This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: A Rational Expectations Approach to Macroeconomics: Testing Policy Ineffectiveness and Efficient-Markets Models

Volume Author/Editor: Frederic S. Mishkin

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-53186-4

Volume URL: http://www.nber.org/books/mish83-1

Publication Date: 1983

Chapter Title: The Econometric Methodology

Chapter Author: Frederic S. Mishkin

Chapter URL: http://www.nber.org/chapters/c10243

Chapter pages in book: (p. 9 - 43)

2 The Econometric Methodology

2.1 The Models

The rational expectations hypothesis asserts that the market's subjective probability distribution of any variable is identical to the objective probability distribution of that variable, conditional on all available past information. Following the literature, we will restrict our attention to linear models and focus only on the first moments of distributions; this yields models which are analytically and empirically more tractable. The rational expectations implication central to this book's analysis is the following: the expectation assessed by the market equals the true conditional expectation using all available past information. For a variable X, this can be written as

$$E_m(X_t|\phi_{t-1}) = E(X_t|\phi_{t-1}), \tag{1}$$

where

ϕ_{t-1} = the set of information available at time $t - 1$,
$E_m(\ldots|\phi_{t-1})$ = the subjective expectation assessed by the market,
$E(\ldots|\phi_{t-1})$ = the objective expectation conditional on ϕ_{t-1}.

The application of rational expectations to financial markets—where it is referred to as market efficiency—shows why the rational expectations hypothesis should be taken seriously in explaining empirical phenomena. Tests of market efficiency usually focus on either holding period returns or prices of securities. For example, let y_t denote the return from holding a particular security from $t - 1$ to t, where the return includes both capital gains and intermediate cash income such as dividends or interest payments. Rational expectations, as in equation (1), then implies the following condition:

$$E[y_t - E_m(y_t|\phi_{t-1})|\phi_{t-1}] = 0. \tag{2}$$

The condition above is too general to be testable. To give it empirical content we must specify a model of market equilibrium that relates $E_m(y_t|\phi_{t-1})$ to some subset of past information, Ω_{t-1},

$$E_m(y_t|\phi_{t-1}) = f(\Omega_{t-1}) = \tilde{y}_t, \tag{3}$$

where Ω_{t-1} is contained in ϕ_{t-1}. For ease of exposition, $f(\Omega_{t-1})$, the representation of the model of market equilibrium, will be denoted by $\tilde{y}_t$. Combining equations (2) and (3) yields the efficient markets condition

$$E(y_t - \tilde{y}_t|\phi_{t-1}) = 0. \tag{4}$$

This condition implies that $y_t - \tilde{y}_t$ should be uncorrelated with any past available information. When $\tilde{y}$, the equilibrium return (or, in loose parlance, a "normal" return), is viewed as determined by such factors as risk and the covariance of y_t with the overall market return (see Fama 1976*a*), the above condition can be stated in a slightly different way. Market efficiency, or equivalently rational expectations, implies that no unexploited profit opportunities will exist in securities markets: at today's price, market participants cannot expect to earn a higher than normal return by investing in that security.

The condition in (4) is analogous to an arbitrage condition. Arbitrageurs who are willing to speculate may perceive unexploited profit opportunities and will purchase or sell securities until the price is driven to the point where this condition holds approximately. An example may clarify the intuition behind this argument. Assume that for a security, $y_t - \tilde{y}_t$, which is sometimes called an "excess" return, is positively correlated with some piece of past information known at time $t - 1$, let us say the company's past earnings. If today the company's past earnings are known to be high, then a higher return than normal for this security is to be expected over the subsequent period. This is a contradiction because an unexploited profit opportunity would now exist. Market efficiency implies that, if this opportunity occurred, the security would be bid up in price until the expected return fell to the normal return. The positive correlation between past earnings and the "excess" return for this security would then disappear.

Several costs involved in speculating could drive a wedge between the left- and right-hand side of (4). Because the collection of information is not costless, arbitrageurs would have to be compensated for that cost and others incurred in their activities, as well as for the risk they bear. (Indeed, as Grossman and Stiglitz (1976) point out, if (4) held exactly, efficient-markets theory would imply a paradox. If all information were fully reflected in a market as eq. [4] specifies, obtaining information would have a zero return. Since there would be no incentive to collect information, it would remain uncollected and unknown. The market would then not reflect this information.) Transaction and storage costs would also result in violations of equation (4). Yet financial securities have the key feature of homogeneity, for they are merely paper claims to income on real assets. Transactions and storage costs will then be small, while compensation of arbitrageurs and the cost of collecting information should not be large relative to the total value of securities traded. Thus deviations from the condition in (4) should not be large.

There are two conclusions to be drawn from the discussion above. First, although the efficient markets or rational expectations condition in (4) may not hold exactly, it is an extremely useful approximation for macroeconomic analysis. Second, this condition should be a useful approximation even if not all market participants have expectations that are rational. Indeed, even if most market participants were irrational, we would still expect the market to be rational as long as some market participants stand ready to eliminate unexploited profit opportunities. It is important to emphasize this point when discussing whether survey forecasts should be used in analyzing market behavior, as Chapter 4 indicates.

A model that satisfies the efficient-markets condition in (4) is

$$y_t = \tilde{y}_t + (X_t - X_t^e)\beta + \epsilon_t, \tag{5}$$

where

ϵ_t = a disturbance with the property $E(\epsilon_t|\phi_{t-1}) = 0$—thus ϵ_t is serially uncorrelated and uncorrelated with X_t^e;
X_t = the vector containing variables relevant to the pricing of the security at time t;
X_t^e = the vector of one-period-ahead rational forecasts of X_t, that is, $X_t^e = E_m(X_t|\phi_{t-1}) = E(X_t|\phi_{t-1})$;
β = vector of coefficients.

That the model above satisfies (4) is easily verified by taking expectations conditional on ϕ_{t-1} of both sides of (5). This yields

$$E(y_t|\phi_{t-1}) = E(\tilde{y}_t|\phi_{t-1}) + E(X_t - X_t^e|\phi_{t-1})\beta + E(\epsilon_t|\phi_{t-1}) = \tilde{y}_t \tag{6}$$

which clearly satisfies (4).

For expositional convenience, we refer to model (5) as "the efficient-markets model." Note, however, that the model embodies not only market efficiency (or, equivalently, rational expectations) but also a model of market equilibrium. This model stresses that only when new information hits the market will y_t differ from $\tilde{y}_t$. This is equivalent to the proposition that only unanticipated changes in X_t can be correlated with $y_t - \tilde{y}_t$.

As the empirical work later in the book demonstrates, the efficient-markets model is useful in attacking such interesting questions as the rationality of interest rate and inflation forecasts in the bond market and the relationship of monetary policy to interest rates. The econometric methodology outlined here is worth studying for this reason alone. Yet it is also worth studying because there are many other applications of the efficient markets model (e.g., Dornbusch 1980; French, Ruback, and Schwert 1981; Frenkel 1981; Hartley 1983; Hoffman and Schlagenhauf 1981*b*; Plosser 1982; Rozeff 1974; Schwert 1977*a*, 1977*b*).

The other model analyzed in the empirical section of this book displays the neutrality property that only unanticipated and not anticipated countercyclical policy will have an effect on business cycle fluctuations. This model displays the policy ineffectiveness proposition of Sargent and Wallace (1975) that a constant money growth rule is not dominated by any rule with feedback. As usually estimated, it has the form

$$y_t = \tilde{y}_t + \sum_{i=0}^{N} \beta_i (X_{t-i} - X^e_{t-i}) + \epsilon_t, \tag{7}$$

where

y_t = unemployment or real output at time t;
$\tilde{y}_t$ = natural or equilibrium level of unemployment or real output at time t;
X_t = an aggregate demand variable, such as money growth, inflation or nominal GNP growth;
X^e_t = anticipated X_t conditional on information available at $t - 1$;
β_i = coefficients;
ϵ_t = error term which might be serially correlated but is assumed to be uncorrelated with the right-hand-side variable.

In the case where the number of lags, N, equals zero and $\tilde{y}_t$ is a distributed lag on past y_t, this is the model estimated by Sargent (1976*a*). The Barro (1977, 1978) model has $N > 0$ and $\tilde{y}_t$ is represented as a time trend or a linear combination of such variables as the minimum wage and a measure of military conscription. Other empirical applications of this model include Barro 1979; Barro and Hercowitz 1980; Barro and Rush 1980; Björkland and Holmlund 1981; Germany and Srivastava 1979; Gordon 1979; Grossman 1979; Hoffman and Schlagenhauf 1981*a*; Leiderman 1979, 1980; Makin 1982; Sheffrin 1979; Small 1979; and Wogin 1980. Following Modigliani (1977), this model will be referred to as the Macro Rational Expectations (MRE) model.

The methodology discussed here is also worth studying for its useful applications in many recent empirical studies which analyze the differential effects of anticipated versus unanticipated movements in explanatory variables. These studies make use of the general model

$$y_t = \tilde{y}_t + \sum_{i=0}^{N} \beta_i (X_{t-i} - X^e_{t-i}) + \sum_{i=0}^{N} \delta_i X^e_{t-i} + \epsilon_t \tag{8}$$

for different definitions of y_t, $\tilde{y}_t$, and X_t. They include Bernanke 1982; Bilson 1980; Bodie 1976; Carr and Darby 1981; Fama and Schwert 1977, 1979; Fischer 1981; Flavin 1981; Jaffee and Mandelkar 1976; Makin 1981; Nelson 1976; Schwert 1981; and Shiller 1980.

2.2 The Methodology

2.2.1 Estimation and Testing

The form of the efficient-markets equation is just a special case of the MRE equation where $N = 0$; hence the discussion here needs to focus only on the estimation and testing methodology for equations (7) and (8). To simplify the exposition, we will limit ourselves to the case where X_t is a single variable. Modifications of the analysis for the case where X_t is a vector of variables is straightforward.

Rational expectations implies that the anticipations of X_t will be formed optimally, using all available information, and, as is usual in the literature, forecasting models are assumed to be linear. A forecasting equation that can be used to generate these anticipations is

(9) $$X_t = Z_{t-1}\gamma + u_t,$$

where

Z_{t-1} = a vector of variables used to forecast X_t which are available at time $t - 1$ (this includes variables known at $t - 1, t - 2, t - 3$, etc.),

γ = a vector of coefficients,

u_t = an error term which is assumed to be uncorrelated with any information available at $t - 1$ (which includes Z_{t-1} or u_{t-1} for all $i \geq 1$, and hence u_t is serially uncorrelated).

