2 Econometric Considerations

2.1 Introduction

Three of the major features that are likely to characterize an aggregate quarterly time series model such as the one developed in this study are the simultaneous determination of the endogenous variables, the inclusion of lagged endogenous variables among the predetermined variables, and serial correlation of the error terms in the equations. Serial correlation is likely in the present model because of the relatively simple specification of the equations. Many relevant variables have doubtlessly been excluded from the equations, and if these excluded variables are autocorrelated, it is likely that the error terms in the equations will be autocorrelated as well.

With respect to the estimation of dynamic multiperiod forecasting models, Klein in his paper on economic prediction points out that there is a contradiction between the assumptions of traditional estimation theory and those of prediction theory.[1] In estimation theory lagged endogenous variables are treated as though they were predetermined, whereas for multiperiod forecasts lagged endogenous variables can be considered to be predetermined only for the first period forecasts, since after the first period the values of the lagged endogenous variables must be generated within the model. Klein suggests that these models might be estimated in a nontraditional manner by minimizing the sum of the multiperiod forecast errors. An alternative suggestion would be to minimize a weighted average of these errors if more distant forecast errors were of less concern to the investigator than the more recent ones. Klein further points out, however, the properties of these types of techniques are not well understood, and the techniques are not easy to use.[2]

Because of these difficulties, no attempt was made in this study to use the less-traditional techniques. The technique that has been used is based on the assumptions of traditional estimation theory. The technique is described in Fair [17]. It is designed for the estimation of simultaneous equation models with lagged endogenous variables and first order serially correlated errors—the properties that are assumed to characterize the money GNP sector in the present model. The technique yields consistent parameter estimates and

[1] Klein [33], p. 56.
[2] Ibid., p. 65.

allows an estimate of the serial correlation coefficient of the error terms to be made for each equation of the model.

In this chapter the statistical properties that are assumed to characterize the money GNP sector of the model will be made more explicit and the technique that has been used to estimate the sector will be briefly described. The data that have been used in the money GNP sector will then be discussed and the notation that has been used for each of the variables will be presented. The chapter concludes with a discussion of the periods of estimation used. Some of the discussion in the next section follows closely the discussion in Fair [17], Sections 2 and 3.

2.2 The Technique Used to Estimate the Money GNP Sector[3]

The money GNP sector of the model can be written in matrix notation as follows:

$$AY + BX = U, \tag{2.1}$$

where

$$U = RU_{-1} + E. \tag{2.2}$$

Y is an $h \times T$ matrix of the endogenous variables of the sector; X is a $k \times T$ matrix of the predetermined (i.e., both exogenous and lagged endogenous) variables included in the sector; U and E are $h \times T$ matrices of error terms; and A, B, and R are $h \times h$, $h \times k$, and $h \times h$ coefficient matrices respectively. T is the number of observations, h is the number of endogenous variables in the sector, and k is the number of predetermined variables. The subscript -1 for U_{-1} denotes the one-quarter-lagged values of the terms of U.

The following basic assumptions about the sector are made. Write E as

$$E = (e(1)\, e(2) \dots e(T)), \tag{2.3}$$

where $e(t) = (e_1(t)\, e_2(t) \dots e_h(t))'$ is an $h \times 1$ vector of the tth value of the error terms. It is assumed that

$$\mathcal{E}(E) = 0 \tag{2.4}$$

$$\mathcal{E}e(t)\, e'(t) = \textstyle\sum, \qquad (t = 1, 2, \dots, T), \tag{2.5}$$

$$\mathcal{E}e(t)\, e'(t') = 0, \qquad (t, t' = 1, 2, \dots, T;\ t \neq t'). \tag{2.6}$$

[3] This section is more difficult than the rest of the text, and it can be skipped without too much loss of continuity.

In other words, it is assumed that the error terms in E have zero expected values and are uncorrelated with their own past values and with each other's past values. The contemporaneous error terms, however, can be correlated across equations.

