# 4

# Estimating and Testing Single Equations

## 4.1 Notation

This chapter discusses the estimation and testing of single equations. The notation that will be used is the same as that used in Section 1.2. The model is written as

$$f_i(y_t, x_{t,}\alpha_i) = u_{it}, \quad (i = 1, \ldots, n), \quad (t = 1, \ldots, T) \tag{4.1}$$

where $y_t$ is an $n$–dimensional vector of endogenous variables, $x_t$ is a vector of predetermined variables (including lagged endogenous variables), $\alpha_i$ is a vector of unknown coefficients, and $u_{it}$ is the error term for equation $i$ for observation $t$. It will be assumed that the first $m$ equations are stochastic, with the remaining $u_{it}$ $(i = m + 1, \ldots, n)$ identically zero for all $t$.

The following notation is also used. $u_i$ denotes the $T$–dimensional vector $(u_{i1}, \ldots, u_{iT})'$. $G_i'$ denotes the $k_i \times T$ matrix whose $t$th column is $\partial f_i(y_t, x_t, \alpha_i)/\partial\alpha_i$, where $k_i$ is the dimension of $\alpha_i$. $\alpha$ denotes the vector of all the unknown coefficients in the model: $\alpha = (\alpha_1', \ldots, \alpha_m')$. The dimension of $\alpha$ is $k$, where $k = \sum_{i=1}^{m} k_i$. Finally, $Z_i$ denotes a $T \times K_i$ matrix of predetermined variables that are to be used as first stage regressors for the two stage least squares technique.

It will sometimes be useful to consider the case in which the equation to be estimated is linear in coefficients. In this case equation $i$ in 4.1 will be written as

$$y_{it} = X_{it}\alpha_i + u_{it}, \quad (i = 1, \ldots, n), \quad (t = 1, \ldots, T) \tag{4.2}$$

where $y_{it}$ is the left hand side variable and $X_{it}$ is a $k_i$–dimensional vector of explanatory variables in the equation. $X_{it}$ includes both endogenous and predetermined variables. Both $y_{it}$ and the variables in $X_{it}$ can be nonlinear functions of other variables, and thus 4.2 is more general than the standard linear model. All that is required is that the equation be linear in $\alpha_i$. Note from the definition of $G_i'$ above that for equation 4.2 $G_i' = X_i'$, where $X_i'$ is the $k_i \times T$ matrix whose $t$th column is $X_{it}$.

Each equation in 4.1 is assumed to have been transformed to eliminate any autoregressive properties of its error term. If the error term in the untransformed version, say $w_{it}$ in equation $i$, follows a $r$th order autoregressive process, $w_{it} = \rho_{1i} w_{it-1} + \ldots + \rho_{ri} w_{it-r} + u_{it}$, where $u_{it}$ is *iid*, then equation $i$ is assumed to have been transformed into one with $u_{it}$ on the right hand side. The autoregressive coefficients $\rho_{1i}, \ldots, \rho_{ri}$ are incorporated into the $\alpha_i$ coefficient vector, and the additional lagged values that are involved in the transformation are incorporated into the $x_t$ vector. This transformation makes the equation nonlinear in coefficients if it were not otherwise, but this adds no further complications to the model because it is already allowed to be nonlinear. It does result in the "loss" of the first $r$ observations, but this has no effect on the asymptotic properties of the estimators. $u_{it}$ in 4.1 can thus be assumed to be *iid* even though the original error term may follow an autoregressive process.

Many nonlinear optimization problems in macroeconometrics can be solved by general purpose algorithms like the Davidon-Fletcher-Powell (DFP) algorithm. This algorithm is discussed in Fair (1984), Section 2.5, and this discussion will not be repeated here. Problems for which the algorithm seems to work well and those for which it does not are noted below.

Unless otherwise stated, the goodness of fit measures have not been adjusted for degrees of freedom. For the general model considered here (nonlinear, simultaneous, dynamic) only asymptotic results are available, and so if any adjustments were made, they would have to be based on analogies to simpler models. In many cases there are no obvious analogies, and so no adjustments were made. Fortunately, in most cases the number of observations is fairly large relative to numbers that might be used in the subtraction, and so the results are not likely to be sensitive to the current treatment.