An optimal forecast for X_t then simply involves taking expectations of equation (2) conditional on information available at $t - 1$. Hence

(10) $$X_t^e = Z_{t-1}\gamma,$$

and, substituting into equation (7), we have

(11) $$y_t = \tilde{y}_t + \sum_{i=0}^{N} \beta_i(X_{t-i} - Z_{t-1-i}\gamma) + \epsilon_t.$$

Two identification problems occur in the equation (11) model. Some assumption about the correlation of the error term, ϵ, and the right-hand-side variables is necessary in order to identify the β coefficients. The usual assumption—the one that is used in the tests here as well as in previous empirical work on this subject—holds that all the right-hand-side variables are exogenous and are uncorrelated with the error term. This assumption, that (11) is a true reduced form, implies that least-squares estimation methods will yield consistent estimates of the β's.

The other identification problem has been raised by Sargent (1976*b*). If Z_{t-1} includes only lagged values of X_t and there are no restrictions on the lag length N, the MRE model in (11) is observationally equivalent to "an unnatural rate model" where anticipated aggregate demand policy also matters. Hence, in this case, we cannot distinguish between the two competing hypotheses. To see Sargent's point, we can write the forecasting equation where only lagged X's are explanatory variables as,

$$X_t = \gamma(L)X_{t-1} + u_t, \tag{12}$$

where

$$\gamma(L) = \text{polynomial in the lag operator } L = \sum_{i=1}^{\infty} \gamma_i L^i.$$

Taking expectations of (12) conditional on ϕ_{t-1} and substituting into (7) where N is not restricted, we have the MRE model

$$y_t = \tilde{y}_t + \sum_{i=0}^{\infty} \beta_i[X_{t-i} - \gamma(L)X_{t-i-1}] + \epsilon_t, \tag{13}$$

and this can be written as

$$y_t = \tilde{y}_t + \sum_{i=0}^{\infty} \psi_i X_{t-i} + \epsilon_t, \tag{14}$$

where

$$\psi_o = \beta_o,$$
$$\psi_i = \beta_i - \sum_{j=0}^{i-1} \beta_j \gamma_{i-j-1} \text{ for } i \geq 1.$$

If the forecasting equation in (12) is used to derive expectations in equation (8) where anticipated as well as unanticipated aggregate demand matters, we have

$$\begin{aligned} y_t = \tilde{y}_t &+ \sum_{i=0}^{\infty} \beta_i[X_{t-i} - \gamma(L)X_{t-i-1}] \\ &+ \sum_{i=0}^{\infty} \delta_i \gamma(L) X_{t-i-1} + \epsilon_t \end{aligned} \tag{15}$$

which also can be written as (14), where

$$\psi_o = \beta_o$$
$$\psi_i = \beta_i + \sum_{j=0}^{i-1} (\delta_j - \beta_j)\gamma_{i-j-1} \text{ for } i \geq 1.$$

Because both models can be written down as (14), the two models are observationally equivalent: that is, the data cannot discriminate between them because parameters are unidentified.

The problem of observational equivalence has arisen in empirical work on whether anticipated aggregate demand policy matters, in particular, Grossman (1979). Grossman analyzes the MRE model where the aggregate demand variable is nominal GNP growth. His forecasting equation, however, includes only lags of nominal GNP growth as explanatory variables. Because of the resulting observational equivalence problem, Grossman cannot and does not test whether the anticipated nominal GNP growth variables have significant additional explanatory power. Instead, he reports results supporting the MRE hypothesis which rely on flimsy grounds for identification, namely, the assumption that the lag length on nominal GNP growth cuts off at six quarters.

It is possible to discriminate between the two competing models by means of identifying restrictions. These are derived by checking what conditions must be imposed to keep the MRE model and the model in which anticipated aggregate demand matters from being observationally equivalent. This exercise is carried out in Appendix 2.1. The observational equivalence problem is overcome, parameters are identified, and tests of the MRE model are feasible, by either of two conditions particularly important in this book's empirical applications. They are: (1) N is known to be zero, as in the efficient markets model; or (2) Z_{t-1} includes lagged values of at least one other variable besides X which does not enter equation (11) separately from the $\beta_i(X_{t-i} - Z_{t-1-i}\gamma)$ terms.

The method for estimating the MRE model involves joint, nonlinear estimation of the equations (9) and (11) system, which we rewrite as

$$(16) \qquad \begin{aligned} X_t &= Z_{t-1}\gamma + u_t, \\ y_t &= \tilde{y}_t + \sum_{i=0}^{N} (X_{t-i} - Z_{t-1-i}\gamma)\beta_i + \epsilon_t. \end{aligned}$$

System (16) embodies two sets of constraints. Rationality of expectations is imposed since the coefficient γ which appears in the equation for X_t also appears in the equation for y_t. The neutrality property, that anticipated policy is not correlated with $y_t - \tilde{y}_t$, is also imposed because the δ coefficients on X^e_{t-i} are constrained to be zero. Relaxing the neutrality and rationality constraints, the system (16) becomes

$$(17) \qquad \begin{aligned} X_t &= Z_{t-1}\gamma + u_t, \\ y_t &= \tilde{y}_t + \sum_{i=0}^{N} (X_{t-i} - Z_{t-1-i}\gamma^*)\beta_i \\ &\quad + \sum_{i=0}^{N} Z_{t-1-i}\gamma^*\delta_i + \epsilon_t. \end{aligned}$$

A likelihood ratio test comparing both the constrained system (16) and the unconstrained system (17) provides a joint test of both the rationality constraints $\gamma = \gamma^*$ and the neutrality constraints $\delta_i = 0$, conditional on the maintained hypothesis of the model of equilibrium output. Note that (17) can also be written as

(18) $$X_t = Z_{t-1}\gamma^* + u_t,$$
$$y_t = \tilde{y}_t + \sum_{i=0}^{N} \beta_i'(X_{t-i} - Z_{t-1-i}\gamma) + \sum_{i=0}^{N} \delta_i X_{t-i} + \epsilon_t,$$

where $\beta_i' = \beta_i - \delta_i$. This is the form used by Barro (1977) in his tests of neutrality.

As an alternative to relaxing both the neutrality and the rationality constraints, we can relax one set of constraints only. For example, maintaining the hypothesis of rationality but relaxing the assumption of neutrality, system (16) becomes

(19) $$X_t = Z_{t-1}\gamma + u_t,$$
$$y_t = \tilde{y}_t + \sum_{i=0}^{N} (X_{t-i} - Z_{t-1-i}\gamma)\beta_i + \sum_{i=0}^{N} Z_{t-1-i}\gamma\delta_i + \epsilon_t.$$

Under the maintained hypothesis of rational expectations, the null hypothesis of neutrality, that is, $\delta_i = 0$, can be tested by comparing the estimated systems (16) and (19).

Rather than maintain the hypothesis of rationality of expectations and then test for neutrality, one can maintain the hypothesis of neutrality and then test for rationality. The unconstrained system used to perform this test is:

(20) $$X_t = Z_{t-1}\gamma + u_t,$$
$$y_t = \tilde{y}_t + \sum_{i=0}^{N} (X_{t-i} - Z_{t-1-i}\gamma^*)\beta_i + \epsilon_t.$$

A comparison of the estimated systems (16) and (20) provides a test of the null hypothesis of rationality, that is, $\gamma = \gamma^*$, under the maintained hypothesis of neutrality. In the efficient-markets case where $N = 0$, neutrality is a reasonable maintained hypothesis since the absence of neutrality would indicate the presence of unexploited profit opportunities. It must be noted, however, that a rejection of the null hypothesis that $\gamma = \gamma^*$ may result from a breakdown of rationality, neutrality, or the model of market equilibrium. Furthermore, as is demonstrated in Appendix 2.1, when $N = 0$ this test is equivalent to that generated by comparing the systems (16) and (17), which jointly tests $\delta_0 = 0$ and $\gamma = \gamma^*$.

The χ^2 statistic for the joint hypothesis of rationality and neutrality can be partitioned into the contribution from each component hypothesis by relaxing the constraints sequentially. These constraints can be relaxed in two different orders. A priori economic reasoning may suggest an appropriate sequence for relaxing constraints. For example, in testing whether anticipated policy is correlated with output, it seems appropriate first to relax $\delta_i = 0$ and test neutrality under the maintained hypothesis of rationality. Then, without maintaining neutrality, the constraint $\gamma = \gamma^*$ can be relaxed, and rationality can be tested. This is the procedure that is followed in tests of the MRE hypothesis in Chapter 6.

Under the alternative sequence for relaxing constraints, we first relax the constraint $\gamma = \gamma^*$ and test for rationality under the maintained hypothesis of neutrality. The next step in relaxing constraints permits a test of neutrality without maintaining the hypothesis of rational expectations. Yet neutrality has meaning only if we have a theory of expectations such as rational expectations. Realize that the test of neutrality is conducted on the assumption that the expectations of X_t in the second equation of the system (17) are formed with the same information set Z_{t-1} as the time-series model of X_t in the first equation. Yet, if we are not willing to assume that expectations are rational, there seems to be no reason to assume that the same set of variables belongs in Z in both equations in (17). Therefore, it is not clear that this test yields useful information.

One way to generate the likelihood ratio statistics for the above tests is to estimate both the constrained and unconstrained systems with full-information-maximum-likelihood (FIML). Estimation proceeds under the identifying assumption used in previous research on the MRE hypothesis, that the y equation is a true reduced form.[1] This assumption implies that the covariance of the error terms in the two equations of the system is zero. The estimated variance-covariance matrix of the residuals is then

(21)
$$\hat{\Sigma} = \begin{bmatrix} \frac{SSR_X}{n} & 0 \\ 0 & \frac{SSR_y}{n} \end{bmatrix}$$

where

SSR_X = the sum of squared residuals of the X equation,
SSR_y = the sum of squared residuals of the y equation,
n = the number of observations.

The systems analyzed here are triangular, so FIML involves maximizing the concentrated log likelihood function,

(22)
$$\log L = \text{constant} - \frac{n}{2} \log (\det \hat{\Sigma}),$$

1. In the case where $N = 0$, even if the asumption is untrue, its imposition will not invalidate the test statistics (see Chap. 3). However, it is not clear that this desirable result—that the assumption does not matter to the tests of interest here—carries over to the case where $N > 0$.

where det $\hat{\Sigma}$ = determinant of $\hat{\Sigma}$. The resulting likelihood ratio statistic

$$(23) \qquad -2\log\left[\frac{L^c(\hat{\Sigma}^c)}{L^u(\hat{\Sigma}^u)}\right] = n\log(\det \hat{\Sigma}^c/\det \hat{\Sigma}^u)$$

is distributed asymptotically as $\chi^2(q)$, where

q = the number of constraints (Appendix 2.1 discusses how to count them),
L^c = maximized likelihood of the constrained system,
L^u = maximized likelihood of the unconstrained system,
$\hat{\Sigma}^c$ = the resulting estimated $\hat{\Sigma}$ for the constrained system,
$\hat{\Sigma}^u$ = the resulting estimated $\hat{\Sigma}$ for the unconstrained system.