It is further assumed that

$$\text{plim}\ T^{-1}XE' = \text{plim}\ T^{-1}X_{-1}E' = \text{plim}\ T^{-1}Y_{-1}E' = 0, \tag{2.7}$$

i.e., that the error terms in E are uncorrelated with the contemporaneous and one-quarter-lagged values of the predetermined variables and with the one-quarter-lagged values of the endogenous variables. Finally, the inverse of A is assumed to exist, and R is assumed to be a diagonal matrix of elements between minus one and one. The assumption made in equation (2.2) of first order serial correlation and the further assumption that R is a diagonal matrix are not as general as one might hope for, but they are necessary if the following estimation technique is to yield consistent estimates.

The technique used to estimate the money GNP sector is a version of the two-stage least squares technique. It is essentially a combination of the standard two-stage least squares technique for dealing with simultaneous equation bias and the Cochrane–Orcutt iterative technique for dealing with serially correlated errors. Care must be taken when using the technique to be sure that the proper instrumental variables are included in the first stage regression to insure consistent estimates. In the following explanation of the technique, attention will be concentrated on the estimation of the first equation of (2.1).

The first equation of (2.1) will be written as

$$y_1 = -A_1 Y_1 - B_1 X_1 + u_1, \tag{2.8}$$

where

$$u_1 = r_{11} u_{1_{-1}} + e_1. \tag{2.9}$$

y_1 is a $1 \times T$ vector of the values of y_{1t}, the endogenous variable explained by the first equation; Y_1 is an $h_1 \times T$ matrix of the endogenous variables (other than y_1) included in the first equation; X_1 is a $k_1 \times T$ matrix of predetermined variables included in the first equation; u_1 and e_1 are $1 \times T$ vectors of error terms; r_{11} is the element in the first row and first column of R; and A_1 and B_1 are $1 \times h_1$ and $1 \times k_1$ vectors of coefficients corresponding to the relevant elements of A and B respectively.

Two further equations will be useful in the following analysis. First, from (2.1) and (2.2) the reduced form for Y, expressed in terms of the error E only, is

$$Y = -A^{-1}BX + A^{-1}RAY_{-1} + A^{-1}RBX_{-1} + A^{-1}E. \tag{2.10}$$

Secondly, equations (2.8) and (2.9) imply for any value of r:

$$y_1 - ry_{1_{-1}} = -A_1(Y_1 - rY_{1_{-1}}) - B_1(X_1 - rX_{1_{-1}}) + [(r_{11} - r)u_{1_{-1}} + e_1]. \tag{2.11}$$

In equation (2.11), e_1 is correlated with Y_1, and $u_{1_{-1}}$ is correlated with $Y_{1_{-1}}$ and with the lagged endogenous variables in X_1 and $X_{1_{-1}}$. The equation can be consistently estimated, however, by the following procedure:

First stage regression: Choose a set of instrumental variables that are uncorrelated with e_1 and that at least include $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$. Regress each row of Y_1 on this set and calculate the predicted values of Y_1 (denoted as $\hat{Y}_1$) from these regressions.

Second stage regression: For a given r, estimate equation (2.11) by ordinary least squares, using $\hat{Y}_1 - rY_{1_{-1}}$ in place of $Y_1 - rY_{1_{-1}}$, and calculate the sum of squared residuals of the regression.

Scanning or iterative procedure: Repeat the second stage regression for various values of r between minus one and one (or use an iterative procedure) and choose that r and the corresponding estimates of A_1 and B_1 that yield the smallest sum of squared residuals of the second stage regression. An iterative procedure that can be used here—and which was in fact used in this study—is the following. From initial estimates of A_1 and B_1 (say, $A_1^{(0)}$ and $B_1^{(0)}$), calculate