## 4.2 Two Stage Least Squares[1]

Probably the most widely used estimation technique for single equations that produces consistent estimates is two stage least squares (2SLS).[2] The 2SLS estimate of $\alpha_i$ (denoted $\hat{\alpha}_i$) is obtained by minimizing

$$S_i = u_i' Z_i (Z_i' Z_i)^{-1} Z_i' u_i = u_i' D_i u_i \tag{4.3}$$

with respect to $\alpha_i$. $Z_i$ can differ from equation to equation. An estimate of the covariance matrix of $\hat{\alpha}_i$ (denoted $\hat{V}_{2ii}$) is

$$\hat{V}_{2ii} = \hat{\sigma}_{ii} (\hat{G}_i' D_i \hat{G}_i)^{-1} \tag{4.4}$$

where $\hat{G}_i$ is $G_i$ evaluated at $\hat{\alpha}_i$, $\hat{\sigma}_{ii} = T^{-1} \sum_{t=1}^{T} \hat{u}_{it}^2$, and $\hat{u}_{it} = f_i(y_t, x_t, \hat{\alpha}_i)$.

The 2SLS estimate of the $k \times k$ covariance matrix of all the coefficient estimates in the model (denoted $\hat{V}_2$) is

$$\hat{V}_2 = \begin{bmatrix} \hat{V}_{211} & . & . & . & \hat{V}_{21m} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ \hat{V}_{2m1} & . & . & . & \hat{V}_{2mm} \end{bmatrix} \tag{4.5}$$

where

$$\hat{V}_{2ij} = \hat{\sigma}_{ij} (\hat{G}_i' D_i \hat{G}_i')^{-1} (\hat{G}_i' D_i D_j \hat{G}_j') (\hat{G}_j' D_j \hat{G}_j')^{-1} \tag{4.6}$$

and $\hat{\sigma}_{ij} = T^{-1} \sum_{t=1}^{T} \hat{u}_{it} \hat{u}_{jt}$.

## 4.3 Estimation of Equations with Rational Expectations[3]

With only slight modifications, the 2SLS estimator can be used to estimate equations that contain expectational variables in which the expectations are formed rationally. As discussed later in this chapter, this estimation technique

[1]See Fair (1984), Section 6.3.2, for a more detailed discussion of the two stage least squares estimator, especially for the case in which the equation is linear in coefficients and has an autoregressive error.

[2]Ordinary least squares is used a lot in practice in the estimation of commercial models even when the estimator does not produce consistent estimates. This lack of care in the estimation of such models is undoubtedly one of the reasons there has been so little academic interest in them.

[3]The material in this section is taken from Fair (1993b), Section 3 and Appendix A.

can be used to test the rational expectations hypothesis against other alternatives. The modifications of the 2SLS estimator that are needed to handle the rational expectations case are discussed in this section.

It will be useful to begin with an example. Assume that the equation to be estimated is

$$y_{it} = X_{1it}\alpha_{1i} + E_{t-1}X_{2it+j}\alpha_{2i} + u_{it}, \quad (t = 1, \ldots, T) \tag{4.7}$$

where $X_{1it}$ is a vector of explanatory variables and $E_{t-1}X_{2it+j}$ is the expectation of $X_{2it+j}$ based on information through period $t-1$. $j$ is some fixed positive integer. This example assumes that there is only one expectational variable and only one value of $j$, but this is only for illustration. The more general case will be considered shortly.

A traditional assumption about expectations is that the expected future values of a variable are a function of its current and past values. One might postulate, for example, that $E_{t-1}X_{2it+j}$ depends on $X_{2it}$ and $X_{2it-1}$, where it assumed that $X_{2it}$ (as well as $X_{2it-1}$) is known at the time the expectation is made. The equation could then be estimated with $X_{2it}$ and $X_{2it-1}$ replacing $E_{t-1}X_{2it+j}$ in 4.7. Note that this treatment, which is common to many macroeconometric models, is not inconsistent with the view that agents are "forward looking." Expected future values do affect current behavior. It's just that the expectations are formed in fairly simply ways—say by looking only at the current and lagged values of the variable itself.

Assume instead that $E_{t-1}X_{2it+j}$ is rational and assume that there is an observed vector of variables (observed by the econometrician), denoted here as $Z_{it}$, that is used in part by agents in forming their (rational) expectations. The following method does not require for consistent estimates that $Z_{it}$ include all the variables used by agents in forming their expectations.