Comparison of this statistic with the critical $\chi^2(q)$ then tests the null hypothesis.

Test procedures used in this book proceed in a slightly different way from that described above, in that they make use of nonlinear least-squares estimation. Estimation is conducted with nonlinear least squares primarily for algorithmic reasons. FIML computer packages are usually not capable of handling large numbers of parameters, and this is required in some of the models analyzed empirically in this book. In addition, FIML packages do not allow us easily to impose the necessary covariance restrictions in (21) or to make a desirable degree of freedom correction, described below, which results in more conservative likelihood ratio statistics. In contrast, the nonlinear least-squares procedure outlined here easily implements the covariance restriction and degrees-of-freedom correction and makes use of a computer package (SAS Institute 1979) that can estimate systems with large numbers of parameters.

The procedure is as follows. Given an initial estimate for the variance-covariance matrix of the residuals, $\hat{\Sigma}$, estimate the system with nonlinear generalized least squares (GLS). (The initial $\hat{\Sigma}$ can be obtained from unconstrained ordinary least-squares estimates of the X and y equations or from previously estimated systems.) Given the particular diagonal form of the $\hat{\Sigma}$ matrix, nonlinear GLS is equivalent to nonlinear weighted least squares (WLS) using the estimates from $\hat{\Sigma}$: that is, the observations for the X forecasting equation are weighted by $\sqrt{SSR_X/SSR_y}$. Appendix 2.2 contains an annotated computer program describing this procedure in great detail. A new $\hat{\Sigma}$ matrix can be estimated using the resulting residuals and the system reestimated again with nonlinear WLS. This iterative procedure is continued until there is little change in the $\hat{\Sigma}$ matrix. Because the system is triangular, this procedure will converge to maximum-likelihood estimates, since theorems showing that iterative three-stage-least-squares is equivalent to FIML then apply to this nonlinear case as well. High computation costs required that iterations were continued only until the estimated variance of all the weighted equations in the system differed by less than 5 percent. Some experimentation indicated that further iterations would have altered the likelihood ratio statistics in the book by at most 1 or 2 percent. This would only lead to a negligible effect on the inference drawn from these statistics.[2]

If the same procedure is followed for estimating the unconstrained system, then the likelihood ratio statistic in (23) is easily calculated and can be used to test the null hypothesis. Although (23) yields valid asymptotic tests, it could be misleading in a small sample like that used here. The problem is that, in the maximum-likelihood calculation of the $\hat{\Sigma}^u$ matrix of the unconstrained system, no correction is made for substantial relative differences in the degrees of freedom in estimates of each unconstrained equation. Thus the finite sample distribution of the test statistic might differ substantially from the asymptotic distribution. For example, in Chapter 6, model 2.1 (see tables 6.1 and 6.2), the unconstrained money growth equation is estimated with 79 degrees of freedom, while the unconstrained output equation is estimated with only 70 degrees of freedom. This is a difference of over 10 percent. The problem is even more severe in the case of model A16.1 (see tables 6.A.16 and 6.A.17) in Appendix 6.3 of Chapter 6: the degrees of freedom for the unconstrained money growth and output equations are now 79 and 32, respectively, a difference of 50 percent. Another way of stating this problem is to say that the weighting matrix for GLS will have a biased estimate of the variance of one equation relative to another. The bias occurs because the estimated variances are the maximum-likelihood estimates (the sum of squared residuals divided by the number of observations in each equation) rather than the unbiased estimates (the sum of squared residuals divided by the degrees of freedom).

The likelihood ratio statistics reported here are corrected for the small sample problem as follows: the constrained system is estimated with the iterative procedure, and the resulting $\hat{\Sigma}^c$ matrix from the constrained system is then again used with nonlinear WLS to estimate the unconstrained system. This corrects the degrees-of-freedom problem because, in the systems where there are cross-equation constraints, the degrees of freedom do not differ across equations. Thus $\hat{\Sigma}^c$ does not suffer from the degrees-of-freedom problem of $\hat{\Sigma}^u$. The resulting likelihood ratio statistic, which is also distributed asymptotically as $\chi^2(q)$ under the null hypothesis, is (Goldfeld and Quandt 1972)

$$(24) \qquad -2\log\left[\frac{L^c(\hat{\Sigma}^c)}{L^u(\hat{\Sigma}^c)}\right] = 2n\log\left(\mathrm{SSR}^c/\mathrm{SSR}^u\right),$$

2. In the empirical analysis in this book, when Goldfeld-Quandt (1965) tests revealed the presence of heteroscedasticity within an equation, the time-trend procedure outlined by Glesjer (1969) was used to weight each observation to eliminate this heteroscedasticity.

where the superscripts on the $\hat{\Sigma}$ indicate that the maximized likelihoods of both the constrained and unconstrained systems were estimated with the same weighting matrix $\hat{\Sigma}^c$ and

SSR^c = the sum of squared residuals from the constrained weighted system,
SSR^u = the sum of squared residuals from the unconstrained weighted system.

Although asymptotically the two test statistics are equivalent, in finite samples the likelihood ratio statistic in (24) is smaller than the alternative in (23) and is more conservative on rejecting the null hypothesis.[3] To see this, realize that $L^u(\hat{\Sigma}^u) \geq L^u(\hat{\Sigma}^c)$, which implies

$$(25) \qquad -2 \log \left[\frac{L^c(\hat{\Sigma}^c)}{L^u(\hat{\Sigma}^c)}\right] \leq -2 \log \left[\frac{L^c(\hat{\Sigma}^c)}{L^u(\hat{\Sigma}^u)}\right].$$

Using (24) rather than (23) will thus give more credibility to rejections if they occur.

One issue concerning estimation remains to be discussed. Since the standard test statistics assume serially uncorrelated error terms, we need to eliminate serial correlation from the residuals. If this is not done, then, as Granger and Newbold (1974) and Plosser and Schwert (1978) have pointed out, we are likely to encounter the spurious regression phenomenon, where significant relationships appear in the data only because there has been no correction for serial correlation. As long as we include lagged dependent variables in the forecasting equation there should be little serial correlation in the u_t residuals and no serial correlation correction will be needed. In the case of the efficient-markets model, theory specifies that $E(\epsilon_t|\phi_{t-1}) = 0$ and hence ϵ_t should be serially uncorrelated. Again no serial correlation correction is needed. However, in the MRE output or unemployment model there is no theoretical argument guaranteeing that the error term is serially uncorrelated. To correct for potential serial correlation and thus avoid the spurious regression problem, the

3. The likelihood ratio statistics here are frequently not appreciably different whether they are calculated using (23) or (24). E.g., in Chapter 6's model 2.1 the likelihood ratio statistic for the joint hypothesis calculated from (23) is 22.81 vs. the value 22.69 reported in table 6.1. In the models found in Appendix 3 (Chapter 6), which use up more degrees of freedom, the difference between statistics calculated from (23) and (24) is more appreciable: e.g., in model A.16.1 the likelihood ratio statistic for the joint hypothesis calculated from (23) is 76.33 vs. the value 66.90 reported in table 6.A.13. Note that the statistic in (24) is essentially the statistic for a Lagrange multiplier test where the percentage change in the sum of squares is approximated by a change in the logs. It is well known that the Lagrange multiplier test is less likely to reject the null hypothesis than a likelihood ratio test, so these results are not surprising. For a further discussion of the Lagrange multiplier test, see Engle (1980).

error term in the MRE output or unemployment models estimated later is assumed to be a fourth-order autoregressive process. This specification for the error term was chosen because fourth-order autoregressions usually eliminate most serial correlation in quarterly, macro time series. Indeed, Durbin-Watson statistics and the residual autocorrelations of the estimated models indicate that this correction for serial correlation is successful in reducing the residuals to white noise.

2.2.2 Specification of the Forecasting Equation

Rational expectations theory implies that X_t^e is an optimal, one-period-ahead forecast, conditional on available information. Thus an appropriate forecasting equation for X_t should rely only on lagged explanatory variables. Economic theory may not be very valuable in generating an accurate model of expectations formation because it is difficult on theoretical grounds to exclude any piece of information available at time $t - 1$ from the Z vector as a useful predictor of a policy variable. Any particular variable may be a useful predictor of X_t even if there is no strong theoretical reason to include it in the Z_{t-1} vector, because the personalities involved in policymaking may be such that they react to this variable nonetheless. For example, if the Board of Governors of the Federal Reserve System were to link monetary policy to the level of unemployment, even though there is no good reason for doing so in a world where the policy ineffectiveness proposition holds, we would still expect to find that the unemployment rate would be highly useful in predicting money growth. This suggests that an atheoretical statistical procedure may be superior to economic theory for deciding on the forecasting equation's specification.

Two procedures are used in this book to specify the forecasting equations. The simplest uses univariate time-series models of the autoregressive type. In the empirical studies later in the book these models are usually subject to unstable coefficients and, more important, should only be used in the efficient-markets model where $N = 0$ because of the observational equivalence problem discussed earlier. Multivariate forecasting models are therefore needed. The Granger (1969) "causality" concept is a natural way to approach the specification of the multivariate models. A variable Z is said to Granger-cause another variable X, if X can be predicted better from past values of Z and X than from past values of X alone. Our forecasting equation for X should definitely include lagged values of X to eliminate any serial correlation in the residuals. If Z Ganger-causes X, then it should be used also in an optimal forecast of X. Hence, as is also argued in Sargent (1981), it belongs as an explanatory variable in the forecasting equation. Note that the issue here is the predictive content of information—which is what Granger-causality is really meant to analyze—and does not involve the tricky concept and issue of economic causality which has led to so much confusion in the literature (see Zellner 1979).

The Granger-criterion for specifying the multivariate forecasting equation is as follows. The X variable is regressed on its own four lagged values (four lags usually ensure white noise residuals in the quarterly data used in this study) as well as on four lagged values of a wide-ranging set of macro variables. The four lagged values of each of these variables are retained in the equation only if they are jointly significant at some marginal significance level (the 5 percent level is one choice). This procedure has the advantage of imposing a discipline on the researcher that prevents his searching for a forecasting equation specification that yields results confirming his prior on the validity of the null hypothesis. Note that a stepwise regression procedure might miss significant explanatory variables because of the order that it chooses to run the regressions. Some judgment must be used in conducting a more general search to find a specification that includes any variables with significant explanatory power.

