DISCUSSION PAPER NO. 444

ESTIMATION OF MACROECONOMETRIC MODELS UNDER RATIONAL EXPECTATIONS: A SURVEY

by

FUMIO HAYASHI*/

October 1980

Department of Economics
Northwestern University

*/ I am grateful to Olivier-Jean Blanchard, Dale W. Jorgenson, and Kazuo Ueda for comments and discussions.

1. Introduction

Most behavioral equations in macroeconomics depend critically on expectations about current and future economic variables. Consumption function and investment function are obvious examples. Since expectations are not usually observable, econometricians who want to estimate such behavioral equations have to specify the way expectations are formed. The conventional practice is to assume autoregressive expectations and write expectations as a distributed lag function of their own past values. This practice is criticized by Lucas (1976) who argued that expectations made by reasonably intelligent economic agents may not be represented by distributed lags whose coefficients are fixed. The alternative hypothesis about expectations formation, advanced by Muth (1961), is that expectations are rational, i.e., they are the same as the predictions made by the relevant macroeconomic model that econometricians wish to estimate.

In the past several years, various methods for estimating behavioral equations that explicitly involve rational expectations have been proposed in various contexts (Sargent (1973, 1976), Abel (1977), McCallum (1976 a,b), Fair (1979), Taylor (1979), Wallis (1980), Hansen and Sargent (1980)). Some of them require a very strong set of assumptions for the consistency of parameter estimates. Some of them do not give appropriate measures of the sampling error of parameter estimates. The purpose of this survey paper is to place them in a common perspective and indicate their logical extensions. More specifically, we consider two general classes of simultaneous equations models that explicitly involve rational expectations and show how the available estimation methods can be applied. In so doing we can compare the available methods and their appropriate modifications in terms of their general applicability, computational ease, and practical feasibility. The paper should serve as a complement to Schiller (1978) which emphasizes the analytical side of rational expectations models.

The organization of the paper is as follows. Section 2 formally states a simultaneous equations model that explicitly involves expectations and shows how it can be solved for the expectational variables under rational expectations. The relevant information variable in the model is the expectation of current endogenous variables conditional on information available at the end of the previous period. Section 3 discusses three methods to estimate a single equation in the model. Two of them require that the policy rule -- a rule according to which the value of policy instruments are determined -- do not change in the sample period. Section 4 presents three methods to estimate the model as a whole. They can be thought of as the system estimation counterparts of the single-equation estimation methods in section 3. In section 5, we indicate what modifications are necessary in order for the preceding estimation methods to be applicable to a model whose expectational variables are replaced by the expectation of future endogenous variables conditional on current information. Section 6 examines the estimation of behavioral equations that involve the present discounted value of future rational expectations. This is an important topic because most behavioral equations derived from intertemporal optimization considerations invariably involve the present discounted value. Until very recently it was impossible except under special circumstances to obtain a consistent estimate of the asymptotic variance of parameter estimates when the disturbance term is autocorrelated. Section 7 presents recently proposed methods for correcting autocorrelation. Section 8 contains concluding remarks.

2. The Model and Its Rational Expectations Solutions

The simultaneous equations model we are interested in estimating is

$$C(L)\ Y_t + A\ {}_{t-1}Y_t + B\ X_t = u_t,\ (t=1,\ldots,\ T) \qquad (2.1)$$

where Y_t = vector of endogenous variables (size: g x 1)

${}_{t-1}Y_t$ = expectation as of t-1 of Y_t

X_t = vector of forcing variables (size: k x 1)

u_t = behavioral error[1] (size g x 1)

$C(L) = C_0 + C_1L + \ldots + C_rL^r$, polinomial in the lag operator L of degree r.

The matrices A,B, $C(=(C_0,\ C_1,\ \ldots,\ C_r))$ are the parameters of the structural equations. Exclusion restrictions are already imbedded in these matrices so that some of their elements are a priori set equal to 0. The only difference between our model (2.1) and the conventional model of simultaneous equations is that our model involves the expectational variables ${}_{t-1}Y_t$. Therefore the structural parameters exclusively represent the elements of economic behavior that are unaffected by a change in expectations.[2] Let I_t be the set of information consisting of $(X_t,\ X_{t-1},\ \ldots,\ Y_t,\ Y_{t-1},\ \ldots,\ u_t,\ u_{t-1},\ \ldots)$. The rational expectations hypothesis states that ${}_{t-1}Y_t$ is the mathematical expectation of Y_t conditional on I_{t-1}:

$${}_{t-1}Y_t = E(Y_t|I_{t-1}). \qquad (2.2)$$

The form of the conditional expectation operator $E(.|I_{t-1})$ depends on: the true value of the structural parameters (A,B,C), the stochastic specification of the behavioral error u_t, and the stochastic process of the X_t process. In the present model, economic agents do not know the realized value of current variables $(X_t,\ Y_t,\ u_t)$ and their behavior has to be based on their perception

${}_{t-1}Y_t$ about the current endogenous variables.[3] An example of this type of model is Sargent's (1976) classical macroeconomic model of the U.S. which embraces the Lucas supply function (Lucas (1973)) as the central feature.

To complete the system for (X_t, Y_t), the stochastic process for X_t has to be coupled with (2.1). We write it as

$$X_t = G(X_{t-1}, X_{t-2}, \ldots, Y_{t-1}, Y_{t-2}, \ldots; \theta_t) + \varepsilon_t , \qquad (2.3)$$

where θ_t is an exogenous shift variable for the function G, and ε_t is a white noise process that is independent of the u_t process. Some of the forcing variables may be purely exogenous, i.e., independent of the u_t process. The other elements of X_t are policy instruments. The function G for such elements in X_t are called the <u>policy rule</u>. A change in the shift variable θ_t implies a shift in the policy rule. We assume that the stochastic process (X_t, Y_t) generated by $(u_t, \varepsilon_t, \theta_t)$ through (2.1) and (2.3) is stationary.

Many of the estimation methods to be presented in this paper require that the function G be explicitly specified. The particular stochastic process for X_t we will consider is

$$X_t = F(L)X_t + H(L)Y_t + \varepsilon_t \qquad (2.4)$$

where F(L) and G(L) are polinomials in the lag operator L of degrees p and q, respectively:

$$F(L) = F_1L + F_2L^2 + \ldots + F_pL^p ,$$

$$H(L) = H_1L + H_2L^2 + \ldots + H_qL^q .$$

If X_{it} is purely exogenous, H(L) in the i-th equation is zero. For convenience we will refer to (2.3) or (2.4) as the policy rule. (2.4) is perhaps the most popular policy rule in the literature.[4] The assumption here is that F_i and H_i do not change their values in the sample period. In this respect the estimation methods based on (2.4) is not free from the Lucas (1976) critique.

Having laid out the model, we now seek to express the rational expectation ${}_{t-1}Y_t$ for the model consisting of (2.1) and (2.4) in terms of the past observable variables (X_{t-1}, X_{t-2}, ..., Y_{t-1}, Y_{t-2}, ...). Applying the conditional expectation operator to (2.1) and solving for ${}_{t-1}Y_t$, we obtain

$$
\begin{aligned}
{}_{t-1}Y_t = -(C_0 + A)^{-1} B\, {}_{t-1}X_t - \sum_{i=1}^{r}(C_0 + A)^{-1} C_i Y_{t-i} \\
+ (C_0 + A)^{-1}\, {}_{t-1}u_t \, , \qquad (2.5)
\end{aligned}
$$

where ${}_{t-1}u_t = E(u_t | I_{t-1})$ and ${}_{t-1}X_t = E(X_t |_{t-1})$. (2.5) involves unobservable expectational variable ${}_{t-1}X_t$. Applying the conditional expectation operator to (2.4) gives