Let the expectation error for $E_{t-1}X_{2it+j}$ be

$${}_{t-1}\epsilon_{it+j} = X_{2it+j} - E_{t-1}X_{2it+j} \quad (t = 1, \ldots, T) \tag{4.8}$$

where $X_{2it+j}$ is the actual value of the variable. Substituting 4.8 into 4.7 yields

$$\begin{aligned} y_{it} &= X_{1it}\alpha_{1i} + X_{2it+j}\alpha_{2i} + u_{it} -_{t-1} \epsilon_{it+j}\alpha_{2i} \\ &= X_{it}\alpha_i + v_{it} \quad (t = 1, \ldots, T) \end{aligned} \tag{4.9}$$

where $X_{it} = (X_{1it}\ X_{2it+j})$, $\alpha_i = (\alpha_{1i}\ \alpha_{2i})'$, and $v_{it} = u_{it} -_{t-1} \epsilon_{it+j}\alpha_{2i}$.

Consider now the 2SLS estimation of 4.9, where the vector of first stage regressors is the vector $Z_{it}$ used by agents in forming their expectations. A necessary condition for consistency is that $Z_{it}$ and $v_{it}$ be uncorrelated. This will be true if both $u_{it}$ and ${}_{t-1}\epsilon_{it+j}$ are uncorrelated with $Z_{it}$. The assumption that $Z_{it}$ and $u_{it}$ are uncorrelated is the usual 2SLS assumption. The assumption that $Z_{it}$ and ${}_{t-1}\epsilon_{it+j}$ are uncorrelated is the rational expectations assumption. If expectations are formed rationally and if the variables in $Z_{it}$ are used (perhaps along with others) in forming the expectation of $X_{2it+j}$, then $Z_{it}$ and ${}_{t-1}\epsilon_{it+j}$ are uncorrelated. Given this assumption (and the other standard assumptions that are necessary for consistency), the 2SLS estimator of $\alpha_i$ in equation 4.9 is consistent.

The 2SLS estimator does not, however, account for the fact that $v_{it}$ in 4.9 is a moving average error of order $j-1$, and so it loses some efficiency for values of $j$ greater than 1. The modification of the 2SLS estimator to account for the moving average process of $v_{it}$ is Hansen's (1982) generalized method of moments (GMM) estimator, which will now be described.

Write 4.9 in matrix notation as

$$y_i = X_i\alpha_i + v_i \tag{4.10}$$

where $X_i$ is $T \times k_i$, $\alpha_i$ is $k_i \times 1$, and $y_i$ and $v_i$ are $T \times 1$. Also, let $Z_i$ denote, as above, the $T \times K_i$ matrix of first stage regressors. The assumption in 4.9 that there is only one expectational variable and only one value of $j$ can now be relaxed. The matrix $X_i$ can include more than one expectational variable and more than one value of $j$ per variable. In other words, there can be more than one led value in this matrix.

The 2SLS estimate of $\alpha_i$ in 4.10 is

$$\hat{\alpha}_i = [X_i'Z_i(Z_i'Z_i)^{-1}Z_i'X_i]^{-1}X_i'Z_i(Z_i'Z_i)^{-1}Z_i'y_i \tag{4.11}$$

This use of the 2SLS estimator for models with rational expectations is due to McCallum (1976).

As just noted, this use of the 2SLS estimator does not account for the moving average process of $v_{it}$, and so it loses efficiency if there is at least one value of $j$ greater than 1. Also, the standard formula for the covariance matrix of $\hat{\alpha}_i$ is not correct when at least one value of $j$ is greater than 1. If, for example, $j$ is 3 in 4.9, an unanticipated shock in period $t+1$ will affect ${}_{t-1}\epsilon_{it+3}$, ${}_{t-2}\epsilon_{it+2}$, and ${}_{t-3}\epsilon_{it+1}$, and so $v_{it}$ will be a second order moving average. Hansen's GMM estimator accounts for this moving average process. The GMM estimate in the present case (denoted $\tilde{\alpha}_i$) is

$$\tilde{\alpha}_i = (X_i'Z_iM_i^{-1}Z_i'X_i)^{-1}X_i'Z_iM_i^{-1}Z_i'y_i \tag{4.12}$$