2.2.3 Specification of the Lag Length, N

The theory of efficient markets indicates that only contemporaneous surprises will be correlated with $y_t - \tilde{y}_t$, and hence $N = 0$. However, the theoretical framework for the MRE model does not specify what the lag length, N, should be. For example, McCallum (1979*a*) argues that if all the state variables are included in the MRE output or unemployment equation, then the theory does imply that $N = 0$. However, since relevant state variables are almost surely excluded from estimated MRE equations, the lag length is not known. In studying the MRE model here, a primary objective is to obtain information on the robustness of results. As discussed in Leamer (1978), experimenting with plausible, less restrictive models is a necessary strategy for verifying robustness of results.

The addition of irrelevant variables to an estimated model only has the disadvantage of a potential decrease in power of the likelihood ratio tests so that we would be less likely to reject the null hypothesis if it were untrue. It will not result in invalid test statistics; that is, the test statistics will have the assumed asymptotic distributions. However, excluding relevant variables will render test statistics invalid. Furthermore, because rejections of the null hypothesis are less likely when the power of a test is reduced by the addition of irrelevant variables, a rejection in this case at a standard significance level is even stronger evidence against the null hypothesis. This is the rationale behind Leamer's (1978) suggestion that when the power of a test decreases—that is, the probability of Type II error increases—then the significance level used to signify rejection should be increased as well. The reasoning above suggests that less restrictive models with longer lags are worth studying, and they are a feature of the later empirical work.

2.2.4 Specification of $\tilde{y}$

Depending on the model studied, many different specifications of $\tilde{y}_t$ may be appropriate. This becomes apparent in the empirical analysis later in the book. Is a correct specification of $\tilde{y}_t$ always a necessary requirement for generating reliable tests of the models described here? This question is particularly important because some specifications of $\tilde{y}_t$ used in the empirical studies in this book are crude, which makes us suspect that they may not be entirely accurate.

The answer to this question is central to an understanding of much of the empirical literature on efficient markets and the policy ineffectiveness proposition. For example, tests of market efficiency have often assumed that $\tilde{y}_t$, the equilibrium nominal return on a security such as a stock or bond, is constant. This is clearly a very crude model of market equilibrium, and we might expect that it will result in a rejection of the efficient-markets model. Yet this often does not occur. Why? The answer is that as long as the variation of $\tilde{y}_t$ is small relative to the variation of $y_t - \tilde{y}_t$, then the specification of $\tilde{y}_t$ will have little impact on tests of the efficient-markets model. The reason in this case is that the correct $\tilde{y}_t$ model will only explain a small percentage of the variation in y_t and thus will have little explanatory power. Then alterations in the $\tilde{y}_t$ specification will make little difference to the fit of the model and hence to its test statistics. This is what we would expect to find in cases where the security is long-lived, such as a long-term bond or common stock, and the holding period is short, say three months. Then the actual return, y_t, has large variation, while any reasonable model of market equilibrium for $\tilde{y}_t$ indicates that it has only small variation. It is exactly in such cases as these where the crude model of the constancy of $\tilde{y}_t$ does not lead to rejections of market efficiency.

The interested reader can find a further discussion of this issue along with clarifying figures in Fama (1976*a*). The point raised here has been made in a different context by Nelson and Schwert (1977) in their comment on Fama (1975). They stress that, if $\tilde{y}_t$ has little variation relative to that of $y_t - \tilde{y}_t$, then tests for the specifications of $\tilde{y}_t$ have little statistical power.

Proponents of equilibrium or natural rate models in which the policy ineffectiveness proposition holds usually emphasize deviations from the natural rate in their explanations of unemployment or output. This emphasis makes sense because they believe that the bulk of the cyclical variation in unemployment or output can be attributed to these deviations. This is exactly the case in which the variation in $\tilde{y}_t$ (removing its trend, if there is one, as in the output case) is small relative to the variation in $y_t - \tilde{y}_t$. Then, as is argued above, tests of the policy ineffectiveness proposition are insensitive to the specification of the model for the natural rate of unemployment or output (as long as the trend is removed).

2.3 A Comparison with Previous Methodology

Previous empirical work has tested the neutrality implications of the MRE hypothesis. How does the methodology of this chapter compare with that used in the work cited earlier?

Barro (1977, 1979), Barro and Rush (1980), and Small (1979), among others, use a two-step procedure. They first estimate a forecasting equation by ordinary least squares (OLS) over the sample period and calculate the residuals, that is,

$$\hat{u}_t = X_t - Z_{t-1}\hat{\gamma} \tag{26}$$

Then the residuals are used as the unanticipated aggregate demand variable in the MRE y equation,

$$y_t = \tilde{y}_t + \sum_{i=0}^{N} \beta_i \hat{u}_{t-i} + \epsilon_t \tag{27}$$

which is then also estimated by OLS. Another way of describing this two-step procedure is to say that the γ in the y equation is assumed to equal the OLS estimate of γ from the forecasting equation. Tests of the neutrality proposition then involve adding current and lagged values of X to the y equation to yield an equation similar to (18),

$$y_t = \tilde{y}_t + \sum_{i=0}^{N} \beta_i' \hat{u}_{t-i} + \sum_{i=0}^{N} \delta_i X_{t-i} + \epsilon_t, \tag{28}$$

and testing with a standard F test the null hypothesis that the δ coefficients of X_{t-i} are equal to zero.

This methodology raises several issues, the most important of which deal with the econometrics. The two-step procedure will yield consistent parameter estimates. However, it does not generate valid F test statistics. This procedure implicitly assumes that there is no uncertainty in the estimate of $\hat{\gamma}$. This results in inconsistent estimates of the standard errors of the parameters and hence test statistics that do not have the assumed F distribution. This can lead to inappropriate inference (see Pagan [1981] for a formal proof of this statement).

The joint estimation procedure generates valid test statistics because it does not ignore the uncertainty in the estimate of γ. It has two other advantages over the two-step procedure. The joint procedure will result in more efficient estimates of parameters because the X and y equations each make use of the other's information in the estimation process. The joint procedure also generates tests of both the neutrality and rationality implications of the MRE hypothesis, whereas the two-step procedure cannot test for rationality and is capable of testing only for neutrality.

What relationship exists between tests of neutrality using the joint versus the two-step procedure? Is the joint procedure more likely than the two-step procedure to lead to a rejection of neutrality? The answer is no: the opposite is true. By the nature of likelihood maximization in constrained systems, the joint procedure must attain as high or higher a likelihood than if the forecasting equation is forced to remain unchanged, as in the two-step procedure. The likelihood ratio statistic from the joint procedure should be smaller than the corresponding statistic from the two-step procedure. Therefore, the joint estimation procedure used in this book will be even more favorable to the neutrality hypothesis.

That the two-step procedure is biased toward rejecting neutrality and is less favorable to this null hypothesis than the joint procedure is borne out by a comparison of actual neutrality tests using both procedures. For example, in Chapter 6, model 4.1 (see table 6.4), the likelihood ratio statistic from the two-step procedure testing the neutrality constraints is $\chi^2(4) = 22.14$, with a marginal significance level of .0002 rejecting neutrality. The corresponding F statistic is $F(4,78) = 5.31$ with a marginal significance level of .0008. (The marginal significance level is the probability of obtaining as high a value of the test statistic or higher under the null hypothesis. A marginal significance level less than .01 indicates rejection of the null hypothesis at the 1 percent level.) In table 6.1, the test statistic using the joint procedure is only $\chi^2(4) = 15.45$, with a marginal significance level of .0039. Obviously, the bias of the two-step procedure against the neutrality null hypothesis is not negligible.

The two-step procedure suffers also from a conceptual problem more minor than the econometric criticisms of the procedure. It assumes that the OLS $\hat{\gamma}$, the estimate of γ which minimizes the mean-squared forecasting error, is used in forming expectations in the y equation. Rationality of expectations implies only that subjective probability distributions do not differ from the true probability distributions. This implies that the γ which is *expected* to minimize the mean-squared forecasting error is used in forming expectations and not the *actual* $\hat{\gamma}$ which minimizes the mean-squared error. Thus, in finite samples, the two-step procedure makes an overly strong assumption about expectations formation. This criticism is another way of stating the conceptual difficulty with using regression equations to measure anticipations of variable values early in the sample period when later data are used in estimating the regression relationship. Anticipations are made with information from the future as well as from the past, which clearly goes beyond the rational expectations principle. Note that the joint estimation procedure does not suffer from this problem. As rationality implies in this case, the γ which is expected to minimize the mean-squared forecasting error is used to form expectations in the y equation. As a practical matter, however, this criticism of the two-step procedure is not extremely important, because the OLS $\hat{\gamma}$'s are not very different from the jointly estimated $\hat{\gamma}$'s and asymptomatically they will not differ.

One last point about estimation methodology is worth discussing. Someone used to analyzing the neutrality proposition with the two-step procedure will tend to focus on the deterioration in fit from the imposition of the neutrality constraints of the y equation alone. Such a tendency will be highly misleading in the case of the estimated equations from the joint procedure. In the joint estimation procedure, if constraints are imposed on the y equation, the deterioration in fit is spread over both this equation and the forecasting equation. Thus the deterioration in the y equation fit will not be as severe as when the fit of the forecasting equation is not allowed to change, as in the two-step procedure. However, the likelihood ratio statistic in either (23) or (24) demonstrates that the deterioration of fit in both equations is involved in testing constraints. Therefore, strong rejections can occur even though there is only a small decline in R^2 (or rise in the standard error) of the y equation.[4]

The specification of the forecasting equation in previous empirical work sometimes violates a rational expectations principle. The theory of rational expectations implies that X_t^e in the y equation should be an optimal, one-period-ahead forecast conditional on information available at time $t - 1$. Thus, an appropriate forecasting equation should rely only on lagged explanatory variables. The procedure for specifying the forecasting equations here does satisfy this principle. However, this is not true in empirical studies which have used the Barro (1977) specification for the money growth forecasting equation. They include a contemporaneous variable (FEDV$_t$, the deviation of federal expenditures from the