$$
{}_{t-1}X_t = F(L)X_t + H(L)Y_t \, . \qquad (2.6)
$$

Substituting (2.6) into (2.5) we obtain

$$
{}_{t-1}Y_t = M_X(L)X_t + M_Y(L)Y_t + N\, {}_{t-1}u_t, \qquad (2.7)
$$

or

$$
Y_t = M_X(L)X_t + M_Y(L)Y_t + N\, {}_{t-1}u_t + e_t \, , \qquad (2.8)
$$

where

$$
\begin{aligned}
M_X(L) &= M_{X1}L + \ldots + M_{Xp}L^p \, , \\
M_Y(L) &= M_{Y1}L + \ldots + M_{Ys}L^s, \; s = \max(q, r), \\
M_{Xi} &= -(C_0 + A)^{-1} BF_i \quad (i = 1, \ldots, p), \qquad (2.9)
\end{aligned}
$$

$$M_{Yi} = -(C_0 + A)^{-1} (BH_i + C_i) \quad (i = 1, \ldots, s) \tag{2.10}$$

$$N = (C_0 + A)^{-1} \tag{2.11}$$

$$e_t = Y_t - {}_{t-1}Y_t \ .$$

e_t, the difference between the realized value and its expected value, is called the <u>forecast error</u>. The hypothesis of rational expectations (2.2) implies that e_t is orthagonal to I_{t-1}, i.e., $E(e_t | I_{t-1}) = 0$, which in turn implies that e_t is not autocorrelated. We will call (2.7) or (2.8) the <u>forecasting equation</u> because it does not involve unobservable expectations about X_t or Y_t on the right side. We here note that the coefficients in the forecasting equation depend on the policy rule (2.4).[5] The equation will be used to obtain a consistent estimate of the rational expectations ${}_{t-1}Y_t$.

3. Single-Equation Estimation Methods

3.1. Introduction

In this section, we discuss three methods for estimating a single equation in the system (2.1) without specifying the rest of the system. Suppose we are interested in estimating the first equation of (2.1):[6]

$$y_t = \gamma Y_{1t} + \alpha\ {}_{t-1}Y_t^1 + \beta X_{1t} + u_{1t} \ , \tag{3.1}$$

where y_t = the first element of Y_t ,

Y_{1t} = vector of endogenous variables that appear in the first equation,

Y_t^1 = vector of endogenous variables whose expected values appear in the equation,

${}_{t-1}Y_t^1 = E(Y_t^1 | I_{t-1})$,

X_{1t} = vector of forcing variables in the first equation,

u_{1t} = behavioral error in the equation.

Y_t^1 and Y_{1t} may share common elements of Y_t. The estimation methods to be presented in this section can be grouped into two classes. The first class of estimation methods assumes that the forcing variables X_t are generated by the fixed policy rule (2.4). The second class does not specify the policy rule and leaves it as general as (2.3). The crucial step in either class of the estimation methods is to obtain a proxy for the rational expectation ${}_{t-1}Y_t$ without directly estimating the rest of the system. The first class utilizes the forecasting equation to obtain a consistent proxy for the rational expectation and the second class takes the realized value Y_t^1 as the proxy.

3.2 Estimation with Fixed Policy Rule (2.4)

Write the forecasting equation for Y^1_t, a subset of Y_t, as

$$Y^1_t = M^1_X(L)X_t + M^1_Y(L)Y_t + N^1\, {}_{t-1}u_t + e^1_t, \quad (3.2)$$

where $e^1_t = Y^1_t - {}_{t-1}Y^1_t$ is the forecast error and is orthogonal to I_{t-1}. We recall that the orders of polinomials M^1_X and M^1_Y are p and s, respectively, where s = max(q, r) and r is the order of C(L) in (2.1).

First consider the simplest case where the vector u_t is not autocorrelated. Then we have ${}_{t-1}u_t = 0$, and we can obtain a consistent estimate $\hat{Y}^1_t$ of the rational expectation ${}_{t-1}Y^1_t$ by regressing Y^1_t on $X_{t-1}, \ldots, X_{t-p}$, $Y_{t-1}, \ldots, Y_{t-s}$ and taking the fitted value. The consistency of $\hat{Y}^1_t$ is guaranteed by the fact that e^1_t in (3.2) is orthogonal to all the right side variables. This is the first step of estimation. The second step is to replace ${}_{t-1}Y^1_t$ in (3.1) by $\hat{Y}^1_t$ and estimate (3.1) by 2LS (two-stage least squares) or by IV (instrumental variable estimation technique) treating X_{1t} and $\hat{Y}^1_t$ as predetermined.[7] This yields a consistent estimate of the parameters (α, β, γ) since ${}_{t-1}Y^1_t$ is uncorrelated with u_{1t} and $\hat{Y}^1_t$ is a consistent proxy for ${}_{t-1}Y^1_t$.

This two-step estimation method was proposed by Sargent (1973, 1976) and Wallis (1980).[8] Its most impractical aspect is that the regression in the first step involves enormous number of the right side variables: p lagged values of all the forcing variables and s lagged values of all the endogenous variables must be included to ensure the consistency of the proxy $\hat{Y}^1_t$.[9] The degree of freedom for the regression will be quickly exhausted as the size of the model (2.1) gets large. The difficulty becomes absolutely insurmountable if <u>some</u> of the elements of the vector u_t are autocorrelated. This is because the term ${}_{t-1}u_t$ in (3.2) now depends on lagged values of u_t which in turn depend on lagged values of $(X_t, Y_t, {}_{t-1}Y_t)$ through (2.1). Thus ${}_{t-1}Y_t^1$ in general depends on infinite lags of (X_t, Y_t), making the consistent estimation of ${}_{t-1}Y_t^1$ impossible.

Another problem in the two-step method is that it cannot give appropriate standard errors for the parameter estimates in the second stage.[10] This is because the second stage ignores the uncertainty associated with the fact that $\hat{Y}_t^1$ is estimated. To correctly evaluate the true sampling error, we have to carry out the two steps simultaneously, i.e., we have to estimate the following system <u>jointly</u>:

$$y_t = \gamma Y_{1t} + \alpha[M_X^1(L)X_t + M_Y^1(L)Y_t] + \beta X_{1t} + u_{1t}, \tag{3.3}$$

$$Y_t^1 = M_X^1(L)X_t + M_Y^1(L)Y_t + e_t^1, \tag{3.4}$$

Note that this system (3.3) and (3.4) involves a nonlinear constraint that M_X^1 and M_Y^1 appear across equations. The appropriate estimation technique is NL3LS (nonlinear three-stage least squares) with $(X_t, X_{t-1}, \ldots, X_{t-p}, Y_{t-1}, \ldots, Y_{t-s})$ as instruments.[11] Of course, this joint estimation method is not applicable when u_t is autocorrelated and has the same degrees of freedom problem. As we will see in the next section, it is in principle possible to write down the likelihood function associated with the system consisting of (2.1) and (2.4). It would, however, be very difficult if not impossible to derive the concentrated likelihood function in terms of $(\alpha, \beta, \gamma, M_X^1, M_Y^1)$ and carry out a version of the limited-information maximum likelihood estimation. One of the sources of the difficulty is that the forecast error e_t^1 can be correlated with current and future values of any variables including the purely exogenous variables.

3.3. Estimation without Fixed Policy Rule

In the above two estimation methods, the fixed linear policy rule was necessary only for the purpose of obtaining a consistent proxy for the rational expectation. The third method we now present does not need the assumption of fixed policy rule (2.4) because it does not need to have a consistent proxy. The proxy it uses is the realized value Y_t^1 of the rational expectation ${}_{t-1}Y_t^1$. This is the method proposed by McCallum (1976a,b). Substituting the formula $e_t^1 = Y_t^1 - {}_{t-1}Y_t^1$ into (3.1), we obtain

$$y_t = \gamma Y_{1t} + \alpha Y_t^1 + \beta X_{1t} + v_{1t} , \tag{3.5}$$

where $v_{1t} = u_{1t} - \alpha e_t^1$. The disturbance term v_{1t} consists of the behavioral error and the forecast error e_t^1. (3.5) resembles a typical errors-in-variables equation in that Y_t^1 and v_{1t} are correlated because of the fact that Y_t^1 measures ${}_{t-1}Y_t^1$ with some error. The feature in (3.5) that is not present in the conventional errors-in-variables model is that all the right side variables -- including the ones that are uncorrelated with the behavioral error u_{1t} -- can be correlated with the disturbance term. This is because the forecast error e_t^1 can be correlated with any elements in I_t that are not in I_{t-1}.