where $M_i$ is some consistent estimate of lim $T^{-1}E(Z_i'v_iv_i'Z_i)$. The estimated covariance matrix of $\tilde{\alpha}_i$ is

$$T(X_i'Z_iM_i^{-1}Z_i'X_i)^{-1} \tag{4.13}$$

There are different versions of $\tilde{\alpha}_i$ depending on how $M_i$ is computed. To compute $M_i$, one first needs an estimate of the residual vector $v_i$. The residuals can be estimated using the 2SLS estimate $\hat{\alpha}_i$:

$$\hat{v}_i = y_i - X_i\hat{\alpha}_i \tag{4.14}$$

A general way of computing $M_i$ is as follows. Let $f_{it} = \hat{v}_{it} \otimes Z_{it}$, where $\hat{v}_{it}$ is the $t$th element of $\hat{v}_i$. Let $R_{ip} = (T-p)^{-1}\sum_{t=p}^{T} f_{it}f_{it-p}'$, $p = 0, 1, \ldots, P$, where $P$ is the order of the moving average. $M_i$ is then $(R_{i0} + R_{i1} + R_{i1}' + \ldots + R_{iP} + R_{iP}')$. In many cases computing $M_i$ in this way does not result in a positive definite matrix, and so $\tilde{\alpha}_i$ cannot be computed. I have never had much success in obtaining a positive definite matrix for $M_i$ computed in this way.

There are, however, other ways of computing $M_i$. One way, which is discussed in Hansen (1982) and Cumby, Huizinga, and Obstfeld (1983) but is not pursued here, is to compute $M_i$ based on an estimate of the spectral density matrix of $Z_{it}'v_{it}$ evaluated at frequency zero. An alternative way, which is pursued here, is to compute $M_i$ under the following assumption:

$$E(v_{it}v_{is} \mid Z_{it}, Z_{it-1}, \ldots) = E(v_{is}v_{is}) \quad , \quad t \geq s \tag{4.15}$$

which says that the contemporaneous and serial correlations in $v_i$ do not depend on $Z_i$. This assumption is implied by the assumption that $E(v_{it}v_{is}) = 0, t \geq s$, if normality is also assumed. Under this assumption $M_i$ can be computed as follows. Let $a_{ip} = (T-p)^{-1}\sum_{t=p}^{T}\hat{v}_{it}\hat{v}_{it-p}$ and $B_{ip} = (T-p)^{-1}\sum_{t=p}^{T} Z_{it}Z_{it-p}'$, $p = 0, 1, \ldots, P$. $M_i$ is then $(a_{i0}B_{i0} + a_{i1}B_{i1} + a_{i1}B_{i1}' + \ldots + a_{iP}B_{iP} + a_{iP}B_{iP}')$. In practice, this way of computing $M_i$ usually results in a positive definite matrix.

### The Case of an Autoregressive Structural Error

Since many macroeconometric equations have autoregressive error terms, it is useful to consider how the above estimator is modified to cover this case. Return for the moment to the example in 4.7 and assume that the error term $u_{it}$ in the equation follows a first order autoregressive process:

$$u_{it} = \rho_{1i}u_{it-1} + \eta_{it} \tag{4.16}$$

Lagging equation 4.7 one period, multiplying through by $\rho_{1i}$, and subtracting the resulting expression from 4.7 yields

$$y_{it} = \rho_{1i} y_{it-1} + X_{1it}\alpha_{1i} - X_{1it-1}\alpha_{1i}\rho_{1i} + E_{t-1}X_{2it+j}\alpha_{2i} - E_{t-2}X_{2it+j-1}\alpha_{2i}\rho_{1i} + \eta_{it} \qquad (4.17)$$