$$r^{(1)} = \frac{(y_{1_{-1}} + A_1^{(0)}Y_{1_{-1}} + B_1^{(0)}X_{1_{-1}})(y_1 + A_1^{(0)}Y_1 + B_1^{(0)}X_1)'}{(y_{1_{-1}} + A_1^{(0)}Y_{1_{-1}} + B_1^{(0)}X_{1_{-1}})(y_{1_{-1}} + A_1^{(0)}Y_{1_{-1}} + B_1^{(0)}X_{1_{-1}})'};$$

use this value of $r^{(1)}$ to compute new estimates, $A_1^{(1)}$ and $B_1^{(1)}$, of A_1 and B_1 from the second stage regression; use these estimates to compute $r^{(2)}$; and so on until two successive estimates of r are within a prescribed tolerance level. (This is essentially the standard Cochrane–Orcutt [5] iterative procedure adjusted to take into account simultaneous equation bias.) The tolerance level used in this study was .005, and for almost all of the equations that were estimated, the technique converged in less than five iterations.

Consistency of the above estimating procedure can be seen heuristically as follows. Let $\hat{V}_1 = Y_1 - \hat{Y}_1$. Then the equation estimated in the second stage regression is

$$\begin{aligned} y_1 - ry_{1_{-1}} = {} & -A_1(\hat{Y}_1 - rY_{1_{-1}}) - B_1(X_1 - rX_{1_{-1}}) \\ & + [(r_{11} - r)u_{1_{-1}} + e_1 - A_1\hat{V}_1]. \end{aligned} \tag{2.12}$$

Setting r equal to r_{11}, equation (2.12) can then be written:

$$y_1 = r_{11}y_{1_{-1}} - A_1\hat{Y}_1 + r_{11}A_1Y_{1_{-1}} - B_1X_1 + r_{11}B_1X_{1_{-1}} + (e_1 - A_1\hat{V}_1). \tag{2.13}$$

The general estimation method outlined above consists in choosing estimates of r_{11}, A_1, and B_1 (say, $\hat{r}_{11}$, $\hat{A}_1$, and $\hat{B}_1$) such that the sum of squared residuals in (2.13) is at a minimum. The case where r_{11} is assumed to be zero corresponds to the ordinary two-stage least squares method. The error term $e_1 - A_1\hat{V}_1$ in (2.13) has zero expected value ($\hat{V}_1$ has zero mean by the property of least squares) and is not correlated with $y_{1_{-1}}$, $\hat{Y}_1$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ ($\hat{V}_1$ is orthogonal to these variables by the property of least squares, since $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ are used as instruments in the first stage regression). Equation (2.13) can thus be considered to be a nonlinear equation with an additive error term whose properties are sufficient for insuring consistent estimates by minimizing the sum of squared residuals.

It is now clear why $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ have to be used as instruments in the first stage regression in order to insure consistent estimates. The error term $\hat{V}_1$ in (2.13) must be uncorrelated with $\hat{Y}_1$, $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ in order to insure consistent estimates, and using $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ as instruments in the first stage regression insures that $\hat{V}_1$ will be uncorrelated (in fact, orthogonal) with these variables. Otherwise, there is no guarantee that $\hat{V}_1$ will be uncorrelated with the right-hand-side variables in (2.13).

With respect to the iterative procedure described above, minimizing the sum of squared residuals of (2.13) with respect to r_{11}, A_1, and B_1 yields the following equation for $\hat{r}_{11}$:

$$\hat{r}_{11} = \frac{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_1 + \hat{A}_1\hat{Y}_1 + \hat{B}_1X_1)'}{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})'}.$$

Since $\hat{Y}_1 = Y_1 - \hat{V}_1$ and since $\hat{V}_1$ is orthogonal to $y_{1_{-1}}$, $Y_{1_{-1}}$, and $X_{1_{-1}}$, this equation can be written:

$$\hat{r}_{11} = \frac{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_1 + \hat{A}_1Y_1 + \hat{B}_1X_1)'}{(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})(y_{1_{-1}} + \hat{A}_1Y_{1_{-1}} + \hat{B}_1X_{1_{-1}})'},$$

which is the formula used to calculate successive values of r in the iterative procedure.