This last feature makes consistent estimation more difficult but not impossible. Consistent estimation is achieved if we can find a (column) vector Z_{t-1} in I_{t-1}, to be used as instruments for (Y_{1t}, Y_t^1, X_{1t}), that satisfy the following two conditions:

(1) uncorrelatedness: $E(u_{1t} Z_{t-1}) = 0$,

(2) ergodicity: (X_t, Y_t) is ergodic. (Note that Z_{t-1} is a subset of $(X_{t-1}, X_{t-2}, \ldots, Y_{t-1}, Y_{t-2}, \ldots)$).

The orgodicity condition amounts to the usual regularity condition for instruments that the sample moments matrix for Z_{t-1} and the right side variables to be instrumented converges in probability to a nonsingular matrix as the sample size increases.[12] If the orthogonality of the behavioral error u_{1t} to Z_{t-1} is satisfied, then the whole disturbance term u_{1t} is orthogonal to Z_{t-1} because the forecast error e_t^1 is orthogonal to I_{t-1} which includes Z_{t-1}. Under these conditions, an instrumental variable estimation technique using Z_{t-1} as instruments for Y_{1t}, Y_t^1 and X_{1t} yields a consistent and asymptotically normal estimate of (α, β, γ). See Hansen (1980) for a rigorous proof of this.

If u_{1t} is not autocorrelated, the choice of instruments is quite straightforward: any elements of I_{t-1} can constitute Z_{t-1}. Furthermore, the standard errors routinely calculated by a standard program package are appropriate ones because none of the right side variables are estimated proxies and because the disturbance term v_{1t} is not autocorrelated.[13] If u_{1t} is autocorrelated, the estimation becomes less straightforward in two respects. First, the choice of instruments is now a subtle one. Lagged values of the endogenous variables and policy instruments may not be orthogonal to u_{1t} because of their dependence on u_{t-1}. The set of instruments must be composed of lagged values of purely exogenous variables. Second, the disturbance term is now autocorrelated and hence the standard errors routinely calculated are biased.

As Flood and Garber (1980) have shown, the conventional way of correcting autocorrelation (like the generalized least squares) introduces biases into the parameter estimates. We will discuss in section 7 how to correct autocorrelation without destroying the consistency.

There is one problem in the present estimation method. That is that, if Y_{1it} and Y^1_{jt} refer to the same variable, only $\gamma_i + \alpha_j$ is identifiable and γ_i and α_j cannot be identified separately.[14] This implies, in particular, that we cannot estimate the Lucas (1973) supply function:

$$GNP_t = a_0 + a_1(p_t - {}_{t-1}p_t) + u_{1t} \; . \tag{3.6}$$

If McCallum's idea of using p_t as a proxy for ${}_{t-1}p_t$ is used, the last two terms on the right side of (3.6) is treated as the disturbance term.

3.4. Comparing the Three Methods

We have discussed three estimation methods -- the two-step method, the joint estimation method, and McCallum's instrumental variable method. The third method has three advantages over the first two. (1) It does not require the listing of all the endogenous and forcing variables. This is quite a desirable feature when we are interested in only one or a subset of the equations in the model. (2) It gives consistent estimates even if the behavioral error is autocorrelated. (3) It is applicable to not only the case where the policy rule is linear and fixed but also the case where the (linear or nonlinear) policy rule shifts in the sample period. It thus seems that the case for the third method is very strong, as long as Y_{1t} and Y^1_t share no common elements and hence both γ and α are identifiable. The only remaining issue here is how much more efficient the second method is relative to the third method when the assumption of nonautocorrelated u_t and fixed policy rule is satisfied.[15] It might seem that the extra variance introduced by the forecast error e_t^1 into the disturbance v_{1t} makes the third method inefficient, but that extra variance represents the uncertainty associated with the fact that the second method has to estimate the parameters of the forecasting equations (3.4). The real source of inefficiency of the third method is that it cannot exploit the orthogonality of u_{1t} to <u>current</u> exogenous variables X_t. The following simple example highlights the point.

Suppose (2.1) is written as

$$y_{1t} = \alpha_{t-1}y_{2t} + u_{1t} , \qquad (3.7)$$

$$y_{2t} = \beta x_t + u_{2t} \qquad (3.8)$$

and suppose x_t is generated by a first-order autoregressive process:

$$x_t = \rho x_{t-1} + \varepsilon_t \qquad (3.9)$$

The forecasting equation for y_{2t} is

$$y_{2t} = \gamma x_{t-1} + e_t , \qquad (3.10)$$

where $\gamma = \beta\rho$ and $e_t = u_{2t} + \beta\varepsilon_t$. We assume that u_{1t}, u_{2t}, and ε_t are nonautocorrelated and stationary, and that (u_{1t}, u_{2t}) is uncorrelated with ε_t . Let $\hat{\gamma}$ be the OLS estimate of γ obtained from (3.10) and write $\hat{y}_{2t} = \hat{\gamma} x_{t-1}$. The estimate $\hat{\alpha}$ of α given by the two-step method is

$$\hat{\alpha} = \frac{\Sigma\, y_{1t}\hat{y}_{2t}}{\Sigma\, \hat{y}_{2t}^{2}} . \tag{3.11}$$

It is easy to see that $\hat{\alpha}$ is consistent and its asymptotic variance is

$$\frac{\sigma_1^{2} - 2\alpha\sigma_{12} + \alpha^{2}\sigma_e^{2}}{\gamma^{2}\sigma_x^{2}} , \tag{3.12}$$

where $\sigma_{12} = \mathrm{Cov}(u_{1t}, e_t)$, $\sigma_e^2 = \mathrm{Var}(e_t)$, $\sigma_1^2 = \mathrm{Var}(u_{1t})$ and $\sigma_x^2 = \mathrm{Var}(x_t)$. However, the standard error of α routinely calculated by any OLS package is

$$\frac{T^{-1}\, \Sigma\, (y_{1t} - \hat{\alpha}\hat{\gamma}\, x_{t-1})^{2}}{\hat{\gamma}^{2}\, \Sigma\, x_{t-j}^{2}} . \tag{3.13}$$

T times (3.13) converges in probability to

$$\frac{\sigma_1^{2}}{\gamma^{2}\sigma_x^{2}} \tag{3.14}$$

which is smaller than (3.12). The joint estimation method estimates

$$y_{1t} = \alpha\gamma x_{t-1} + u_{1t} \tag{3.15}$$

and (3.10) jointly. It can be shown that the asymptotic variance of the estimate of α given by this method is equal to (3.12). The third method estimates

$$y_{1t} = \alpha y_{2t} + (u_{1t} - \alpha e_t) . \tag{3.16}$$

If x_{t-1} is used as the instrument for y_{2t}, then the asymptotic variance of the resulting estimate of α is again equal to (3.12). The third method is asymptotically equivalent to the second method in this example because the exogenous variable x_t does not appear in (3.7).

4. System Estimation Methods

4.1. Introduction

This section presents three estimation methods to estimate the system (2.1) as a whole. Apart from a priori constraints on the structural parameters (A,B,C), there are two sources of efficiency gain of the system estimation methods over the single-equation methods. The first is, of course, the contemporaneous covariance among the behavioral errors. The second is the rational expectations constraint that ${}_{t-1}Y_t$ is the predictions made on the basis of the whole system. If (2.4) is the policy rule, then the rational expectations constraint is that M_X and M_Y in the forecasting equation are written as (2.9) and (2.10). All the system estimation methods presented in this section exploit the first source of efficiency gain.

4.2. Estimation with Fixed Policy Rule (2.4)

The simplest estimation method under the assumption that (2.4) is the policy rule is the three-stage least squares (3LS) analogue of the two-step method presented in 3.2. It therefore estimates

$$C(L)Y_t + A\hat{Y}_t + BX_t = u_t \qquad (4.1)$$

by 3LS treating $\hat{Y}_t$ and X_t as predetermined, where $\hat{Y}_t$ is the consistent proxy of ${}_{t-1}Y_t$ given by the forecasting equation (2.8). The assumption that u_t is not autocorrelated is necessary for the consistency of the resulting parameter estimates, because otherwise $\hat{Y}_t$ is not a consistent proxy, as we have seen in 3.2. Note that the rational expectations constraint (2.9) and (2.10) is not imposed. This method does not give appropriate standard errors because the estimated value $\hat{Y}_t$ is used in the 3LS estimation.

