation or historical period, followed by the ex post forecasting period, with the poorest fit likely to be for the ex ante forecast period.

⁵ The process of simulation is an econometric procedure. In it, a model consisting of one or more equations, each defining a relationship between a single **dependent** variable and one or more **independent** variables, is used to generate estimates of the values of these **dependent** variables. The simulation of the values of the dependent variables may be done for the historical period over which the model relationship was estimated (the in-sample period), or it can be done for some out-of-sample period, in which case the simulation results are called forecasts.

Using these concepts, the simulation properties of the model were tested in a series of experiments. In the first experiment, the nominal PCE section of the model was estimated using data from the second quarter of 1967 (1967:2) to the fourth quarter of 1972 (1972:4). The model was then simulated over two periods; first over the estimation period 1967:2 to 1972:4, and then over the period 1973:1 to 1978:2, the ex post forecast period.

In the second experiment, the nominal PCE section was also initially estimated from 1967:2 to 1972:4. However, it was then simulated one quarter forward, as it would have been if used, as designed, for making one quarter ahead forecasts of PCE. A repeated process of reestimation and resimulation was then followed. In all, this procedure was repeated 20 times, with the model being reestimated every quarter, using a sample period one quarter longer than in the prior estimation, and then again simulated for only a single quarter forward. The result of this procedure was a data set with 22 simulated values over the period 1973:1 to 1978:2 for each of the endogenous variables. In subsequent comparisons, these results will be referred to as results of the nominal iteration model, while the results of the first experiment will be referred to as those of the single-estimation model. A third experiment, described subsequently, combined the nominal iteration model with the deflator section of the model to provide a simulation of real personal consumption expenditures.

Analysis of Simulation Results

After the simulation results of the various experiments were obtained, they were evaluated as follows. First, the simulated values were compared to the actual values for each of the dependent variables. The errors (i.e., the differences between actual and simulated values) were inspected for both magnitude and

sign patterns. Second, the root-mean-square (RMS) simulation error was calculated over the periods used in the simulation. The RMS error measures the deviation of the simulated variable from its actual time path, and is the most widely used comparison of simulation results. The remainder of the article examines and analyzes the simulation results for total personal consumption expenditures in current dollars and in constant dollars.

Simulation Results: Nominal PCE

The Single-Estimation Model. This model was simulated for two periods, the estimation period 1967:3 to 1972:4, and the ex post forecasting period 1973:1 to 1978:2. A comparison of the simulation output indicates the existence of a negative bias in the latter period. While the simulation errors of this single-estimation model are distributed fairly evenly about zero in the earlier period (mean error of 0.05), they begin to develop a distinctly negative bias in the ex post forecasting period (mean error of -3.38). This worsening of the

⁶ The RMS error for a variable Y_t is defined as:

RMS error =
$$\frac{1}{T_t} \sum_{t=1}^{T} (Y_t^s - Y_t^a)^2$$

where Y^S = the simulated value of Y_t

 Y_{+}^{a} = the actual value

T = the number of periods in the simulation

See Robert S. Pindyck and Daniel L. Rubinfeld, section 10.2, for this and other methods for evaluating simulation models.

⁷ In all simulations presented in this article, lagged dependent variables were constrained to equal actual historical values rather than prior period simulated values. This procedure yielded a somewhat better simulation fit than that provided by the unconstrained simulated procedure.

model's performance is further illustrated by the behavior of the root-mean-square error, which almost doubles from 3.32 in the historical period to 5.19 in the ex post forecasting period. Furthermore, in the historical period, only about half of the errors (10 out of 22) are negative. In the ex post forecasting period, however, fully 17 out of 22 errors are negative. A systematic, negative error thus appeared in the simulation output of the single-estimation experiment—that is, the model regularly underestimates the actual levels of nominal consumer spending after the end of 1972.

The negative bias of the model in the ex post forecasting period is readily explained. A model used for forecasting will retain its accuracy beyond the estimation period only to the extent that the behavior of the economy does not change significantly between the two periods. For the two periods employed, however, this assumption cannot be made. The oil embargo in late 1973, the quadrupling of oil prices, the subsequent large increases in other energy and petroleum related product prices, the worldwide food shortage in 1973, and the most severe postwar recession all took place after the estimation period. As a result of these factors, price increases—especially those for services accelerated sharply and continued to rise more rapidly in the forecasting period than in the estimation period. The trend rate of current dollar expenditures on both goods and services could thus be expected to rise. In addition, the percentage of disposable income spent by Americans rose above its range observed in the estimation period. As a result, the relationships previously estimated in the single-estimation model over the period 1967:2 to 1972:4 were no longer accurate for the period following 1972. Therefore, using the model to forecast services expenditures in the later period resulted in the underestimation of actual services expenditures in 19 of 22 cases, and total personal consumption expenditures were systematically underestimated.