Note that this transformation yields a new viewpoint date, $t-2$. Let the expectation error for $E_{t-2}X_{2it+j-1}$ be

$$_{t-2}\epsilon_{it+j-1} = X_{2it+j-1} - E_{t-2}X_{2it+j-1} \qquad (4.18)$$

Substituting 4.8 and 4.18 into 4.17 yields

$$\begin{aligned} y_{it} = \rho_{1i} y_{it-1} + X_{1it}\alpha_{1i} - X_{1it-1}\alpha_{1i}\rho_{1i} + X_{2it+j}\alpha_{2i} - X_{2it+j-1}\alpha_{2i}\rho_{1i} \\ +\eta_{it} -_{t-1} \epsilon_{it+j}\alpha_{2i} +_{t-2} \epsilon_{it+j-1}\alpha_{2i}\rho_{1i} \\ = \rho_{1i} y_{it-1} + X_{it}\alpha_i - X_{it-1}\alpha_i\rho_{1i} + v_{it} \end{aligned} \qquad (4.19)$$

where $X_{it}$ and $\alpha_i$ are defined after 4.9 and now $v_{it} = \eta_{it} -_{t-1}\epsilon_{it+j}\alpha_{2i} +_{t-2}\epsilon_{it+j-1}\alpha_{2i}\rho_{1i}$. Equation 4.19 is nonlinear in coefficients because of the introduction of $\rho_{1i}$. Again, $X_{it}$ can in general include more than one expectational variable and more than one value of $j$ per variable.

Given a set of first stage regressors, equation 4.19 can be estimated by 2SLS. The estimates are obtained by minimizing

$$S_i = v_i' Z_i (Z_i' Z_i)^{-1} Z_i' v_i = v_i' D_i v_i \qquad (4.20)$$

4.20 is just 4.3 rewritten for the error term in 4.19. A necessary condition for consistency is that $Z_{it}$ and $v_{it}$ be uncorrelated, which means that $Z_{it}$ must be uncorrelated with $\eta_{it}$, $_{t-1}\epsilon_{it+j}$, and $_{t-2}\epsilon_{it+j-1}$. In order to insure that $Z_{it}$ and $_{t-2}\epsilon_{it+j-1}$ are uncorrelated, $Z_{it}$ must not include any variables that are not known as of the beginning of period $t-1$. This is an important additional restriction in the autoregressive case.[4]

In the general nonlinear case 4.20 (or 4.3) can be minimized using a general purpose optimization algorithm. In the particular case considered here, however, a simple iterative procedure can be used, where one iterates between

[4]There is a possibly confusing statement in Cumby, Huizinga, and Obstfeld (1983), p. 341, regarding the movement of the instrument set backward in time. The instrument set must be moved backward in time as the order of the autoregressive process increases. It need not be moved backward as the order of the moving average process increases due to an increase in $j$.

estimates of $\alpha_i$ and $\rho_{1i}$. Minimizing $v_i' D_i v_i$ with respect to $\alpha_i$ and $\rho_{1i}$ results in the following first order conditions:

$$\hat{\alpha}_i = [(X_i - X_{i-1}\hat{\rho}_{1i})' D_i (X_i - X_{i-1}\hat{\rho}_{1i})]^{-1} (X_i - X_{i-1}\hat{\rho}_{1i})' D_i (y_i - y_{i-1}\hat{\rho}_{1i}) \tag{4.21}$$

$$\hat{\rho}_{1i} = \frac{(y_{i-1} - X_{i-1}\hat{\alpha}_i)' D_i (y_i - X_i\hat{\alpha}_i)}{(y_{i-1} - X_{i-1}\hat{\alpha}_i)' D_i (y_{i-1} - X_{i-1}\hat{\alpha}_i)} \tag{4.22}$$

where the $-1$ subscript denotes the vector or matrix of observations lagged one period. Equations 4.21 and 4.22 can easily be solved iteratively. Given the estimates $\hat{\alpha}_i$ and $\hat{\rho}_{1i}$ that solve 4.21 and 4.22, one can compute the 2SLS estimate of $v_i$, which is

$$\hat{v}_i = y_i - y_{i-1}\hat{\rho}_{1i} - X_i\hat{\alpha}_i + X_{i-1}\hat{\alpha}_i\hat{\rho}_{1i} \tag{4.23}$$

Regarding Hansen's estimator, given $\hat{v}_i$, one can compute $M_i$ in one of the number of possible ways. These calculations simply involve $\hat{v}_i$ and $Z_i$. Given $M_i$, Hansen's estimates of $\alpha_i$ and $\rho_{1i}$ are obtained by minimizing[5]

$$SS_i = v_i' Z_i M_i^{-1} Z_i' v_i = v_i' C_i v_i \tag{4.24}$$

Minimizing 4.24 with respect to $\alpha_i$ and $\rho_{1i}$ results in the first order conditions 4.21 and 4.22 with $C_i$ replacing $D_i$. The estimated covariance matrix is

$$T(G_i' C_i G_i)^{-1} \tag{4.25}$$

where $G = (X_i - X_{i-1}\hat{\rho}_{1i} \quad y_{i-1} - X_{i-1}\hat{\alpha}_i)$.