The most efficient estimation method that assumes the fixed policy rule (2.4) is the maximum likelihood method proposed by Wallis (1980). Substituting the forecasting equation (2.7) into (2.1), we obtain

$$C(L)Y_t + A[M_X(L)X_t + M_Y(L)Y_t + N_{t-1}u_t] + BX_t = u_t \ . \qquad (4.2)$$

The system consisting of (4.2) and (2.4) can be estimated by the maximum likelihood method assuming (u_t, ε_t) is jointly normal, with the rational expectations constraint (2.9) - (2.11) and the constraint that u_t is independent of ε_t. This can handle the case u_t is autocorrelated.

This method requires enormous amount of computation because of the highly nonlinear nature of the rational expectations constraint. Some iterative procedure that is asymptotically equivalent to the maximum likelihood method is necessary for practical applications. The procedure proposed by Fair (1979) iterates on the series for the rational expectation ${}_{t-1}Y_t$. For a given series for ${}_{t-1}Y_t$ (t=1,2, ..., T), we can estimate (A,B,C,F,H) using 3LS on (2.1) and the multivariate regression on (2.4).[16] We then compute ${}_{t-1}Y_t$ from (2.7), (2.9)-(2.11) using the estimated parameter value. This new series for ${}_{t-1}Y_t$ is then used to obtain a new set of parameter values, and so on. Obviously, this idea of iterating on expectations becomes infeasible as the sample size increases. An alternative iteration procedure is to iterate on parameters (A,B,C,F,H). If it converges, this iteration procedure will yield parameter estimates asymptotically equivalent to the maximum likelihood estimates.[17]

We have described two system estimation methods, the simplest one and the most efficient one. We can easily think of several methods that lie in between the two polar cases. We do not discuss them here because they should be pretty obvious by now.

4.3. System Estimation without Fixed Policy Rule

McCallum's idea of using the realized value as a proxy for the rational expectations can be easily extended to system estimation. The following estimation method is based on the idea suggested by Wickens (1977). We assume u_t is not autocorrelated. Substituting $Y_t = {}_{t-1}Y_t + e_t$ into (2.1) we obtain

$$(C(L) + A)\ Y_t + BX_t = v_t, \tag{4.3}$$

where $v_t = u_t - Ae_t$. As we noted in 3.3, C_0 and A cannot be identified separately. Since e_t is not autocorrelated, the disturbance term v_t is not autocorrelated. The variance-covariance matrix of v_t is

$$\mathrm{Var}(v_t) = \mathrm{Var}(u_t) + A\ \mathrm{Var}(e_t)A' - 2A\ \mathrm{Cov}(e_t, u_t). \tag{4.4}$$

The stationarity of (X_t, Y_t) implies that e_t is stationary. So Var (e_t) is constant over time. The first step in the estimation is to apply the single-equation method in 3.3 to each of the equations in (4.3) and obtain a consistent estimate of $v_t (t = 1,2, \ldots, T)$. This gives a consistent estimate S of $\mathrm{Var}(v_t)$. The second step is to apply FIVE (full-information instrumental variable technique) to (4.3).[18] That is, we minimize the quadratic form $v'(S^{-1} \otimes Z(Z'Z)^{-1}Z')v$ with respect to (A,B,C), where $v = (v_{11}, \ldots, v_{1T}, v_{21}, \ldots, v_{2T}, \ldots, v_{g1}, \ldots, v_{gT})$ is the estimated vector of the disturbance and $Z' = (Z_0, \ldots, Z_{T-1})$ is the matrix of instruments. Z_{t-1} should not include X_t because X_t can be correlated with e_t.[19]

Efficiency gain of this method over the single equation method in 3.3 could be substantial because of the extra variance term in (4.4). The maximum likelihood estimation cannot be used on (4.3) because, as we have seen in 3.2, the forecast error e_t can be correlated with X_{t+j} $(j \geq 0)$ and its covariance depends not only on the structural parameters (A,B,C) but also on the policy rule which is left unspecified here.

If u_t is autocorrelated, this method still yields a consistent estimate of (A,B,C) with appropriately chosen instruments, but does not give appropriate standard errors. A correct treatment of autocorrelation for this situation will be given at the end of section 7.

5. The Perfect Information Model

In the model we dealt with in the previous sections, the relevant expectational variable is the expectation of current endogenous variables conditional on information available in the preceding period. For ease of reference, we will henceforth call this model the imperfect information model. The alternative model, equally popular in the literature, posits that the relevant expectational variable is the expectation of future endogenous variables conditional on information available in the current period. We will call it the perfect information model. It is written formally as

$$C(L)Y_t + A\ {}_tY_{t+1} + BX_t = u_t \tag{5.1}$$

with
$$ {}_tY_{t+1} = E(Y_{t+1}|Y_t) \ , \tag{5.2}$$

where $I_t = \{X_t, X_{t-1}, \ldots, Y_t, Y_{t-1}, \ldots, u_t, u_{t-1}, \ldots, \varepsilon_t, \varepsilon_{t-1}, \ldots\}$.

The only difference between (2.1) and (5.1) is that ${}_{t-1}Y_t$ is now replaced by ${}_tY_{t+1}$ here. The X_t process is still written as (2.4), but it may be natural to allow policy instruments to respond to current endogenous and exogenous variables. Thus F_0 in $F(L)$ and H_0 in $H(L)$ may not be zero.

It is harder in the perfect information model than in the imperfect information model to derive a forecasting equation in a form suitable for the consistent estimation of the rational expectation. Under the assumption of stationality, it is possible to express the rational expectation ${}_tY_{t+1}$ as:

$$ {}_tY_{t+1} = M_Y(L)Y_t + M_X(L)X_t + N \sum_{k=0}^{\infty} P^k\ {}_tu_{t+k+1}, \tag{5.3}$$

where $M_Y(L)$ and $M_X(L)$ are polinomials in the lag operator L. The first two terms in the right side of (5.3) are distributed lags of Y_t and X_t that includes <u>current</u> values of (X_t, Y_t). The degree of $M_Y(L)$ is the maximum of the degrees of C(L) and H(L) minus one, and the degree of $M_X(L)$ is equal to that of F(L) minus one. It should be noted that (5.3) does not involve the current value of u_t which is included in the information set. The rational expectations hypothesis implies that $(M_Y(L), M_X(L), N, P)$ are expressed in terms of (A,B,C,F,H). The expression can be obtained as follows: substitute (5.3) into (5.1), lead the resulting equations and (2.4) one period, take the expectations conditional on I_t, compare the resulting equation for ${}_tY_{t+1}$ with (5.3). We do not write the expression for $(M_Y(L), M_X(L), N,P)$ because it is extremely complicated.

The estimation method analogous to the maximum likelihood method in 4.2 would be to substitute (5.3) into (5.1) and maximize the likelihood function for (5.1) and (2.4) subject to the rational expectations constraint that $(M_Y(L), M_X(L), N,P)$ are functions of (A,B,C,F,H). This method is likely to be infeasible in practical applications because of the complexity of the constraint. If the behavioral error u_t is not autocorrelated, a method analogous to the first method in 4.2 is available. A consistent proxy $\hat{Y}_{t+1}$ for ${}_tY_{t+1}$ can be obtained from the regression of Y_{t+1} on current and lagged values of <u>all</u> the endogenous and exogenous variables (X_t, Y_t), since the last term in the forecasting equation (5.3) is zero and since the forecast error $e_{t+1} = Y_{t+1} - {}_tY_{t+1}$ is orthogonal to I_t. The lag length in the regression should be long enough to cover the degrees of the polinomials in (5.3). (5.1) can be estimated consistently by 3LS using $\hat{Y}_{t+1}$ as a proxy for ${}_tY_{t+1}$. $\hat{Y}_{t+1}$ must be treated as endogenous because it can be correlated with u_t. Since the estimated value $\hat{Y}_{t+1}$ is used in the 3LS estimation, the standard errors routinely computed are not consistent estimates of the asymptotic variance of parameter estimates. Likewise, application of the two single-equation methods in 3.2 to the perfect information model should be straightforward. So much for the estimation methods with fixed policy rule.