4. The most striking example in Chapter 6 occurs when results of the model 2.1 (see tables 6.1 and 6.2) are compared with the 5.1 results (see table 6.5). The comparison is a little tricky because the model 2.1 is not strictly nested in model 5.1 because of the polynomial distributed lag specification, but it is still interesting to see what test statistics arise if we ignore this problem. The pseudolikelihood ratio statistic using (23) of the null hypothesis $\delta_o = \delta_1 \ldots = \delta_{20} = 0$ and $\beta_8 = \ldots \beta_{20} = 0$ equals 11.69 with a marginal significance level of .0199. Thus the hypothesis is rejected at the 5 percent level even though there is only a small change in the R^2 and standard error of the output equation in going from 2.1 to 5.1. A numerical explanation of the pseudolikelihood ratio statistic illustrates the point in the text. The maximum likelihood estimates of the standard errors of the 2.1 and 5.1 output equations are, respectively, .00796 and .00774. The percentage difference, calculated as the change in the logs, is 2.8 percent. The maximum likelihood standard errors for 2.1 and 5.1 money growth equations are, respectively, .00409 and .00394, with a percentage difference of 3.7 percent. Both of these percentage differences are added up in calculating the likelihood ratio statistic in (23), which is $92[2\ (.028 + .037)] \approx 12$.

normal level) as an explanatory variable in the forecasting equation. Yet it is unlikely that the market has complete knowledge of this variable at time $t - 1$. That this is a possibly serious misspecification can be seen as follows. Denoting the contemporaneous variable by A_t, the forecasting equation can be written as

$$X_t = Z_{t-1}\gamma + \xi A_t + u_t. \tag{29}$$

Using rational expectations and denoting $E(\ldots | \phi_{t-1})$ by E_{t-1}, unanticipated X_t is

$$\begin{aligned} X_t - X_t^e &= X_t - E_{t-1}X_t = X_t - (Z_{t-1}\gamma + \xi E_{t-1}A_t) \\ &= (X_t - Z_{t-1}\gamma - \xi A_t) + \xi(A_t - E_{t-1}A_t) \\ &= u_t + \xi(A_t - E_{t-1}A_t). \end{aligned} \tag{30}$$

Expression (30) is not equivalent to the residual from the forecasting equation, for it differs by an expression involving unanticipated A_t. It is valid to use residuals from the forecasting equation to proxy for unanticipated X only if there are no errors in forecasting A_t. As is shown in the next chapter, this misspecification can render test statistics for rationality invalid. Note, however, the more accurately A_t can be predicted, the less serious this misspecification becomes.

This chapter's discussion of the specification of the lag length N suggests that MRE models with fairly long lags deserve study. The criterion for specifying the lag length N in earlier studies, on the other hand, results in a fairly short lag length—on the order of two years. The lag length is chosen by cutting off the lags when the coefficients on the unanticipated variables are no longer statistically significant in the MRE equation. If the MRE hypothesis is not valid, then choosing the lag length from an MRE equation is inappropriate for testing this hypothesis. This is then a further justification for experimenting with MRE models with longer lag lengths, as is done in Chapter 6.

Appendix 2.1: Identification and Testing

The various tests discussed in this chapter depend on estimation of the parameters δ_i and γ^* in the unconstrained system (17). More specifically, neutrality requires that the estimate of δ_i not differ significantly from zero, and rationality requires that the estimate of γ^* not differ significantly from γ. These restrictions are testable only if the relevant parameters are identified, that is, if observational equivalence is avoided. If not all of the parameters are identified, then only some of the restrictions or linear combinations of restrictions are testable.

Appendix 2.1 is based on joint work with Andrew Abel (Abel and Mishkin 1983, sec. 5).

A procedure is outlined here for determining identification by analyzing an interesting special case of systems (16)–(20), where Z_{t-1} is rewritten as shown below in system (17):

$$\text{(A1)} \qquad X_t = \sum_{i=1}^{M} Z_{t-i}\gamma_i + u_t$$

$$y_t = \bar{y}_t + \sum_{j=0}^{N}\left(X_{t-j} - \sum_{i=1}^{M} Z_{t-j-i}\gamma_i^*\right)\beta_j + \sum_{j=0}^{N}\left(\sum_{i=0}^{M} Z_{t-j-i}\gamma_i^*\right)\delta_j + \epsilon_t,$$

where

X_t = a k-element row vector of variables relevant for determining y_t; $k \geq 1$.
Z_{t-i} = a $(p + k)$-element row vector of variables dated $t - i$ which are used in predicting X_t. It contains the k elements of X_{t-i} as well as p other variables; $p \geq 0$.
y_t = a scalar.

γ_i and γ_i^* = $(p + k)$ matrices of parameters.
β_i and δ_j = $k \times 1$ column vectors of parameters.

Observe that this system embodies the exclusion restriction that Z_{t-i} does not enter the y equation except as it enters terms representing X^e_{t-i}. The exclusion restriction is crucial to the discussion of identification and hypothesis testing. Note that (A1) embodies the following simplifying assumptions: (*a*) the same lag length applies to all variables used to predict X_t in the first equation; and, (*b*) in the second equation the same lag length, N, is used for both anticipated and unanticipated X_t. These assumptions, which are made for expositional clarity, can be relaxed and the following discussion can be generalized in a straightforward manner. Note also that the row vector Z_{t-i}, which is used in the time-series model for predicting X_t, contains the k-element row vector X_{t-i}, since lagged values of the dependent variable are often useful in prediction. In addition, the row vector Z_{t-i} contains p other variables at time $t - i$, where $p \geq 0$. It is assumed that u_t and ϵ_t are uncorrelated and that $E(u_t|\phi_{t-1}) = E(\epsilon_t|\phi_{t-1}) = 0$. Finally, recall that the rationality restriction is $\gamma_i = \gamma_i^*$, $i = 1, \ldots, M$, and the neutrality restriction is $\delta_j = 0$, $j = 0, \ldots, N$.

The first step in determining identification is to analyze the order condition. Consider, for example, the most unconstrained system (A1) in which γ_i, γ_i^*, β_j, and δ_j are the free parameters to be estimated. Observe that γ_i can be estimated by OLS on the first equation in (A1). The remaining parameters γ_i^*, β_j, and δ_j are estimated from the second equation in (A1). The most constrained form of this second equation is

(A2) $$y_t = \sum_{j=0}^{N} \hat{u}_{t-j}\beta_j + \sum_{l=1}^{M+N} Z_{t-l}\theta_l + \epsilon_t$$

where

$$\hat{u}_{t-j} = X_{t-j} - \sum_{i=1}^{M} Z_{t-j-i}\hat{\gamma}_i$$

θ_ℓ = a $(p + k) \times 1$ column vector of parameters which is zero if $\delta_j = 0, j = 0, \ldots, N$ and $\gamma_i^* = \gamma_i, i = 1, \ldots, M$

$= \sum_{t+j=l} [(\gamma_i - \gamma_i^*)\beta_j + \gamma_i^*\delta_j]$, $1 \leq i \leq M$ and $0 \leq j \leq N$.

Note that for $j = 1, \ldots, N$, the residual $\hat{u}_{t-j}$ can be expressed as a linear combination of the other right-hand-side variables $Z_{t-1}, \ldots, Z_{t-M-N}$. That is, only the residual at time t, $\hat{u}_t$, is not perfectly correlated with the other right-hand-side variables. Hence, the most unconstrained form of this equation that can be estimated by OLS is

(A3) $$y_t = \hat{u}_t\beta_o + \sum_{l=1}^{M+N} Z_{t-l}\theta_l + \epsilon_t.$$

Since there are k elements in β_o and $(M + N)(p + k)$ elements in the θ coefficients, equation (A3) can be used to estimate at most $k + (M + N)(p + k)$ parameters. As long as this number of estimable parameters exceeds the number of free parameters contained in the β, δ, and γ^* coefficients, the order condition is satisfied.

Identification depends on the rank condition as well as the order condition. The rank condition is particularly important in the identification of (A3) because, in general, it need not be satisfied at the same time as the order condition. This failure to satisfy the rank condition becomes clear if we rewrite (A1) as

(A4) $$X_t^1 = \sum_{i=1}^{M} Z_{t-i}\gamma_i^1 + u_t^1$$

.

.

.

$$X_t^k = \sum_{i=1}^{M} Z_{t-i}\gamma_i^k + u_t^k$$

$$y_t = \sum_{s=1}^{k} \left[\sum_{j=0}^{N} (X_{t-j}^s\beta_j^s) + \sum_{j=0}^{N} (\delta_j^s - \beta_j^s) \sum_{i=1}^{M} Z_{t-i-j}\gamma_i^{*s} \right],$$

where X_t^s, γ_i^s, γ_i^{*s}, and u_t^s are the sth columns of X_t, γ_i, γ_i^*, and u_t, respectively. The scalars β_j^s and δ_j^s are the sth elements of β_j and δ_j, respectively.

Note that for any particular s, say s_o, the system will be unchanged by a doubling of all the elements of $\gamma_i^{*s_o}$ for all i and a halving of $\delta_j^{s_o} - \beta_j^{s_o}$ for all j. Because of this observational equivalence, the parameters $\delta_j^{s_o} - \beta_j^{s_o}$ and $\gamma_i^{*s_o}$ are not identified even when the order condition is satisfied. A restriction on any element of $\delta_j^{s_o}$ or $\gamma_i^{*s_o}$ is sufficient to identify these parameters. If we apply this argument to each of the k values of s, it is clear that k additional restrictions are needed for identification. The restrictions will be provided if either neutrality ($\delta_j = 0$) or rationality ($\gamma_i = \gamma_i^*$) is treated as a maintained hypothesis. Thus, only if neither neutrality nor rationality is maintained will the rank condition fail to be satisfied in situations when the order condition is satisfied.

Tests of hypotheses are conducted by comparing the residual sums of squares from constrained and unconstrained systems. The number of restrictions tested (and hence the number of degrees of freedom in the χ^2 statistic) equals the number of identified parameters estimated in the unconstrained system, less the number of identified parameters estimated in the constrained system. To illustrate this calculation using the procedures above, consider in the efficient-markets case in which $N = 0$, the test of rationality under the maintained hypothesis of neutrality. The last equation in the constrained system (where $\delta_o = 0, \gamma_i = \gamma_i^*$) contains k parameters (the elements of β), all of which are identified. The last equation in the unconstrained system (where $\delta_o = 0$) contains $k + Mk(p + k)$ parameters. However, as explained above, only $k + M(p + k)$ parameters can be estimated. Only if $k = 1$ will all of the parameters in the unconstrained system be identified. However, even if $k > 1$, there are $M(p + k)$ testable restrictions. These restrictions are linear combinations of the restrictions $\gamma - \gamma^* = 0$ (see the next chapter for an example).

Another test which may be conducted in the efficient-markets framework ($N = 0$) is a test of the null hypothesis of neutrality under the maintained hypothesis of rationality. Recall that the last equation of the constrained system ($\gamma_i = \gamma_i^*, \delta_o = 0$) contains k parameters (the elements of β), and observe that the last equation of the unconstrained system ($\gamma_i = \gamma_i^*$) contains $2k$ parameters (the elements of β and δ_o). In both the constrained and unconstrained systems, all parameters are identified and all k neutrality restrictions are testable.