The choice of instruments to be used in the first stage regression is discussed in Fair [17]. It has already been seen that $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ must be used as instruments to insure consistent estimates. In a method proposed by Sargan [39], all of the predetermined and lagged variables in the model are used as instruments (i.e., all of the variables in X, X_{-1}, and Y_{-1}).[4] From (2.10) it is seen that these are all of the variables that enter the reduced form for Y_1. In general, some lagged endogenous variables are included in both Y_{-1} and X, but they are obviously counted only once as instruments. The disadvantage of Sargan's method for even moderately sized models is the large number of instrumental variables that are used, and the question of how the number of instrumental variables can be decreased with zero or perhaps small loss of asymptotic efficiency is discussed in [17]. The small sample properties of the estimators are also briefly discussed in [17], and the asymptotic covariance matrices are presented.

For the money GNP sector in this study the need to decrease the number of instrumental variables from that proposed by Sargan is not as critical as it would be for larger models, since the number of variables in the present model is not large. Nevertheless, not all of the instrumental variables proposed by Sargan were used in estimating the equations of the sector. One of the exogenous variables in the sector, for example, is the sum of government spending and farm housing investment (denoted below as G_t), and this variable is included in the GNP identity. Sargan's choice of instrumental variables would thus suggest that G_{t-1} should be used as an instrument. Since the identity has a zero error term, however, and since lagged GNP is not among the predetermined variables of the sector, G_{t-1} does not enter the reduced form equation (2.10) and so does not need to be used as an instrument.

Sargan's method also suggests that all of the lagged endogenous variables should be included as instruments. In this study, however, those lagged endogenous variables that were not among the predetermined variables (i.e., those that were not in the X matrix above) were not used as instruments except in those equations in which it was necessary to do so to insure consistent estimates. (As mentioned above, when estimating an equation like (2.8), the variables $y_{1_{-1}}$, $Y_{1_{-1}}$, X_1, and $X_{1_{-1}}$ must be included as instruments to insure consistent estimates, and this was always done in the study.) The analysis in [17] indicated that using a large number of instruments is likely to increase the small sample bias of the estimates, and thus an attempt was made in this study to avoid the use of instruments that did not have to be included to insure consistent estimates and that on theoretical grounds were not considered to be too important in the explanation of the endogenous variables. The instrumental variables that were used for each of the equation estimates in the following chapters are listed in brackets underneath each

[4] See Sargan [39], p. 422.

equation. Also, a "hat" is put over each endogenous variable in the equation (i.e., over each variable for which fitted values rather than actual values were used in the second stage regression). The basic instrumental variables that were used for nearly all of the estimates are discussed in Section 2.3.

In [17] the asymptotic variance-covariance matrix of the above estimator is presented and a suggestion is made as to how the approximate variance-covariance of the estimator may be estimated. With respect to equation (2.8), let $\hat{A}_1$, $\hat{B}_1$, and $\hat{r}_{11}$ denote the estimates of A_1, B_1, and r_{11} respectively. Let $\hat{C}_1$ denote the $1 \times (h_1 + k_1)$ vector $(\hat{A}_1 \quad \hat{B}_1)$. Then the suggestion made in [17] is that the approximate variance-covariance matrix of $\hat{C}_1$ be estimated as $\hat{\sigma}_{11}(\hat{Q}_1\hat{Q}_1')^{-1}$, where $\hat{Q}_1$ is the $T \times (h_1 + k_1)$ matrix $(\hat{Y}_1 - r_{11}Y_{1_{-1}} \quad X_1 - r_{11}X_{1_{-1}})$, where

$$\hat{\sigma}_{11} = T^{-1}\hat{u}_1\hat{u}_1',$$
$$\hat{u}_1 = y_1 - \hat{r}_{11}y_{1_{-1}} + \hat{A}_1(Y_1 - r_{11}Y_{1_{-1}}) + \hat{B}_1(X_1 - \hat{r}_{11}X_{1_{-1}}),$$

and that the approximate variance of $\hat{r}_{11}$ be estimated as $T^{-1}(1 - \hat{r}_{11}^2)$. These are the formulas that have been used to estimate the approximate variances and covariances in this study. The t-statistic of a coefficient estimate is defined in this study to be the ratio of the coefficient estimate to the estimate of its approximate standard error, the approximate standard error being defined as the square root of the approximate variance.