McCallums' idea of using the realized value as a proxy for rational expectations can easily be applied to the perfect information model. Suppose we want to estimate the first euqation of (5.1):[20]

$$y_t = \gamma Y_{1t} + \alpha_t Y^1_{t+1} + \beta X_{1t} + u_{1t}, \tag{5.4}$$

where $Y^1_{t+1} = E(Y^1_{t+1}|I_t)$. Using $e^1_{t+1} = Y^1_{t+1} - Y^1_{t+1}$ we can rewrite (5.4) as

$$y_t = \gamma Y_{it} + \alpha Y^1_{t+1} + \beta X_{1t} + v_{1t} \tag{5.5}$$

where $v_{1t} = u_{it} - \alpha e^1_{t+1}$ is the disturbance term. Let Z_t be a set of variables in I_t that are uncorrelated with the behavioral error u_{1t}. Z_t serves a valid set of instruments since Z_t, being in I_t, is uncorrelated with the forecast error e^1_{t+1}.[21] Current value of purely exogenous variables are obvious candidates for Z_t. This errors-in-variables method has two features that were not present when it was applied to the imperfect information model. First both γ and α can be identified even if Y_{1t} and Y^1_t share common elements. Second, the uncorrelatedness of purely exogenous variables with u_{1t} can now be exploited. The disadvantage of the method is that the disturbance term is autocorrelated even if the behavioral error u_{1t} is not. This is due to a possible correlation between e^1_{t+1} and u_{1t+1}. We will discuss the correction of autocorrelation in section 7.

6. The Present Discounted Value of Rational Expectations

So far, we have assumed that the expectation horizon - the number of periods extending from the current period over which expectations are defined - is one. The estimation methods with fixed policy rule can be extended to the case in which the expectation horizon is finite. The key step in the estimation is to obtain the forecasting equation for future rational expectations (${}_{t}Y_{t+j}$ or ${}_{t-1}Y_{t+j}$), which can be done by solving a difference equation whose order is equal to the expectations horizon. The rational expectations constraint which relates the structural parameters to the forecasting equation will be quite complicated. We can also extend McCallum's errors-in-variables method (which does not assume fixed policy rule) to the finite horizon case by simply replacing future rational expectations by their realized values.

However, neither of the two methods does not have a straightforward extension to the infinite horizon case in which the relevent expectational variable is the present discounted value of rational expectations extending to the infinite future. The infinite horizon case undoubtedly deserves a separate discussion because most behavioral equations that are derived from intertemporal optimization involve the present discounted value of future rational expectations. Prime examples of such behavioral equations are: Friedman's (1957) permanent income hypothesis, Lucas and Rapping's (1969) labor supply equation, Tobin's q-type investment function (Lucas and Prescott (1971), Abel (1977)), and Hansen and Sargent's (1980) labor demand equation. This section discusses the estimation methods for the infinite horizon case. The entire discussion will be carried out in terms of the simple example considered by Lucas and Prescott (1971) and further elaborated by Sargent (1979). This makes the discussion simple and concrete, while bringing up the essential issues that will always show up in more complicated models.

Consider a representable firm seeking to maximize

$$E_t \sum_{j=o}^{\infty} \lambda^j [p_{t+j}(\alpha\, k_{t+j}) - b_{t+j}(k_{t+j}-k_{t+j-1}) - \frac{\beta}{2}\, (k_{t+j}-k_{t+j-1}-u_t)] ,$$

where $E_t x = E(x|I_t)$, p_t = price of the firm's output, b_t = price of investment goods, and k_t = capital stock. The information set I_t includes

$\{k_t, k_{t-1},\ldots,p_t,p_{t-1},\ldots, \quad b_t,b_{t-1},\ldots,u_t,u_{t-1},\ldots\}$.

The discount factor λ obeys $0 < \lambda < 1$. The firm is competitive in the sense that the stochastic process for (p_t,b_t) is independent of the firm's action. The quadratic term in (6.1) represents the costs of adjustment associated with investment $(k_{t+j}-k_{t+j-1})$. The term u_{t+j} is a random shock to the adjustment cost. u_{t+j} is seen by the firm in period $t + j$ but unobserved by the econometrician. The first order condition for optimality (i.e., the Euler equation) is

$$\alpha p_{t+j} - b_{t+j} - \beta y_{t+j} + \lambda E_{t+j}(b_{t+j+1}+\beta y_{t+j+1}) = \lambda\beta\, {}_{t+j}u_{t+j+1} - \beta u_{t+j}. \quad (6.2)$$

where ${}_{t+j}u_{t+j+1} = E(u_{t+j+1}|I_{t+j})$ and $y_{t+j} = k_{t+j} - k_{t+j-1}$. The solution to (6.2) is[22]

$$y_t = - \frac{1}{\beta}\, b_t + \frac{\alpha}{\beta}\, q_t + u_t, \quad (6.3)$$

where

$$q_t = \sum_{j=o}^{\infty} \lambda^j\, E(p_{t+j}|I_t) \quad (6.4)$$

is the present discounted value of rational expectations. The question is: how to deal with q_t?

The solution offered by Hansen and Sargent (1980) is the following. Assume (p_t,b_t) are the first two elements of an n-dimensional vector autoregressive process x_t that satisfies

$$B(L)x_t = \eta_t, \tag{6.5}$$

where $B(L) = I - B_1L - \ldots - B_rL^r$. The information set I_t includes $\{x_t, x_{t-1}, \ldots\}$, and any variable that Granger cause (p_t, b_t) must be included in x_t. Under (6.5), q_t can be written as

$$q_t = UB(\lambda)^{-1}[I + \sum_{j=1}^{r-1} (\sum_{i=j+1}^{r} \lambda^{i-j} B_i)L^j]x_t, \tag{6.6}$$

where U is the n-dimensional vector with 1 in the first place (which corresponds to p_t) and zeroes elsewhere. Equation (6.6) corresponds to the forecasting equation (3.2). The parameters $(\alpha, \beta, \lambda, B_1, \ldots, B_r)$ can be estimated by the method of maximum likelihood for (y_t, x_t) with q_t written as (6.6).[23]

This Hansen-Sargent method shares the same difficulties that the joint estimation method in 3.2 has. First, the stochastic process (6.5) may change if the policy rule changes.[24] Second, the list of variables in x_t that Granger cause (p_t, b_t) may be a large one. This will be a serious problem if b_t is actually a vector of many variables.[25] Third, the rational expectations constraint -- the constraint that B_i's appear both in (6.3) through (6.6) and in (6.5) -- implies a lot of computation in the maximization of the likelihood function.

An alternative estimation method, which does not assume a fixed stochastic process like (6.5), is available. The method is based on the works by Abel (1977), Kennan(1979), and Hayashi (1979). The central idea is to use the Euler equation (6.2) instead of the behavioral equation (6.3) to estimate α, β and λ. Given the behavioral equation (6.3), we can recover the Euler equation by noting that q_t defined by (6.4) follows the following stochastic difference equation:

$$q_t - \lambda q_{t+1} = p_t - e_{t+1} \tag{6.7}$$

where

$$e_{t+1} = \sum_{j=1}^{\infty} \lambda^j ({}_{t+1}p_{t+j} - {}_t p_{t+j}). \tag{6.8}$$

e_{t+1} represents the revision of expectations about p_{t+j} that will be made on the arrival of new information as period t + 1 rolls around. The rational expectations hypothesis implies that e_{t+1} is orthogonal to I_t, i.e., $E(e_{t+1}|I_t) = 0$. From (6.3) and (6.7) we can derive the following equation:

$$y_{t+1} = \frac{1}{\lambda} y_t + \frac{1}{\lambda\beta} b_t - \frac{1}{\beta} b_{t+1} - \frac{\alpha}{\lambda\beta} p_t + v_t , \tag{6.9}$$

where

$$v_t = u_{t+1} - \frac{1}{\lambda} u_t + \frac{\alpha}{\lambda\beta} e_{t+1} . \tag{6.10}$$

Equation (6.9) is equivalent with the Euler equation with $e_{t+1} = (y_{t+1} - {}_t y_{t+1}) + \beta^{-1}(b_{t+1} - {}_t b_{t+1}) - (u_{t+1} - {}_t u_{t+1})$. We can estimate (α,β,λ) consistently by the instrumental variable technique. Any variable Z_t in I_t that are uncorrelated with u_t and u_{t+1} are valid instruments because e_{t+1} is orthogonal to Z_t. However, a simple application of this instrumental variable method to (6.9) does not yield a consistent estimate of the asymptotic variance of parameter estimates. This is because the disturbance term v_t is autocorrelated even if u_t is not. We will discuss the correction of autocorrelation in the next section.