To summarize, Hansen's method in the case of a first order autoregressive structural error consists of: 1) choosing $Z_{it}$ so that it does not include any variables not known as of the beginning of period $t-1$, 2) solving 4.21 and 4.22, 3) computing $\hat{v}_i$ from 4.23, 4) computing $M_i$ in one of the number of possible ways using $\hat{v}_i$ and $Z_i$, and 5) solving 4.21 and 4.22 with $C_i$ replacing $D_i$.

## 4.4 Two Stage Least Absolute Deviations[6]

Another single equation estimator that is of interest to consider is two stage least absolute deviations (2SLAD). This estimator is used for comparison purposes in Chapter 8. The following is a brief review of it.

[5]The estimator that is based on the minimization of 4.24 is also the 2S2SLS estimator of Cumby, Huizinga, and Obstfeld (1983).

[6]See Fair (1984), Sections 6.3.6 and 6.5.4, for a more detailed discussion of the two stage least absolute deviations estimator.

It is assumed for the 2SLAD estimator that the model in 4.1 can be written:

$$y_{it} = h_i(y_t, x_t, \alpha_i) + u_{it}, \quad (i = 1, \ldots, n), \quad (t = 1, \ldots, T) \tag{4.26}$$

where in the $i$th equation $y_{it}$ appears only on the left hand side.

Let $\hat{y}_i = D_i y_i$ and $\hat{h}_i = D_i h_i$, where, as above, $D_i = Z_i(Z_i' Z_i)^{-1} Z_i'$, where $Z_i$ is a matrix of first stage regressors. There are two ways of looking at the 2SLAD estimator. One is that it minimizes

$$\sum_{t=1}^{T} \left| \hat{y}_{it} - \hat{h}_{it} \right| \tag{4.27}$$

and the other is that it minimizes

$$\sum_{t=1}^{T} \left| y_{it} - \hat{h}_{it} \right| \tag{4.28}$$

Amemiya (1982) has proposed minimizing

$$\sum_{t=1}^{T} \left| q y_{it} + (1-q)\hat{y}_{it} - \hat{h}_{it} \right| \tag{4.29}$$

where $q$ is chosen ahead of time by the investigator. The estimator that is based on minimizing 4.29 will be called the 2SLAD estimator. For the computational results in Chapter 8, $q = .5$ has been used.

The 2SLAD estimator weights large outliers less than does 2SLS, and so it is less sensitive to these outliers. It is a robust estimator in the sense that its properties are less sensitive to deviations of the distributions of the error terms from normality than are the properties of 2SLS.

## 4.5 Chi-Square Tests

Many single equation tests are simply of the form of adding a variable or a set of variables to an equation and testing whether the addition is statistically significant. Let $S_i^{**}$ denote the value of the minimand before the addition, let $S_i^*$ denote the value after the addition, and let $\hat{\sigma}_{ii}$ denote the estimated variance of the error term after the addition. Under fairly general conditions, as discussed in Andrews and Fair (1988), $(S_i^{**} - S_i^*)/\hat{\sigma}_{ii}$ is distributed as $\chi^2$ with $k$ degrees of freedom, where $k$ is the number of variables added. For the 2SLS estimator the minimand is defined in equation 4.3, i.e., $S_i = u_i' D_i u_i$.

For Hansen's estimator the minimand is defined in equation 4.24, i.e., $SS_i = v_i' C_i v_i$. In this case $(SS_i^{**} - SS_i^*)/T$ is distributed as $\chi^2$, where $T$ is the number of observations. When performing this test the $M_i$ matrix that is used in the construction of $C_i$ must be the same for both estimates. For the results in Chapter 5, $M_i$ was always estimated using the residuals for the unrestricted case (i.e., using the residuals from the equation with the additions).

The following is a list of tests of single equations that can be made by adding various things to the equations and performing $\chi^2$ tests.