The Iteration Model. Though the ex post forecasting ability of the single-estimation version of the PCE model has been found lacking, the real test of the model is how well and how consistently it is able to forecast one quarter ahead—the task for which it was designed. To answer these questions, the iteration model was used. The iteration model was reestimated each quarter and simulated only one quarter ahead to forecast nominal consumer spending for the period 1972:4 to 1978:2.

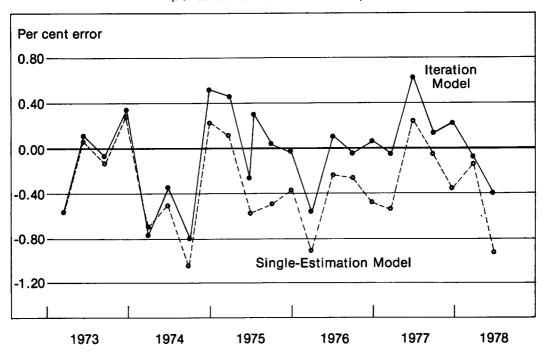
How well does the simulation output from the iteration model compare with that from the single-estimation model? Over the ex post forecasting period, the iteration model is easily the superior of the two models. This is illustrated in Chart 2, which compares the per cent errors of the two simulation outputs. Unlike the single-estimation model, the iteration model shows no particular patterns to its errors, which are both much smaller (RMS of 3.79 versus 5.19) and much more evenly distributed about zero (mean error of -0.41 versus -3.38). Thus, to avoid the simulation inaccuracy introduced by a change in the behavior of the economy between the historical and forecasting periods, the iteration approach is preferable to a single fitting of the model. Reestimating and resimulating the model every quarter appears to be the best method for obtaining good short-run forecasts of nominal personal consumption expenditures.

Simulation Results: Real PCE

Final evaluation of the simulation performance of the personal consumption model requires the generation and analysis of simulation output for real personal consumption expenditures. For this third simulation experiment, both the deflator section and the

Chart 2
FORECAST OUTPUT OF NOMINAL PERSONAL CONSUMPTION EXPENDITURES
1973:1 TO 1978:2

(Per Cent Error From Actual Values)



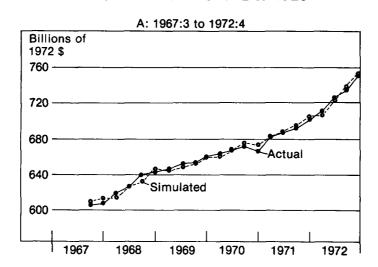
nominal PCE section were estimated over the period 1967:2 to 1978:1, corresponding to the last iteration of the iteration model. This approach to the one-quarter-ahead forecasting of real PCE follows directly from the results of the first two experiments, namely, that a better ex ante forecast of nominal PCE results from the use of the iteration model reestimated each quarter, than from the use of the single-estimation model. Having established that result and therefore accepting the last iteration of the iteration model as the appropriate simulation of nominal PCE, the next step was to estimate and simulate the price deflator section of the model.

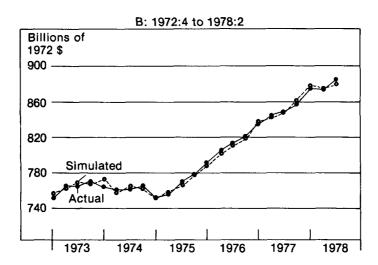
The simulation results from the price equations were extremely good. Over the entire

time period, 1967:3 to 1978:2, the RMS simulation error for the third month of the CPI was only 0.16. The per cent errors for each observation were also quite small, ranging from 0.25 per cent to -0.42 per cent, with most errors far smaller. It should be noted that simulation of the first equation of the deflator section alone generates an estimate of the CPI for the last month of each quarter a full month before the results are published. Finally, excellent simulation results were also obtained for the price deflator: the RMS simulation error equaled 0.25, and the range of per cent errors was from 0.48 per cent to -0.39 per cent

How well does the total model predict real consumer spending? To answer this question,