7. Correcting Autocorrelation[26]

7.1 Introduction

In the previous sections, we have seen that the errors-in-variables method can yield consistent parameter estimates in a variety of situations. The task we have not done is to indicate how to obtain a consistent estimate of the (asymptotic) variance of such estimates or how to incorporate autocorrelation of the disturbance term in the errors-in-variables estimation method. This section takes up that remaining task. The errors-in-variables models we considered in the previous sections have the following common representation:[27]

$$y_t = x_t\beta + v_t \ , \qquad (7.1)$$

where x_t is a row vector, β is the column vector of parameters, and v_t is the disturbance term which is the sum of the behavioral error u_t and the expectation error e_t:

$$v_t = u_t + e_t \ . \qquad (7.2)$$

We require that (i) there exists a (row) vector of instruments z_t that satisfies

$$E(v_t z_t) = 0, \ (t = 1, 2, \ldots, T) \qquad (7.3)$$

and (ii) $\{y_t, x_t, z_t\}$ are jointly stationary and ergodic, and (iii) $E(z_t' x_t)$ and $E(z_t' z_t)$ are matrices of full rank. For example, the errors-in-variables model (3.5) is a special case of this model when x_t is replaced by (Y_{1t}, Y_t^1, X_{1t}), β by (γ, α, β), z_t by Z_{t-1}, u_t by u_{1t} and e_t by $-\alpha e_t^1$.

The same model emerges from a different context. For example, consider a simple model relating the k - period forward rate as of t, x_t, to the spot rate at time t + k, $\tilde{y}_{t+k}$, by

$$E(\tilde{y}_{t+k}|I_t) = \alpha_0 + \alpha_1 x_t. \qquad (7.4)$$

If x_t is the unbiased forecast of y_{t+k}, we expect (α_0, α_1) to be (0, 1). As Hansen and Hodrick (1980) and Hakkio (1980) have shown, (7.4) can be written as

$$\tilde{y}_{t+k} = \alpha_0 + \alpha_1 x_t + v_t \qquad (7.5)$$

where v_t is represented by a moving average process of order k-1:

$$v_t = \varepsilon_t + c_1 \varepsilon_{t+1} + \dots + c_{k-1}\varepsilon_{t+k-1} \text{ with } E(\varepsilon_{t+j}|I_t) = 0.$$

This is a special case of our model (7.1) - (7.3) with $y_t = \tilde{y}_{t+k}$, $z_t = (1, x_t)$ and $u_t = 0$.

If we examine the errors-in-variables models in the previous sections and the k-period-ahead forecast model just described more carefully, we notice that we can impose more restrictions on the disturbance term v_t than is embodied in (7.3). In the k-period-ahead forecast model, v_t is uncorrelated not only with current z_t but also with past values of z_t. In the errors-in-variables model, it is reasonable to expect that the behavioral error u_t is uncorrelated with both current and past values of z_t. Furthermore, the rational expectations hypothesis implies that the expectation error e_t is orthogonal to z_t, z_{t-1}, ..., i.e., $E(e_t|z_t, z_{t-1}, \dots) = 0$. Thus, at least in the models we considered so far, we can assume

$$E(v_t z_{t-j}) = 0, \; j = 0, 1, 2, \dots \qquad (7.3')$$

The methods we present in the next subsection exploit (7.3) and do not exploit (7.3') for $j \geq 1$. We will then present a method that exploits

(7.3').

7,2. Estimation under (7.3)

We noted in the previous sections that application of the instrumental variable technique to (7.1) with z_t as instruments yields a consistent estimate $\hat{\beta}$:

$$\hat{\beta} = (X'Z(Z'Z)^{-1} Z'X)^{-1} X'Z(Z'Z)^{-1} Z'y, \qquad (7.6)$$

where $X' = (x_1', \ldots, x_T')$, $Z' = (z_1', \ldots, z_T')$, $y' = (y_1, \ldots, y_T)$. We also noted that the formula routinely calculated by an instrumental variables package

$$T\hat{\sigma}_v^2 (X'Z(Z'Z)^{-1} Z'X)^{-1}, \qquad (7.7)$$

where $\hat{\sigma}_v^2$ is the sample variance of the estimated residuals, is not a consistent estimate of the asymptotic variance matrix of $\hat{\beta}$ if v_t is autocorrelated. Hansen (1980) has shown that the asymptotic variance matrix of (7.6) is[28]

$$H^{-1} \Sigma_{XZ} \Sigma_{ZZ} M\Sigma_{ZZ}^{-1} \Sigma_{XZ}' H^{-1}, \qquad (7.8)$$

where $H = \Sigma_{XZ} \Sigma_{ZZ}^{-1} \Sigma_{XZ}'$, $\Sigma_{XZ} = E(x_t' x_t)$, $\Sigma_{ZZ} = E(z_t' z_t)$, $M = \lim T^{-1} E(Z'vv'Z)$, and $v' = (v_1, \ldots, v_T)$. If we have $E(v_t v_{t-j} | z_t, z_{t-1}, \ldots) = E(v_t v_{t-j})$, i.e., if the conditional covariance is equal to the unconditional covariance, then $M = \text{plim } T^{-1}Z'\bar{V}Z$, where $\bar{V} = \text{Var}(v)$. If v_t is a stochastic process with finite number of parameters, then a consistent estimate $\hat{\bar{V}}$ of $\bar{V}$ can be obtained from the estimated residuals. Thus, a consistent estimate of the asymptotic variance of (7.6) is obtained by replacing M in (7.8) by $T^{-1} Z'\hat{\bar{V}}Z$, Σ_{XZ} by $T^{-1} X'Z$ and Σ_{ZZ} by $T^{-1}Z'Z$.

However, (7.6) is not most efficient in the class of estimators of β that exploits (7.3). Hansen (1980) has shown that the best estimator in that class is

$$\hat{\beta} = (X'Z \hat{M}^{-1} Z'X)^{-1} X'Z \hat{M}^{-1} Z' y \tag{7.9}$$

where $\hat{M}$ is a consistent estimate of M. The asymptotic variance matrix of (7.9) is

$$(\Sigma_{XZ} M^{-1} \Sigma'_{XZ})^{-1} \tag{7.10}$$

whose consistent estimate is given by $(X'Z \hat{M}^{-1} Z'X)^{-1} T^2$.

7.3. Estimation under (7.3')

The reader may by now wonder why the GLS (Generalized least squares) version of (7.6)

$$\hat{\beta}_{GLS} = (X'\hat{\bar{V}}^{-1} Z(Z'\hat{\bar{V}}^{-1} Z)^{-1} Z'\hat{\bar{V}}^{-1} X)^{-1} X'\hat{\bar{V}}^{-1} Z(Z'\hat{\bar{V}}^{-1} Z)^{-1} Z'\hat{\bar{V}}^{-1} y \tag{7.11}$$

is not viable. As Flood and Barber (1980) have shown, this estimator is not consistent. To see this, we first note that consistency requires $T^{-1} Z' \hat{\bar{V}}^{-1} v$ converges to zero in probability. But $T^{-1} Z' \hat{\bar{V}}^{-1} v$ involves terms like $v_t z_{t+j}$ $(j \geq 1)$ whose expectation is not necessarily zero under (7.3) or (7.3'). To make (7.11) consistent, we would have to assume that z_t is exogenous to v_t, i.e., $E(v_t Z_s) = 0$ for all t and s. We now present the estimation method proposed by Hayashi and Sims (1980) which exploits (7.3'). In order to ensure consistency, we must construct an estimator that does not involve terms like $v_t Z_{t+j}$ $(j \geq 1)$. This is accomplished by filtering y_t, x_t, v_t forward before applying instruments z_t.