A third test in the efficient-markets framework is a test of the joint hypothesis of neutrality and rationality. As in the first two tests, all k parameters of the last equation in the constrained system are identified. In the unconstrained system the last equation contains $2k + Mk(p + k)$ parameters (k elements of β, k elements of δ_o and $Mk(p + k)$ elements of γ_i^*, $i = 1, \ldots, M$), but, as explained above, only $k + M(p + k)$ parameters can be estimated. Therefore, under no circumstances will all parameters of this equation be identified. However, there are $M(p + k)$ testable restrictions that are linear combinations of the restrictions $\gamma - \gamma^* = 0$ and $\delta_o = 0$.

The interpretation of these efficient-markets tests depends on what hypothesis is maintained. In particular, the test statistic associated with the joint test of rationality and neutrality is identical to the test statistic for the test of rationality, under the maintained hypothesis of neutrality. This follows because, although the free parameters in the unconstrained systems are different, the estimated coefficients are identical. Furthermore, the constrained systems are the same. Because of the equivalence of the two tests, one cannot determine whether a rejection is due to a violation of rationality alone or a violation of both rationality and neutrality.

Tests of policy neutrality under the maintained hypothesis of rationality as in Barro (1977, 1978) and in Chapter 6 furnish another interesting case. These models assume that the deviation of current output from its natural level is affected only by the current and N lagged surprises in a single policy variable (i.e., $k = 1$ and $N > 0$). To obtain identification of the coefficients on surprises in the policy variable, these studies implicitly place restrictions on the covariance of ϵ_t with both u_t and with lagged disturbances. There are two alternative conditions sufficient for identification of the δ coefficients, that is, the coefficients on anticipated policy. One condition, discussed and used by Barro (1977, 1978, 1979), Leiderman (1980), and in Chapter 6, is the exclusion restriction $p \geq 1$. That is, the time-series model for the policy variable X_t contains at least one variable that is not directly included in the y equation. The y equation in the constrained system (where $\delta_i = 0$ and $\gamma_i = \gamma_i^*$) contains $N + 1$ parameters ($\beta_o, \ldots, \beta_N$), and in the unconstrained system (where $\gamma_i = \gamma_i^*$) it contains $2(N + 1)$ parameters ($\beta_o, \ldots, \beta_N$ and $\delta_o, \ldots, \delta_N$). In each of these systems, all of the parameters are identified because the number of free parameters is less than the number of estimable parameters, $1 + (M + N)(p + 1)$. Therefore all of the $N + 1$ neutrality restrictions are testable.

The alternative sufficient condition for identification is $M > N$; that is, the number of lags in the time-series model for the policy variable X_t exceeds the number of lagged surprises in the y equation. Although this condition formally leads to identification, it requires strong a priori knowledge of lag lengths. Without this prior knowledge we are faced with the observational equivalence problem raised by Sargent (1976*b*).

To identify δ_i at least one of the two conditions above must hold. One recent example in which this does not occur is in Grossman (1979). His specification of the time-series equation describing his policy variable (nominal GNP growth) does not include any variable other than lagged dependent variables. Moreover, the number of lags in the output equation exceeds that in the time-series equation for the policy variable. Therefore, the δ coefficients in his model are not identified, with the result that not all the neutrality constraints can be tested.

Appendix 2.2: An Annotated Computer Program

The computer program here demonstrates how the models discussed in this book can be estimated. The particular example is chosen from Chapter 6 to illustrate the general principle of estimating models where (1) current and lagged values of *both* anticipated and unanticipated variables have explanatory power, and (2) the error term is specified to follow an autoregressive process. The program makes use of the PROC NLIN nonlinear estimation procedure in the widely available computer package SAS, described in the *SAS User's Guide* (1979). The detailed discussion of this sample program should not only allow a user of SAS to exploit the techniques described in this book, but also should provide enough of the program's logic so that it can be modified for use with other econometric packages with nonlinear estimation capabilities. It should be noted that the PROC NLIN procedure of SAS does have one major advantage: it can handle extremely large problems that are beyond the capability of other packages. This is not important for a small estimation problem, but it is crucial for estimation of models such as those found in Chapter 6 which have over fifty parameters. My experience with SAS's nonlinear estimation routine has been a happy one: it converges quickly and is not prohibitively expensive to use.

The program here estimates over the period 1954:1–1976:4 a model consisting of (A5), a forecasting equation for money growth, and (A6), an output equation in which both anticipated and unanticipated money growth matter.

(A5)
$$M1G = \gamma_o + \sum_{i=1}^{4} \gamma_i M1G_{t-i} + \sum_{i=1}^{4} \gamma_{i+4} \mathrm{RTB}_{t-i} + \sum_{i=1}^{4} \gamma_{i+8} \mathrm{SURP}_{t-i} + u_t,$$

(A6)
$$\log(\mathrm{GNP}_t) = c + \tau \mathrm{TIME} + \sum_{i=0}^{7} \beta_i (M1G_{t-i} - M1G^e_{t-i}) + \sum_{i=0}^{7} \delta_i M1G^e_{t-i} + \epsilon_t,$$

where

$$\epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + \rho_3 \epsilon_{t-3} + \rho_4 \epsilon_{t-4} + \eta_t,$$

$$M1G^e_t = \gamma_o + \sum_{i=1}^{4} \gamma_i M1G_{t-i} + \sum_{i=1}^{4} \gamma_{i+4} \mathrm{RTB}_{t-i} + \sum_{i=1}^{4} \gamma_{i+8} \mathrm{SURP}_{t-i}.$$

The cross-equation restrictions are that the γ_i are identical in (A5) and (A6). The variables are as defined in Chapter 6. Note that this example does not make use of the polynomial distributed lag (PDL) restriction. The interested reader is referred to Kmenta (1971) to see how the PDL restriction can be imposed by "scrambling" variables.

The basic idea of the program is to stack the data so that the system of the two linear equations, (A5) and (A6), can be written as one equation with the appropriate nonlinear constraints. Estimation with the nonlinear procedure PROC NLIN is then fairly straightforward.

Notes for Program Listing in Exhibit A1

The SAS data set ONE contains the data used in estimation. The 120 quarterly observations run from 1947:1 to 1976:4. A number appended to the variable name indicates how many times it is lagged. For example, $M1G$ is unlagged money growth while $M1G1$ is money growth lagged one period. LGNP equals log (GNP) and C is the constant term.

Lines 1–17: The new data set ONEA created from ONE weights the variables in the forecasting equation by HETA in order to correct for the heteroscedasticity across equations. The value of HETA is chosen so that the weighted sum of squared residuals in each equation approach each other. The procedure for doing this will be explained when the output from the program is discussed.

Lines 18–21: The LGNP variable is dropped from the data set and the $M1G$ variable is renamed as LGNP. This operation is necessary for the stacking operation conducted later.

Lines 22–24: The new data set ONER will correspond to the output equation and it adds the constant term to the data set ONE.

Lines 25–76: Here the stacking operation is conducted in order to create the data set EST used in estimation. The outcome of this operation will be discussed first so that we may more easily follow the steps taken to achieve it. Each variable will have 240 observations with the first 120 corresponding to the output equation and the second 120 corresponding to the forecasting equation. If the weighted variables are denoted by the superscript A, then the resulting LGNP variable written in matrix notation is:

$$\text{LGNP} = \begin{pmatrix} \text{LGNP}_{1947:1} \\ \cdot \\ \cdot \\ \cdot \\ \text{LGNP}_{1976:4} \\ M1G^{A}_{1947:1} \\ \cdot \\ \cdot \\ \cdot \\ M1G^{A}_{1976:4} \end{pmatrix}$$

Hence the first 120 observations correspond to the dependent variable of the output equation while the second 120 observations correspond to the dependent variable of the forecasting equation (appropriately weighted for heteroscedasticity). The variables with an A added to their names correspond to the appropriately weighted explanatory variables in the forecasting equation, while those without A (except for LGNP) correspond to the explanatory variables in the output equation. For example,

$$M1G1 = \begin{bmatrix} M1G1_{1947:1} \\ \cdot \\ \cdot \\ \cdot \\ M1G1_{1976:4} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \quad M1G1A = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ M1G1^A_{1947:1} \\ \cdot \\ \cdot \\ \cdot \\ M1G1^A_{1976:4} \end{bmatrix}.$$

In the case of $M1G1$, the 120 observations corresponding to the forecasting equation are set to zero, while in the case of $M1G1A$ the 120 observations corresponding to the output equation are set to zero.

Lines 25–26 conduct the first stacking operation to create data set TWO. All the variables have 240 observations. The operations in lines 18–21 result in a LGNP variable of the form shown above, with the first 120 observations containing the dependent variable of the output equation and the second 120 containing the dependent variable of the forecasting equation. For all other variables, the first 120 observations from the data set ONER correspond to the output equation, and the second 120 observations from the data set ONEA correspond to the forecasting equation. Lines 27–35 add to a new data set THREE the variables with an A which are identical to their counterparts without A. Lines 36–37 have data set EST created from data set THREE. Lines 38–63 set to zero the second 120 observations of the variables with no A, and lines 64–76 set to zero the first 120 observations of the variables with an A. The stacked variables described above are the outcome of these operations.

Lines 77–78: These lines set the first twenty-eight observations of both sets of 120 observations in LGNP to a missing value. This ensures that when PROC NLIN is used in the following lines, the 1947:1–1953:4 observations are excluded from the sample period and estimation over the 1954:1–1976:4 sample period results.

Lines 79–247: Here the actual estimation is carried out with PROC NLIN. The parameters have slightly different names than in (A5) and (A6) above: CO corresponds to c, T to τ, $M0$–$M7$ to β_0–β_7, $E0$–$E7$ to δ_0–δ_7, $A0$–$A12$ to γ_0–γ_{12}, and $RH01$–$RH04$ to ρ_1–ρ_4.

Lines 79–80: The convergence criterion is set and the residuals from the estimation are stored as the variable RESID in the data set DRESID.

Lines 81–117: The starting values for the parameters are provided.

Lines 118–135: Variables are generated here to facilitate calculations of the derivatives in lines 193–247. If these derivatives are not needed, then these lines can be deleted.

Lines 136–139: Anticipated money growth, *EM*, is generated.

Lines 140–151: Unanticipated money growth, *UM*, and its lags are generated.

Lines 152–162: Lags of *EM* are generated.