In the discussion below a coefficient estimate will be said to be significant if the absolute value of its t-statistic is greater than two. A variable will be said to be significant if its coefficient estimate is significant. Because the distribution of the t-statistic defined above is not known, no precise statistical statements can be made, and so it is merely *assumed* in the work below that a coefficient is different from zero if the absolute value of the coefficient estimate is more than twice the size of the estimate of its approximate standard error.

For each estimated equation in Chapters 3 through 7, the coefficient estimates and the absolute values of their t-statistics are presented, including the estimate and t-statistic of the serial correlation coefficient. The standard error of the regression and the number of observations are also presented. The standard error has been adjusted for degrees of freedom, i.e., it has been estimated as $\sqrt{[\hat{u}_1\hat{u}_1'/(T-K)]}$, where K has been taken to be the number of coefficients estimated in the individual equation not including the serial correlation coefficient, and where $\hat{u}_1$ is defined above. The multiple correlation coefficient R is dependent on what form the variable on the left hand side is in, and the R-squared that is presented in Chapters 3 through 7 is the R-squared taking the dependent variable in first differenced form. This R-squared is thus a measure of the percent of the variance of the *change* in the dependent variable explained by the estimated equation. It will be denoted as $R\Delta^2$. The $R\Delta^2$s have not been adjusted for degrees of freedom.

For some of the equations that were estimated in Chapters 3 through 7 there were no endogenous variables among the explanatory variables. For these equations there were thus no problems of simultaneous equation bias, and they were estimated by the simple Cochrane–Orcutt iterative technique. Also, the technique described above and the Cochrane–Orcutt iterative technique were used in the estimation of the equations in the employment and labor force sector. A nonlinear technique was used in the estimation of the equation in the price sector, and this technique will be described in Chapter 10. A different technique was also used in the estimation of the monthly housing starts equations, and this technique will be described in Chapter 8.

In order to see how the estimates of the seven expenditure equations in the money GNP sector achieved using the technique described at the beginning of this section compare with the estimates achieved using the simple Cochrane–Orcutt iterative technique, both sets of estimates are presented and discussed in Appendix B. The ordinary least squares estimates of the seven equations are also presented in Appendix B. The results in Appendix B should thus indicate how important it is to account for serial correlation problems relative to accounting for problems of simultaneous equation bias.

2.3 The Data Used for the Money GNP Sector

Most of the variables that have been considered in the money GNP sector are listed in Table 2–1. Data for most of the variables are seasonally adjusted at annual rates in billions of current dollars (abbreviated as SAAR in the table). The national income accounts data are based on the July 1969 revisions. The nature of the data for the remaining variables is given in the table. Table 2–1 is meant to be used as a guide for reading Chapters 3 through 7. Each time a variable is introduced for the first time in the following text its symbol is defined, and after that the symbol is used to refer to the variable.

A few adjustments were made in some of the data, and these adjustments will be discussed in the relevant chapters. In Appendix A data on the variables listed in Table 2–1 that are not readily available elsewhere are presented, as well as any adjustments that were made in the data. The variables presented in Appendix A include $MOOD_t$, $PE1_t$, $PE2_t$, $ECAR_t$, $VE1_t$, $VE2_t$, and VH_t.

Table 2–1. List and Description of the Variables Considered in the Money GNP Sector.