To make the exposition simple, we assume that the behavioral error u_t in (7.2) is ARMA (p,q) and the expectation error e_t is MA (r). This includes the errors-in-variables models we considered and the k-period-ahead forecast model as special cases. Thus u_t and e_t are written as

$$u_t + a_1 u_{t-1} + \dots + a_p u_{t-p} = \eta_t + b_1 \eta_{t-1} + \dots + b_q \eta_{t-q}, \tag{7.12}$$

$$e_t = \varepsilon_t + c_1 \varepsilon_{t+1} + \dots + c_r \varepsilon_{t+r} \ . \tag{7.13}$$

Since e_t represents the expectation error, ε_{t+j} $(j \geq 0)$ is orthogonal to $\eta_{t-1}, \eta_{t-2}, \dots$ and possibly to η_t. From this and (7.2), (7.12), (7.13), it is easy to see that v_t can be represented as

$$a_p v_t + a_{p-1} v_t + \dots + v_{t+p} = w_t + d_1 w_{t+1} + \dots + d_{r+q} w_{r+q+t} \tag{7.14}$$

for some nonautocorrelated, constant-variance series $\{w_t\}$ that satisfies

$$E(w_t z_t) = 0 \ . \tag{7.15}$$

Without loss of generality, we normalize the moving average part by requiring that $d(L^{-1}) = 1 + d_1 L^{-1} + \dots + d_{r+q} L^{-r-q}$ is invertible, i.e.,

$$d(L^{-1})^{-1} = 1 + f_1 L^{-1} + f_2 L^{-2} + \dots \tag{7.16}$$

with $\sum_{i=1}^{\infty} |f_i| < \infty$.

Now define a (T - p) x T matrix A by

$$A = \begin{matrix} a_p & a_{p-1} & \cdots & 1 & & & \\ & \ddots & & & \ddots & & \\ & & a_p & a_{p-1} & \cdots & 1 \end{matrix}$$

and a (T -p) x (T - p) matrix D by

$$D = \begin{bmatrix} 1 & d_1 & \cdots & d_{r+q} & & \\ & \ddots & \ddots & & \ddots & \\ & & \ddots & & & d_{r+q} \\ & & & \ddots & & \vdots \\ & 0 & & & \ddots & d_1 \\ & & & & & 1 \end{bmatrix}$$

Assuming the terminal values $w_{T-p+1}, \ldots, w_{T-p+q+r}$ are all zero,[29] we have $D^{-1}Av = \bar{w}$, where $\bar{w}' = (w_1, \ldots, w_{T-p})$. Also note that the (i,j) element of D^{-1} is f_{j-i} if $j \geq i$ and zero if $j < i$. Thus, by applying the forward filter $D^{-1}A$ on (7.1), we obtain

$$D^{-1}A\,y = D^{-1}A\,X\,\beta + \bar{w}\ . \tag{7.17}$$

If we know the true value of D and A, then we can simply apply the instrumental variable technique on the <u>forward-filtered</u> equation (7.17) with <u>unfiltered</u> z_t as instruments. This yields

$$\hat{\beta}(A,D) = (X'A'D^{-1'}\,\bar{Z}(\bar{Z}'\bar{Z})^{-1}\,\bar{Z}'D^{-1}AX)^{-1}X'A'D^{-1'}\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'D^{-1}Ay \tag{7.18}$$

where $\bar{Z}' = (z_1', \ldots, z_{T-p}')$. Since w_t and z_t are uncorrelated, (7.18) is consistent. Furthermore, since $\{w_t\}$ is nonautocorrelated with a constant variance, the formula for the asymptotic variance of (7.18) routinely computed by an instrumental variable package when applied to the filtered equation is consistent. If we don't know the true value of A and D, then we can proceed as follows: The first step is to apply the instrumental variable technique on the unfiltered equation (7.1) with z_t as instruments. We can obtain a $T^{1/2}$ - consistent estimate[30] $(\hat{A}, \hat{D})$ of (A, D) by fitting ARMA(p, q) on the estimated residuals.[31] Second, use $(\hat{A}, \hat{D})$ to obtain the forward-filtered equation corresponding (7.17). The third step is to apply the instrumental variable technique on the filtered equation with z_t as instruments. Hayashi and Sims (1980) have shown that the resulting estimator $\hat{\beta}(\hat{A}, \hat{D})$ is consistent and has the same asymptotic distribution as $\hat{\beta}(A, D)$ does. Therefore, we can just take the matrix given by an instrumental variable package in the third step as a consistent estimate of the asymptotic variance of $\hat{\beta}(\hat{A}, \hat{D})$.

Since this estimator does not belong to the class of estimators in which (7.9) is most efficient, (7.18) may or may not be more efficient than (7.9). However, it has been shown in Hayashi and Sims (1980) that (7.18) is always more efficient than (7.9) if v_t is MA(1) and x_t is a scaler first-order autoregressive process.

8. Conclusion

In this paper we have considered three kinds of models -- the imperfect information model, the perfect information model, and the model that involves the present discounted value. It seems that virtually all the macroeconomic models in the literature fall into either one of the three categories. For each of the three models, we examined two classes of estimation methods -- the methods that require that the policy rule (or more precisely, the X_t process) is linear and fixed, and the errors-in-variables methods that do not require such an assumption.

If one wishes to estimate the imperfect information model which involves the Lucas supply function, the estimation methods with fixed policy rule presented in 3.2 and 4.2 must be used, since the errors-in-variables methods in 3.3 and 4.3 cannot estimate both A and C_o in (2.1). However, the methods in 3.2 and 4.2 are likely to be infeasible in practice, partly because the rational expectations constraint is highly nonlinear and partly because the forecasting equation involves a number of variables on the right hand side. If one wishes to estimate the perfect information model or the model with the present discounted value, the errors-in-variables method in sections 5 and 6 must be preferred. In such models the errors-in-variables model can exploit the fact that the behavioral error is uncorrelated with purely exogenous variables, and can identify both A and C_o in (5.1). Furthermore, we can always obtain a consistent estimate of the asymptotic variance of parameter estimates given by the errors-in-variables methods, by using either of the two methods presented in section 7.

FOOTNOTES

1. The term "disturbance" is reserved for later use.

2. However, policy instruments that alter the incentive structure (e.g. tax rates) can affect the structural parameters.

3. The estimation methods to be presented in this paper can also handle the rational expectation of exogenous variables.

4. Another popular assumption is that X_t is ARMA. Estimation of the structural parameters will in fact be easier under the ARMA assumption.

5. The information structure considered by Taylor (1979) implies that the information set I_{t-1} includes current value of forcing variables X_t. Under this assumption, (2.5) can serve as the forecasting equation because ${}_{t-1}X_t$ is now equal to X_t. The estimation methods to be presented in sections 3 and 4 under the assumption of fixed policy rule can be straightforwardly applied to Taylor's information structure with (2.5) as the forecasting equation. Since the new forecasting equation (2.5) with ${}_{t-1}X_t$ replaced by X_t is independent of the policy rule, the assumption (2.4) of fixed policy rule becomes unnecessary. However, it is easy to show that, if X_t is included in I_{t-1}, there is no way to identify both C_o and A in (2.1).

6. Solely to make the notation simple, we do not include lagged endogenous variables in the equation. If they are present in the equation and if u_{1t} is autocorrelated, then we have to worry about the correlation between u_{1t} and lagged endogenous variables. But this is not the feature unique to the rational expectation models.

7. If lagged endogenous variables are present in (3.1), they should be treated as predetermined since u_{1t} is not autocorrelated.

8. The original assumption made by Sargent (1973, 76) is that X_t is generated by a vector autoregressive process. Wallis' original assumption is that X_t is ARMA. In either case, all the forcing variables are purely exogenous in the sense $E(X_t u_s) = 0$ for all t and s.

9. Sargent's (1976) estimates are inconsistent, because he did not include enough number of lagged variables in his first stage regression.