Lines 163–178: The fourth-order autoregressive correction for serial correlation in the output equation (A6) requires the transformation here of the *UM* and *EM* variables into *RUM* and *REM*, as shown.

Lines 179–192: The model consisting of both the output and forecasting equation is written down here. Note that it incorporates the necessary transformation to allow for the serial correlation correction. The stacking operation in previous lines ensures also that this model captures the cross-equation restrictions and the appropriate heteroscedasticity correction.

Lines 193–247: The derivatives of the model in lines 179–192 are calculated here. The version of SAS used to estimate this model required these derivatives. Later versions of SAS may not require them, in which case these lines and lines 118–135 can be deleted.

Lines 248–259: Here the standard errors of both output and forecasting equations are calculated. They are used, as will be shown below, to calculate HETA for the heteroscedasticity correction and to decide when the last iteration is reached. Lines 248–250 retain only the residuals in the data set DRESID. Lines 251–259 use PROC MEANS to calculate the standard error first of the output equation and then of the weighted forecasting equation.

Discussion of the Output in Exhibit A2

The first page of the SAS output shows the convergence to the minimum sum of squared residuals, and pages 3–5 show the asymptotic correlation matrix of the parameter estimates. Only pages 2, 6, and 7 are displayed as they are of the greatest interest. Page 2 contains the parameter estimates, their asymptotic standard errors, and the sum of squared residuals of the system. For example, the coefficient of the constant term in the output equation is 6.18857905 with an asymptotic standard error of .04752109. The sum of squared residuals of the system, which is needed to calculate the likelihood ratio tests discussed in the chapter, is .01012971. Pages 6 and 7 show the standard errors of the output equation and the weighted forecasting equation, respectively, in the standard deviation column. The standard error of the output equation is .00738342, and the standard error of the weighted forecasting equation is .00753653.

The iterative procedure that corrects for heteroscedasticity across the equations continues as follows. The variables in the forecasting equation are weighted by the ratio from the previous iteration of the standard error of the forecasting equation to the standard error of the output equation. This means that the weighting variable HETA from the previous iteration needs to be multiplied by the standard error of the *weighted* forecasting equation divided by the standard error of the output equation. In the example here, the next iteration would therefore multiply the previous iteration's HETA by .00753653 ÷ .00738432, which equals 1.020612595. That is, line 3 of the program would be modified to insert *1.020612595 just before the semicolon, and the program would then be run. Note that computational costs have been lowered by using the last iteration's parameter estimates as starting values in lines 81–117. The criterion for terminating the iterative procedure can be varied but, in the empirical work reported in this book, if the standard errors of the weighted forecasting equation and the output equation differed by less than 2½ percent, then no further iterations were performed. Thus the results reported in Exhibit A2 are the final iteration.

Procedures for Calculating the Likelihood Ratio Tests

To carry out the tests in Chapter 6, the first system estimated was the most constrained where anticipated money has no effect on output but rationality is still imposed. The only changes needed in the computer program are to eliminate terms involving *REM* and *EM* from the model and derivative statements and to delete lines 92–99 and 203–210. The next, less constrained system estimated has anticipated money affecting output and makes use of the program in Exhibit A1. The first iteration uses the same HETA value used in the final iteration of the most constrained system. The likelihood ratio test of neutrality described in Chapter 2 is conducted by comparing the sum of squares of the less constrained system obtained from the first iteration, with the sum of squares for the final iteration of the most constrained system. Further iterations are then performed for this system in which anticipated money matters until the termination criterion is reached.

The most unconstrained system is subject neither to rationality nor to neutrality, and as there are now no binding constraints across the two equations of the system, each can be estimated separately. The forecasting equation can be estimated by OLS while the output equation is estimated by deleting lines 1–21, 25–76, 78, and 188–191 from the program in Exhibit A1 and modifying the derivatives statements appropriately. Note that the *CO* and *AO* parameters are not identified and so one of them should be set to a constant. As discussed in Appendix 2.1, at least one other parameter will not be identified and PROC NLIN will automatically set it to a constant in estimation. In some cases when more parameters are unidentified, the most unconstrained output equation is even more linear, and so takes an even simpler form.

The likelihood ratio tests of neutrality and rationality jointly, or of rationality alone, compare the sum of squared residuals of the appropriately weighted most unconstrained system with those of the more constrained systems, estimation of which is discussed above. The appropriately weighted sum of squared residuals for the most unconstrained system equals the sum of squared residuals from the most unconstrained output equation, added to the sum of squared residuals from the OLS estimated forecasting equation, divided by the square of the HETA value used in the constrained system's final iteration.

Exhibit A1 Program Listing and Output

Line No.

```
DATA ONEA;
SET ONE;
HETA = .4204183267*1.200367097*1.02580668*1.06044268;
M1G = M1G/HETA;
M1G1 = M1G1/HETA;
M1G2 = M1G2/HETA;
M1G3 = M1G3/HETA;
M1G4 = M1G4/HETA;
RTB1 = RTB1/HETA;
RTB2 = RTB2/HETA;
RTB3 = RTB3/HETA;
RTB4 = RTB4/HETA;
SURP1 = SURP1/HETA;
SURP2 = SURP2/HETA;
SURP3 = SURP3/HETA;
SURP4 = SURP4/HETA;
C = 1/HETA;
DATA ONEA;
SET ONEA;
DROP LGNP;
RENAME M1G=LGNP;
DATA ONER;
SET ONE;
C = 1;
DATA TWO;
SET ONER ONEA;
DATA TWOA;
SET TWO;
RENAME
M1G1=M1G1A M1G2=M1G2A M1G3=M1G3A M1G4=M1G4A
RTB1=RTB1A RTB2=RTB2A RTB3=RTB3A RTB4=RTB4A
SURP1=SURP1A SURP2=SURP2A SURP3=SURP3A SURP4=SURP4A
C=CA;
DATA THREE;
MERGE TWO TWOA;
DATA EST;
SET THREE;
IF _N_>=121 THEN M1G=0;
IF _N_>=121 THEN C=0;
IF _N_>=121 THEN TIME=0;
IF _N_>=121 THEN TIME1=0;
IF _N_>=121 THEN TIME2=0;
IF _N_>=121 THEN TIME3=0;
IF _N_>=121 THEN TIME4=0;
IF _N_>=121 THEN M1G1=0;
IF _N_>=121 THEN M1G2=0;
IF _N_>=121 THEN M1G3=0;
IF _N_>=121 THEN M1G4=0;
IF _N_>=121 THEN M1G5=0;
IF _N_>=121 THEN RTB1=0;
IF _N_>=121 THEN RTB2=0;
IF _N_>=121 THEN RTB3=0;
IF _N_>=121 THEN RTB4=0;
IF _N_>=121 THEN RTB5=0;
IF _N_>=121 THEN SURP1=0;
IF _N_>=121 THEN SURP2=0;
IF _N_>=121 THEN SURP3=0;
IF _N_>=121 THEN SURP4=0;
IF _N_>=121 THEN SURP5=0;
IF _N_>=121 THEN LGNP1=0;
IF _N_>=121 THEN LGNP2=0;
IF _N_>=121 THEN LGNP3=0;
IF _N_>=121 THEN LGNP4=0;
IF _N_<121 THEN M1G1A=0;
IF _N_<121 THEN M1G2A=0;
IF _N_<121 THEN M1G3A=0;
IF _N_<121 THEN M1G4A=0;
IF _N_<121 THEN RTB1A=0;
IF _N_<121 THEN RTB2A=0;
IF _N_<121 THEN RTB3A=0;
IF _N_<121 THEN RTB4A=0;
IF _N_<121 THEN SURP1A=0;
IF _N_<121 THEN SURP2A=0;
IF _N_<121 THEN SURP3A=0;
IF _N_<121 THEN SURP4A=0;
```

```
IF _N_<121 THEN CA=0;
IF _N_<=28 THEN LGNP=.;
IF _N_>120 AND _N_<=148 THEN LGNP=.;
PROC NLIN CONVERGENCE=.0001;
OUTPUT OUT=DRESID PREDICTED=PRED RESIDUAL=RESID;
PARAMETERS
                        C0          =     6.18409321
                        T           =     0.00818261
                        M0          =     0.70203726
                        M1          =     1.02226901
                        M2          =     1.96325236
                        M3          =     2.39452781
                        M4          =     2.81242963
                        M5          =     2.59002714
                        M6          =     2.28592198
                        M7          =     1.31723568
                        E0          =     0.76382117
                        E1          =    -0.19910460
                        E2          =     0.24108277
                        E3          =    -0.40288643
                        E4          =    -0.07456418
                        E5          =    -0.59607729
                        E6          =     0.47914434
                        E7          =     0.95216249
                        A0          =     0.00210000
                        A1          =     0.73200405
                        A2          =     0.02305185
                        A3          =    -0.08331019
                        A4          =    -0.13184558
                        A5          =    -0.00226408
                        A6          =     0.00451728
                        A7          =    -0.00132287
                        A8          =    -0.00035536
                        A9          =    -0.00017881
                        A10         =     0.00019747
                        A11         =    -0.00000643
                        A12         =    -0.00012937
                        RHO1        =     1.19896060
                        RHO2        =    -0.42906225
                        RHO3        =     0.12553469
                        RHO4        =     0.03276445
;
ZC = C*(1-RHO1-RHO2-RHO3-RHO4);
MZC     = ZC*(-M0-M1-M2-M3-M4-M5-M6-M7);
EZC=ZC*(E0+E1+E2+E3+E4+E5+E6+E7);
ZM = M1G1 - RHO1*M1G2 - RHO2*M1G3 - RHO3*M1G4 - RHO4*M1G5;
MZM     = -M0*ZM - M1*LAG1(ZM) -M2*LAG2(ZM) - M3*LAG3(ZM) -M4*LAG4(ZM)
-M5*LAG5(ZM) - M6*LAG6(ZM) -M7*LAG7(ZM);
ZR = RTB1 - RHO1*RTB2 - RHO2*RTB3 - RHO3*RTB4 - RHO4*RTB5;
MZR     = -M0*ZR - M1*LAG1(ZR) -M2*LAG2(ZR) - M3*LAG3(ZR) -M4*LAG4(ZR)
-M5*LAG5(ZR) - M6*LAG6(ZR) -M7*LAG7(ZR);
ZH = SURP1 - RHO1*SURP2 - RHO2*SURP3 - RHO3*SURP4 - RHO4*SURP5;
MZH     = -M0*ZH - M1*LAG1(ZH) -M2*LAG2(ZH) - M3*LAG3(ZH) -M4*LAG4(ZH)
-M5*LAG5(ZH) - M6*LAG6(ZH) -M7*LAG7(ZH);
EZM = E0*ZM + E1*LAG1(ZM) + E2*LAG2(ZM) + E3*LAG3(ZM) + E4*LAG4(ZM)
+ E5*LAG5(ZM) + E6*LAG6(ZM) + E7*LAG7(ZM) ;
EZR = E0*ZR + E1*LAG1(ZR) + E2*LAG2(ZR) + E3*LAG3(ZR) + E4*LAG4(ZR)
+ E5*LAG5(ZR) + E6*LAG6(ZR) + E7*LAG7(ZR) ;
EZH = E0*ZH + E1*LAG1(ZH) + E2*LAG2(ZH) + E3*LAG3(ZH) + E4*LAG4(ZH)
+ E5*LAG5(ZH) + E6*LAG6(ZH) + E7*LAG7(ZH) ;
EM = A0*C + A1*M1G1  + A2*M1G2  + A3*M1G3  + A4*M1G4
+ A5*RTB1  + A6*RTB2  - A7*RTB3  + A8*RTB4
+ A9*SURP1  + A10*SURP2  + A11*SURP3  +
A12*SURP4 ;
UM = M1G - EM;
UM1 = LAG1(UM);
UM2 = LAG2(UM);
UM3 = LAG3(UM);
UM4 = LAG4(UM);
UM5 = LAG5(UM);
UM6 = LAG6(UM);
UM7 = LAG7(UM);
UM8 = LAG8(UM);
UM9 = LAG9(UM);
UM10  = LAG10(UM);
UM11 = LAG11(UM);
EM1 = LAG1(EM);
EM2 = LAG2(EM);
EM3 = LAG3(EM);
EM4 = LAG4(EM);
EM5 = LAG5(EM);
```