Chart 3
REAL PERSONAL CONSUMPTION EXPENDITURES:
SIMULATED AND ACTUAL VALUES



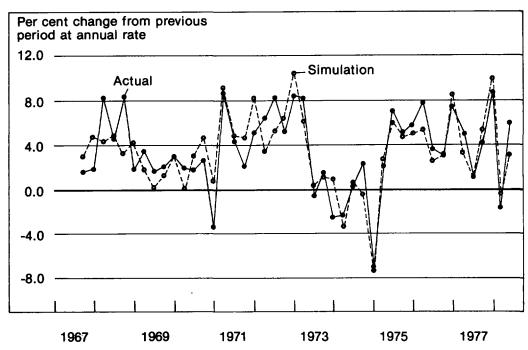


two types of comparisons were made of the tracking behavior of simulated real PCE. The first type, illustrated in Chart 3, compares the **levels** of actual and simulated real PCE over the two periods 1967:3 to 1972:4, and 1972:4 to 1978:2. The second type of comparison,

presented in Chart 4, contrasts the per cent changes from one quarter to the next of the simulated and actual real series. In that chart, each point on the actual per cent change line is calculated as the annual growth rate of one quarter's actual value from the previous

Chart 4
REAL PERSONAL CONSUMPTION EXPENDITURES:
PER CENT CHANGE FROM PREVIOUS QUARTER
1967:3 TO 1978:2

(Annual Rate)



quarter's actual value. The per cent changes on the simulated line, however, are calculated somewhat differently. Since the model was designed to forecast only one quarter forward, actual values of the GNP series for the previous quarter are already known. Thus, rather than calculating the per cent change between two approximate simulated values, the previous quarter's known actual value was used as the initial point for each simulated per cent change calculation in Chart 4.8

Both of the charts allow study of economic turning points. While both upturns and downturns are turning points, the latter have the greater economic significance since economic series such as real PCE normally grow over time. A downturn means that real

expenditures by consumers actually fell following the previous quarter's increase. As the actual series in Chart 3 shows, five such occurrences have taken place in the period under study; in 1970:4, 1973:4, 1974:4, and 1978:1. These downturns can also be seen with the actual series in Chart 4 as points which fall

⁸ This approach would be used by the researcher to forecast quarterly per cent changes each quarter. It yields a closer fit to the actual per cent changes than could be obtained by calculating per cent changes using two quarters of simulated values alone.

⁹ Users of this model would assess whether a turning point had been reached by comparing the most recent quarter's simulated value with the previous quarter's actual value rather than comparing the two quarters' simulated values. Thus, Chart 4 will be used to assess the success of the model in tracking turning points.

below the zero per cent change line, directly after a previous point above the line.

As to overall performance of the model, Chart 3 clearly shows that the simulated levels of real PCE generally track the actual levels very closely. As Chart 4 shows, however, the model's performance in catching downturns is somewhat mixed. While the downturn in 1978:1 was correctly forecast, that in 1974:4 was predicted one quarter early, that in 1973:4 was predicted one quarter late, and those in 1973:2 and 1970:4 were missed altogether. The picture, however, is better than it looks. Though the downturn in 1970:4 was sizable. the downturn in 1973:2 was almost insignificantly different from zero. Furthermore, the level of real PCE predicted for that date was almost exactly equal to its true value. In addition, the simulation model correctly predicted sharp declines in the rate of growth of real PCE from the previous quarter (Chart 4). Finally, the recoveries in the rate of growth of real PCE that followed each of these downturns were also correctly predicted by the model. The model also did quite well in predicting other sharp changes in the rate of growth of PCE, especially in the period following 1973 (Chart 4). For the period before 1974, the model tended to first undershoot and then to overshoot the true per cent changes. However, it appears that the model has stabilized in recent years, and is thus providing very good forecasts of real PCE behavior.

SUMMARY AND CONCLUSIONS

The research described here grew out of two related concerns among economic analysts: how to systematically interpret the vast amount of economic data generated monthly, and how to accurately forecast GNP one quarter ahead using the most up-to-date data available. The solutions to both problems appeared related. If a model could be developed to estimate the

relationship between various monthly series and GNP components, the model could also be used to forecast GNP.

These issues were analyzed through use of a small model of the personal consumption expenditures (PCE) sector, which constitutes nearly two-thirds of GNP. Equations were developed and estimated relating the GNP series of nominal PCE on goods and services, disposable income, and the PCE price deflator to related monthly series such as retail sales, personal income, and other price indexes. Where possible, the equations attempted to measure the GNP variables by as close a monthly approximation as could be specified. Thus, the model differs from the more usual behavioral economic model, in which causal relationships are assumed to exist between the dependent and independent variables.

Using this article's approach, extremely good estimation fits were obtained between the monthly data and their related GNP series. The conclusion of the estimation section of this article is that, despite some data shortcomings in both timeliness and availability, very good models can be constructed using monthly data to estimate personal consumption expenditures and its deflator.