Endogenous Variables	
GNP_t	= Gross National Product, NIA, SAAR.
CD_t	= Personal Consumption Expenditures for Durable Goods, NIA, SAAR.
CN_t	= Personal Consumption Expenditures for Nondurable Goods, NIA, SAAR.
CS_t	= Personal Consumption Expenditures for Services, NIA, SAAR.
IP_t	= Nonresidential Fixed Investment (Plant and Equipment Investment), NIA, SAAR.
IH_t	= Nonfarm Residential Fixed Investment (Housing Investment), NIA, SAAR.
$V_t - V_{t-1}$	= Change in Total Business Inventories (Inventory Investment), NIA, SAAR.
IMP_t	= Imports of Goods and Services, NIA, SAAR.
Exogenous Variables Used in the Final Version of the Sector	
G_t	= Government Expenditures plus Farm Residential Fixed Investment, NIA, SAAR.
EX_t	= Exports of Goods and Services, NIA, SAAR.
$MOOD_t$	= Michigan Survey Research Center Index of Consumer Sentiment in Units of 100.
$PE1_t$	= One-Quarter-Ahead Expectation of Plant and Equipment Investment, OBE–SEC data, SAAR.
$PE2_t$	= Two-Quarter-Ahead Expectation of Plant and Equipment Investment, OBE–SEC data, SAAR.
HSQ_t	= Quarterly Nonfarm Housing Starts, Seasonally Adjusted at Quarterly Rates in Thousands of Units.
Other Variables Considered in the Sector	
$ECAR_t$	= Bureau of the Census Index of Expected New Car Purchases, Seasonally Adjusted in Units of 100.
DPI_t	= Personal Consumption Expenditures plus Personal Saving, NIA, SAAR
	= Disposable Personal Income less Interest Paid and Transfer Payments to Foreigners.
$VE1_t$	= One-Quarter-Ahead Expectation of Manufacturing Inventory Investment, OBE data, SAAR.
$VE2_t$	= Two-Quarter-Ahead Expectation of Manufacturing Inventory Investment, OBE data, SAAR.
VH_t	= Percent of Manufacturing Firms Reporting Inventory Condition as High minus the Percent Reporting as Low, OBE data.

Notes: SAAR = Seasonally Adjusted at Annual Rates in Billions of Current Dollars.
NIA = National Income Accounts Data.
OBE = Office of Business Economics, Department of Commerce.
SEC = Securities and Exchange Commission.

2.4 The Periods of Estimation Used for the Money GNP Sector

The basic period of estimation used for the equations of the model was the first quarter of 1956 through the fourth quarter of 1969.[5] As discussed briefly in Chapter 1, there is always a danger in econometric work of this kind that the structure of the economy (and thus quite likely the simple aggregate relationships that are specified in the present model) has changed from one point in time to another. Within an unchanged structure, it is, of course, desirable to have as large a sample as possible in order to achieve the most efficient estimates possible. There is thus to some extent a trade off between the length of the sample period and the confidence that one places on the assumption of an unchanged structure during the sample period. The choice of the basic sample period for the model was made largely on intuitive grounds. It seemed desirable to exclude the Korean War years and give the economy some time to settle down after the war. In addition, it was felt that 1955 may have been an unusual year in some respects, especially in the demand for automobiles. The first quarter of 1956 was thus chosen as the beginning of the sample period. Some of the data from 1955 were actually used, however, since there were lags in the estimated equations.

There were two significant strikes between 1956 and 1969: the steel strike from 15 July 1959, to 7 November 1959, and the automobile strike from 25 September 1964, to 25 November 1964. These strikes clearly had an effect on GNP and its components, and so it was decided to omit these strike observations from the period of estimation. Since one-quarter-lagged values of GNP and other variables were included in many of the estimated equations, observations for one extra quarter for each of the two strike periods were omitted as well. For the steel strike, observations for 593, 594, and 601 were omitted, and for the automobile strike observations for 644, 651, and 652 were omitted. The reason observations for 652 were omitted, even though the automobile strike ended in 644, was the extremely strong reaction of consumers in 651 (due at least in part to the automobile strike of the previous quarter).