Exhibit A1 (continued)

Line No.

```
EM6 = LAG6(EM);
EM7 = LAG7(EM);
EM8 = LAG8(EM);
EM9 = LAG9(EM);
EM10  = LAG10(EM);
EM11 = LAG11(EM);
RUM= UM - RHO1*UM1 - RHO2*UM2  - RHO3*UM3  - RHO4*UM4 ;
RUM1 = LAG1(RUM);
RUM2 = LAG2(RUM);
RUM3 = LAG3(RUM);
RUM4 = LAG4(RUM);
RUM5 = LAG5(RUM);
RUM6 = LAG6(RUM);
RUM7 = LAG7(RUM);
REM= EM - RHO1*EM1 - RHO2*EM2  - RHO3*EM3  - RHO4*EM4 ;
REM1 = LAG1(REM);
REM2 = LAG2(REM);
REM3 = LAG3(REM);
REM4 = LAG4(REM);
REM5 = LAG5(REM);
REM6 = LAG6(REM);
REM7 = LAG7(REM);
MODEL LGNP =
RHO1*LGNP1      + RHO2*LGNP2      + RHO3*LGNP3      + RHO4*LGNP4
C0*C*(1-RHO1-RHO2-RHO3-RHO4)   +
T*(TIME - RHO1*(TIME1) - RHO2*(TIME2) - RHO3*(TIME3)
- RHO4*(TIME4))
- E0*REM + E1*REM1  + E2*REM2  + E3*REM3  + E4*REM4
+ E5*REM5  + E6*REM6  + E7*REM7
+ M0*RUM + M1*RUM1  + M2*RUM2  + M3*RUM3  + M4*RUM4
+ M5*RUM5  + M6*RUM6  + M7*RUM7
+   A0*CA + A1*M1G1A  + A2*M1G2A  + A3*M1G3A  + A4*M1G4A
+ A5*RTB1A  + A6*RTB2A  + A7*RTB3A  + A8*RTB4A
+ A9*SURP1A  + A10*SURP2A  + A11*SURP3A  +
A12*SURP4A
;
DER.C0 = C*(1-RHO1-RHO2-RHO3-RHO4);
DER.T= (TIME - RHO1*TIME1 -RHO2*TIME2 -RHO3*TIME3 -RHO4*TIME4);
DER.M0 = RUM ;
DER.M1 = RUM1;
DER.M2 = RUM2;
DER.M3 = RUM3;
DER.M4 = RUM4;
DER.M5 = RUM5;
DER.M6 = RUM6;
DER.M7 = RUM7;
DER.E0 = REM ;
DER.E1 = REM1;
DER.E2 = REM2;
DER.E3 = REM3;
DER.E4 = REM4;
DER.E5 = REM5;
DER.E6 = REM6;
DER.E7 = REM7;
DER.A0 = MZC + EZC;
DER.A1 = MZM + EZM + M1G1A;
DER.A2 = LAG1(MZM) + LAG1(EZM) + M1G2A;
DER.A3 = LAG2(MZM) + LAG2(EZM) + M1G3A;
DER.A4 = LAG3(MZM) + LAG3(EZM) + M1G4A;
DER.A5 = MZR + EZR + RTB1A;
DER.A6 = LAG1(MZR) + LAG1(EZR) + RTB2A;
DER.A7 = LAG2(MZR) + LAG2(EZR) + RTB3A;
DER.A8 = LAG3(MZR) + LAG3(EZR) + RTB4A;
DER.A9 = MZH + EZH + SURP1A;
DER.A10 = LAG1(MZH) + LAG1(EZH) + SURP2A;
DER.A11 = LAG2(MZH) + LAG2(EZH) + SURP3A;
DER.A12 = LAG3(MZH) + LAG3(EZH) + SURP4A;
DER.RHO1 = LGNP1 - C0*C - T*(TIME1)
-E0*EM1 - E1*EM2 - E2*EM3 - E3*EM4 - E4*EM5
-E5*EM6 - E6*EM7 - E7*EM8
-M0*UM1 - M1*UM2 - M2*UM3 - M3*UM4 - M4*UM5
-M5*UM6 - M6*UM7 - M7*UM8
;
DER.RHO2 = LGNP2 - C0*C - T*(TIME2)
-E0*EM2 - E1*EM3 - E2*EM4 - E3*EM5 - E4*EM6
-E5*EM7 - E6*EM8 - E7*EM9
-M0*UM2 - M1*UM3 - M2*UM4 - M3*UM5 - M4*UM6
```

```
-M5*UM7 - M6*UM8 - M7*UM9
;
DER.RHO3 = LGNP3 - C0*C - T*(TIME3)
-E0*EM3 - E1*EM4 - E2*EM5 - E3*EM6 - E4*EM7
-E5*EM8 - E6*EM9 - E7*EM10
-M0*UM3 - M1*UM4 - M2*UM5 - M3*UM6 - M4*UM7
-M5*UM8 - M6*UM9 - M7*UM10
;
DER.RHO4 = LGNP4 - C0*C - T*(TIME4)
-E0*EM4 - E1*EM5 - E2*EM6 - E3*EM7 - E4*EM8
-E5*EM9 - E6*EM10 - E7*EM11
-M0*UM4 - M1*UM5 - M2*UM6 - M3*UM7 - M4*UM8
-M5*UM9 - M6*UM10 - M7*UM11
;
DATA DRESID;
SET DRESID;
KEEP RESID  ;
DATA DRESID4;
SET DRESID;
IF _N_<29 THEN DELETE;
IF _N_>120 THEN DELETE;
PROC MEANS;
DATA DRESID4;
SET DRESID;
IF _N_<149 THEN DELETE;
PROC MEANS;
```

Exhibit A2 **SAS Output**

(Page 2 of Output)

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS DEPENDENT VARIABLE LGNP

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	34	4265.59094289	125.45855714
RESIDUAL	150	0.01012971	0.00006753
UNCORRECTED TOTAL	184	4265.60107260	
(CORRECTED TOTAL)	183	2124.35825459	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
CO	6.18857905	0.04752109	6.09468097	6.28247712
T	0.00817790	0.00075394	0.00668817	0.00966763
MO	0.67338908	0.25174720	0.17595567	1.17082248
M1	0.99052411	0.63419173	-0.26259075	2.24363898
M2	1.98333150	0.87999722	0.24452360	3.72213940
M3	2.42454230	1.09553356	0.25985090	4.58923371
M4	2.87331469	1.12105105	0.65820267	5.08842671
M5	2.67999149	1.07604971	0.55379870	4.80618429
M6	2.40977049	0.92363811	0.58473149	4.23480950
M7	1.39301366	0.66200728	0.08493739	2.70108994
EO	0.75851491	0.68319790	-0.59143244	2.10846226
E1	-0.27845622	0.81563359	-1.89008646	1.33317402
E2	0.16799819	0.94744868	-1.70408870	2.04008508
E3	-0.49455089	0.87148992	-2.21654902	1.22744723
E4	-0.16589726	0.84832017	-1.84211370	1.51031918
E5	-0.69837726	0.77518540	-2.23008493	0.83333042
E6	0.47054569	0.69561698	-0.90394081	1.84503220
E7	0.99845498	0.60581543	-0.19859045	2.19550041
AO	0.00210000	0.00000000	0.00210000	0.00210000
A1	0.73786167	0.11300533	0.51457169	0.96115165
A2	0.01750937	0.14832346	-0.27556657	0.31058530
A3	-0.08423813	0.13137921	-0.34383350	0.17535724
A4	-0.13930343	0.10527634	-0.34732151	0.06871465
A5	-0.00201614	0.00083145	-0.00365903	-0.00037325
A6	0.00425483	0.00149704	0.00129679	0.00721287
A7	-0.00122770	0.00156798	-0.00432590	0.00187051
A8	-0.00041573	0.00089441	-0.00218303	0.00135156

A9	-0.00017315	0.00006205	-0.00029576	-0.00005055
A10	0.00020580	0.00007921	0.00004929	0.00036231
A11	-0.00001359	0.00008016	-0.00017199	0.00014480
A12	-0.00012975	0.00006598	-0.00026012	0.00000062
RHO1	1.19154724	0.12115199	0.95216009	1.43093440
RHO2	-0.42871113	0.18992137	-0.80398139	-0.05344087
RHO3	0.13010832	0.21066777	-0.28615525	0.54637190
RHO4	0.03918645	0.14039923	-0.23823180	0.31660470

(Page 6 of Output)

VARIABLE	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE	C.V.
RESID	92	-0.00000250	0.00738342	-0.02660273	0.01635174	0.00076977	-0.00022995	0.00005451	-99999.000

(Page 7 of Output)

VARIABLE	N	MEAN	STANDARD DEVIATION	MINIMUM VALUE	MAXIMUM VALUE	STD ERROR OF MEAN	SUM	VARIANCE	C.V.
RESID	92	0.00003520	0.00753653	-0.02391151	0.01685502	0.00078574	0.00323822	0.00005680	21411.797