When the model was used to forecast PCE developments it was found that, when reestimated each quarter as new data become available, the current dollar section of the model provided excellent simulation results of nominal personal consumption expenditures. Furthermore, good results were also obtained by the equations used to forecast the PCE price deflator. Combined together, the total set of equations was then used to obtain forecasts for real personal consumption expenditures. The resulting output successfully tracked the actual value of real PCE, catching most turning points, and equally important, reflected most of the sharp changes in the series, especially in recent years.

APPENDIX

PERSONAL CONSUMPTION SECTOR MODEL WITH ESTIMATION RESULTS FOR THE FINAL ITERATION (Estimation Range 1967:2 to 1978:1, 44 Observations) EQUATIONS

- A. **Definitions** (by equation number)
 - 1. CONSUMQ ≡ TOTGOODQ + TOTSERVQ
 - 2. PIQ = (PI1 + PI2 + PI3)/3
 - 3. PI3 ≡ WAGEPI3 + NWAGEPI3
 - 9. CPIQ \equiv (CPI1 + CPI2 + CPI3)/3
 - 12. REALPCE ≡ CONSUMQ/GDC
- B. Estimated Equations (t statistics in parenthesis)
 - 4. WAGEPI3 = $1.00885 * WAGEPI3(-1) + 0.59169 * WAGEPI3(-1) * <math>\% \Delta PRIVWAGE3$ (1010.14) (15.3191)

Estimation method: GLS AUTO1 RHO1 = -0.2050, RSQ = 0.99987, DW(\emptyset) = 1.96

Estimation method: OLS, RSQ = 0.99804, $DW(\emptyset)$ = 1.94

Estimation method: GLS AUTO1 RHO1 = -0.31, RSQ = 0.994, DW(\emptyset) = 2.07

7. TOTGOODQ =
$$-1.5801 + 0.8280 * \Delta SALESQ + 1.00715 * TOTGOODQ(-1)$$

(-1.04963)(12.0743) (279.495)

Estimation method: OLS, RSQ = 0.99957, DW(\emptyset) = 2.24

8.
$$TOTSERVQ = -3.67877 + 0.00530 * (DISPQ - TOTGOODQ) + 1.02968 * TOTSERVQ(-1) (-5.41279) (0.31986) (51.5456)$$

Estimation method: GLS AUTO1 RHO1 = -0.3445, RSQ = 0.99989, DW(\emptyset) = 1.85

Estimation method: OLS, RSQ = 0.99997, DW(\emptyset) = 2.06

11. GDC =
$$0.12803 * GDC(-1) + 0.87284 * GDC(-1) * CPIQ/CPIQ(-1)$$

(2.00911) (13.9112)

Estimation method: GLS AUTO1 RHO1 = 0.2627, RSO = 0.99978, $DW(\emptyset)$ = 1.86

(-2.50993)

VARIABLES

Endogenous:

WAGEPI3: Wage personal income for the third month of a quarter.

NWAGEPI3: Nonwage personal income for the third month of a quarter.

DISPIQ: Quarterly disposable personal income (income after taxes).

TOTGOODQ: Quarterly total personal expenditures on goods.

TOTSERVQ: Quarterly total personal expenditures on services.

CPI3: Consumer price index for the third month of a quarter.

GDC: Personal consumption price deflator.

Definition:

PIQ: Quarterly personal income.

PI3: Personal income for the third month of a quarter.

CONSUMQ: Quarterly total personal consumption expenditures on goods

and services.

CPIO: Quarterly average consumer price index.

REALPCE: Real quarterly total personal consumption expenditures.

Exogenous:

WAGEP12, WAGEP11: Wage personal income for the second and first months of a quarter, respectively.

NWAGEPI2, NWAGEPI1: Nonwage personal income for the second and first months of a quarter, respectively.

SALESQ: The sum of three months of retail sales, calculated at an annual rate, and adjusted to reflect only sales of goods to individuals.

(See text.)

PI2, PI1: Personal income for the second and first months of a quarter.

PRIVWAGE3: Total nonsupervisory wages earned in the private nonagricultural workforce in the third month of every quarter.

Calculated as the product of seasonally adjusted hours and earnings from the Labor Department's monthly establishment survey.

TXRATE: The "effective" tax rate, defined as 1 minus the ratio of disposable income (DISPIQ) to total personal income (PIQ), with both numerator and denominator lagged one quarter. (See text.)

CPI1, CPI2: Consumer price indexes for the first and second months of a quarter, respectively.

WPI1, WPI2, WPI3: Consumer finished goods price index from wholesale price index for the first, second, and third months of a quarter.