There were also two significant dock strikes during the 1956–1969 period—one from 16 June 1965, to 1 September 1965, and one from 20 December 1968, to 3 June 1969—which had a serious effect on imports but little overall effect on GNP. Consequently, observations for 653, 684, 691, 692, and 693 were omitted from the sample period for the import equation, as well as the

[5] In the rest of the text the following notation will be adopted. The first quarter of 1956 will be denoted as 561, the second quarter of 1956 as 562, and so on through the fourth quarter of 1969, 694.

already excluded observations for 593, 594, 601, 644, 651, and 652. For the rest of the equations of the sector, observations for 653, 684, 691, 692, and 693 were not omitted.

Because of data limitations, a shorter period of estimation was used for a few of the equations. The data on housing starts before 1959 are notoriously bad, for example; and so these data were not considered in this study. Data on some of the other variables listed in Table 2–1 were also not available before about 1959. Consequently, for the equations that used these variables the period of estimation was taken to begin in 602 rather than in 561 (with the observations for 644, 651, and 652 continuing to be omitted from the sample period). 602 was chosen as the initial quarter, since observations for 593, 594, and 601 were omitted from the basic period of estimation because of the steel strike. In the following chapters the period of estimation that was used is indicated by the number of observations used: 50 for the basic period of estimation, 45 for the import equation, and 36 when data only after 1959 were available.

The procedure of excluding particular quarters from the sample period because of strikes is a little unusual. The common procedure is to use dummy variables that take on values of one during the strike quarters and zero otherwise. Unless a separate dummy variable is used for each quarter, however, the dummy variable procedure implies that only the constant term in the equation is affected by the strike. Using a separate dummy variable for each strike quarter is equivalent to excluding each strike quarter from the sample period, except that the summary statistics (the R-squared, the standard error of the regression, etc.) are different. Unless one has reason to believe that only constant terms are affected by strikes, the most straightforward approach seems to be to just omit the strike quarters from the sample period; and this was the procedure followed here.

Because of the unavailability of data on housing starts before 1959, the quarterly housing starts variables, HSQ_t, HSQ_{t-1}, and HSQ_{t-2}, which were included among the final predetermined variables of the sector, could not be used as instruments in most of the equations estimated. In practice, these variables were used as instruments only in the particular equations in which they appeared. Using the notation in Table 2–1, the following variables were used as instruments in nearly all of the equations: GNP_{t-1}, CD_{t-1}, CD_{t-2}, CN_{t-1}, CN_{t-2}, CS_{t-1}, CS_{t-2}, V_{t-1}, V_{t-2}, G_t, $MOOD_{t-2}$, $PE2_t$, $PE2_{t-1}$, and the constant term (denoted as 1 in the following chapters).

With respect to the basic set of instrumental variables used, government spending plus exports ($G_t + EX_t$) should be used in place of G_t as one of the basic instruments. This was not done in this study, however, because exports were seriously affected by the dock strikes. Unless the (shorter) sample period used for the import equation was used for all of the equations, $G_t + EX_t$ could not be used as an instrumental variable, and it was felt that it was better to omit EX_t from the sum than to use the shorter sample period for all of the equations.

In the final version of the sector, both $MOOD_{t-1}$ and $MOOD_{t-2}$ are included among the predetermined variables, and so following Sargan's suggestion above, $MOOD_{t-1}$, $MOOD_{t-2}$, and $MOOD_{t-3}$ should have been used as instruments for all of the equations estimated. In order to decrease the number of instrumental variables used, however, only $MOOD_{t-2}$ was included in the basic set of instrumental variables. In order to insure consistent estimates, $MOOD_{t-1}$ and $MOOD_{t-3}$ were, of course, used as instruments in those equations in which $MOOD_{t-1}$ and $MOOD_{t-2}$ appeared as explanatory variables.