10. This was pointed out by Durbin (1970), Sims (1977) and Mishkin (1980) in slightly different contexts. This problem has been ignored or unnoticed by many authors. Standard errors reported by Sargent (1973,76), Barro (1977), and Fair (1979) are inappropriate in the sense they are not consistent estimates of the relevant asymptotic variances.

11. See Jorgenson and Laffont (1974) and Amemiya (1977) for NL3LS.

12. A sufficient condition for ergodieity would be that $(u_t, \varepsilon_t, \theta_t)$ are ergodic. In particular, the fixed policy rule assumption (2.4) with ARMA behavioral error u_t is sufficient for ergodicity.

13. Recall $\{e_t^1\}$ is autocorrelated because of its orthogonality property.

14. As noted in footnote 5, the same situation arises if I_{t-1} includes X_t. The identification of γ_i and α_j by the methods in 3.2 (and in 4.2) depends entirely on their ability to distinguish X_t from ${}_{t-1}X_t$.

15. It is unfair to compare the standard errors of parameter estimates by the first method that are routinely computed by a standard computer package with the standard error of parameter estimates by the third method, because the former are downward-biased estimates of the true sampling error.

16. Fair's original model assumes that X_t is ARMA. He used 2LS (two-stage least squares) in his own iteration.

17. See Hausman (1975) for a proof of the equivalence of 3LS and FIML (full information maximum likelihood) in the simultaneous equations model that is linear in variables but nonlinear in parameters.

18. See Brundy and Jorgenson (1971) for FIVE.

19. As Brundy and Jorgenson (1971) have shown, the size of the matrix Z can be reduced without affecting the efficiency of parameter estimates.

20. Lagged endogenous variables can be introduced into (5.4) without altering the estimation methods to be presented in the text. See footnote 6.

21. Of course, the regularity condition (the ergodicity property) must be satisfied. See footnote 12 for a sufficient condition for ergodicity.

22. The transversality condition

$$\lim_{T\to\infty} \lambda^T E_t(\alpha p_T - b_T - \beta y_T) = 0$$

is assumed to obtain the "forward-looking" solution (6.3). See Sargent (1979, chapters XI and XIV) for more details.

23. The reason the maximum likelihood method is applicable here is not applicable in the joint estimation of (3.3) and (3.4) in that the covariance between (u_t, η_t) and b_{t+j} is specified here. If b_t is not included in x_t, NL3LS must be used to estimate (6.3) and (6.5) jointly.

24. This is noted by Sargent himself (see Sargent (1979, chapter XII,4.c)).

25. The model considered by Hansen and Sargent (1980) does not involve a variable like b_t that appears on the right hand side of the behavioral equation (6.3) in addition to the present discounted value q_t.

26. The section draws heavily on Hansen (1980) and Hayashi and Sims (1980).

27. This is the linear case of the model considered by Hansen (1980).

28. Two additional technical assumptions are necessary for the estimator to be asymptotically normal. They are assumptions (iv) and (iii) in Hansen's (1980) Theorem 4.1. If $E(v_t v_{t-j} | z_t, z_{t-1}, \ldots) = E(v_t v_{t-j})$ and if v_t is ARMA, then the two assumptions are satisfied.

29. This assumption does not affect the asymptotic properties of the resulting estimator.

30. An estimate $\hat{\beta}$ of β is said to be $T^{1/2}$-consistent if $T^{1/2}(\hat{\beta}-\beta)$ is bounded in probability.

31. The representation that a standard ARMA package estimates is $v_t + a_1 v_{t-1} + \ldots + a_p v_{t-p} = w_t + \delta_1 w_{t-1} + \ldots + \delta_{r+q} w_{t-r-q}$, with the normalizing restriction that $\delta(L) = 1 + \delta_1 L + \ldots + \delta_{r+q} L^{r+q}$ is invertible. To recover $\{d_i\}$ from $\{\delta_i\}$, we first factorize

$$\delta(L) = (1-\lambda_1 L) \ldots (1-\lambda_{r+q} L).$$

$\{d_i\}$ is obtained from

$$d_{r+q} + d_{r+q-1} L + \ldots + L^{r+q} = (1-\lambda_1^{-1} L) \ldots (1-\lambda_{r+q}^{-1} L)\, \delta_{r+q}.$$

Then $\{d_i\}$ satisfies the invertibility condition (7.16).

References

Abel, A. (1977), Investment and the Value of Capital, Ph.D. Thesis, Massachusetts Institute of Technology.

Amemiya, T. (1977), "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equations Model," Econometrica 45, 955-968.

Barro, R. (1977), "Unanticipated Money Growth and Unemployment in the United States," American Economic Review 67, 101-115.

Brandy, J. and D. Jorgenson (1971), "Efficient Estimation of Simultaneous Equations by Instrumental Variables," Review of Economics and Statistics 53, 207=224.

Durbin, J. (1970), "Testing for Serial Correlation in Least Squares Regression when some of the Regressors are Lagged Dependent Variables," Econometrica 38, 410-421.

Fair, R. (1979), "An Analysis of the Accuracy of Four Macroeconomic Models," Journal of Political Economy 67, 701-718.

Flood, R. and P. Garber (1980), "A Pitfall in Estimation of Models with Rational Expectations," Journal of Monetary Economics 6, 433-435.

Friedman, M. (1957), A Theory of Consumption Function, Princeton University Press.

Hakkio, C. (1980), "Expectations and the Forward Exchange Rate," NBER working paper #439.

Hansen, L. (1980), "Large Sample Properties of Generalized Methods of Moments Estimators," mimeo. , Carnegie-Mellon University.

---------and R. Hedrick (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," Journal of Political Economy 88.

--------and T. Sargent (1980), "Formulating and Estimating Dynamic Linear Rational Expectations Models," Journal of Economic Dynamics and Control 1.

Hausman, J. (1975), "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models," Econometrica 43, 727-738.

Hayashi, F. (1979), "A New Estimation Procedure under Rational Expectations," Economics Letters 4, 41-43.

---------- and C. Sims (1980), "Efficient Estimation of Time Series Models with Predetermined, But Not Exogenous, Instruments," mimeo.

References (continued)

Jorgenson, D. and J. Laffont (1974), "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances," Annals of Economic and Social Measurement 3, 615-640.

Kennan, J. (1979), "The Estimation of Partial Adjustment Models with Rational Expectations," Econometrica 47, 1441-1456.

Lucas, R. and L. Rapping (1969), "Real Wages, Employment and Inflation," Journal of Political Economy 77, 721-754.

-------------and E. Prescott (1971), "Investment under Uncertainty," Econometrica 39, 659-681

Lucas, R. (1973), "Some International Evidence on Output-Inflation Tradeoffs," American Economic Review 68, 326-334.

------------(1976), "Econometric Policy Evaluation:A Critique," in K. Brunner and A. Meltzer, eds., The Phillips Curve and Labor Markets,North-Holland.

McCallum, B., (1976a), "Rational Expectations and the Estimation of Econometric Models: An Alternative Procedure," International Economic Review 17, 484-490.

-----------(1976b), "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates," Econometrica 44, 43-52.

Mishkin, F.(1980), "Does Anticipated Monetary Policy Matter?: An Econometric Investigation," NBEF working paper #506.

Muth, J. (1961), "Rational Expectations and the Theory of Price Movements," Econometrica 29, 315-335.

Sargent, T. (1973), "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment," Brookings Papers on Economic Activity 2, 429-472.

-----------(1976), "A Classical Macroeconomic Model of the United States," Journal of Political Economy 84, 207-238.

----------(1979, Macroeconomic Theory, Academic Press.

Schiller, R. (1978), "Rational Expectations and the Dynamic Structure of Macroeconomic Models," Journal of Monetary Economics 4, 1-44.

Sims, C. (1977),"Comment," Journal of the American Statistical Association 72, 23-24.

Taylor, J. (1979), "Estimation and Control of a Macroeconomic Model with Rational Expectations," Econometrica 47, 1267-1286.

References continued

Wallis, K. (1980), "Econometric Implication of the Rational Expectations Hypothesis," Econometrica 48, 49-74.

Wickens, M. (1977),"The Efficient Estimation of Econometric Modelss with Rational Expectations, mimeo., University of Essex.