**Dynamic Specification**

Many macroeconomic equations include the lagged dependent variable and other lagged endogenous variables among the explanatory variables. A test of the dynamic specification of a particular equation is to add *further* lagged values to the equation and see if they are significant. For equation 4.2, for example, one could add the lagged value of $y_i$ if the lagged value is not already included in $X_i$ and the lagged values of the variables in $X_i$. (If the lagged value of $y_i$ is in $X_i$, then the value of $y_i$ lagged twice would be added for the test.) Hendry, Pagan, and Sargan (1984) show that adding these lagged values is quite general in that it encompasses many different types of dynamic specifications. Therefore, adding the lagged values and testing for their significance is a test against a fairly general dynamic specification.

**Time Trend**

Long before units roots and cointegration became popular, model builders worried about picking up spurious correlation from common trending variables. One check on whether the correlation might be spurious is to add a time trend to the equation. If adding a time trend to the equation substantially changes some of the coefficient estimates, this is cause for concern. A simple test is to add the time trend to the equation and test if this addition is significant.

**Serial Correlation of the Error Term**

As noted in Section 4.1, if the error term in an equation follows an autoregressive process, the equation can be transformed and the coefficients of the autoregressive process can be estimated along with the structural coefficients. Even if, say, a first order process has been assumed and the first order coefficient estimated, it is still of interest to see if there is serial correlation of the (transformed) error term left. This can be done by assuming a more general process for the error term and testing its significance. For the results in Chapters 5 and 6 a fourth order process was used. If the addition of a fourth order process over, say, a first order process results in a significant increase in explanatory power, this is evidence that the serial correlation properties of the error term have not been properly accounted for.

### Other Explanatory Variables

Variables can obviously be added to an equation and their statistical significance tested. This is done for the equations in the next two chapters. If a variable or set of variables that one does not expect from the theory to belong in the equation is significant, this is evidence against the theory.

Variables can also be added that others have found to be important explanatory variables in similar contexts. For example, Friedman and Kuttner (1992), (1993) have found the spread between the six month commercial paper rate and the six month Treasury bill rate is significant in explaining real GNP in a vector autoregressive system. If the spread is significant in explaining real GNP, then it should be in explaining some of the components of real GNP. It is thus of interest to add the spread to equations explaining consumption and investment to see if it has independent explanatory power. This is done in the next chapter for some of the equations in the US model.[7]

### Leads[8]

Adding values led one or more periods and using Hansen's method for the estimation is a way of testing the hypothesis that expectations are rational. Consider the example in equation 4.7 above, and consider testing the RE hypothesis against the simpler alternative that $E_{t-1}X_{2it+j}$ is only a function of $X_{2it}$ and $X_{2it-1}$, where both of these variables are assumed to be known at the time the expectation is made. Under the simpler alternative, $X_{2it}$ and $X_{2it-1}$ are added as explanatory variables to 4.7. Under the RE alternative, $X_{2it+j}$ is added as an explanatory variable, and the equation is estimated using Hansen's method. A test of the RE hypothesis is thus to add $X_{2it+j}$ to the equation with $X_{2it}$ and $X_{2it-1}$ included and test the hypothesis that

[7]The six month commercial paper rate and the six month Treasury bill rate are not variables in the US model, and they are not presented in Appendix A. The data are available from the Federal Reserve.

[8]The material in this subsection is taken from Fair (1993b), Section 3.

the coefficient of $X_{2it+j}$ is zero. The $Z_{it}$ vector used for Hansen's method should include the predetermined variables in $X_{1it}$ in 4.7—including $X_{2it}$ and $X_{2it-1}$—plus other variables assumed to be in the agents' information sets.[9] The test is really whether these other variables matter. If agents do not use more information than that contained in the predetermined variables in $X_{1it}$ in forming their expectation of $X_{2ti+j}$, then the use of the variables in $Z_{it}$ as first stage regressors for $X_{2it+j}$ adds nothing not already contained in the predetermined variables in $X_{1it}$.

The test of the RE hypothesis is thus to add variable values led one or more periods to an equation with only current and lagged values and estimate the resulting equation using Hansen's method. If the led values are not significant, this is evidence against the RE hypothesis. It means essentially that the extra variables in $Z_{it}$ do not contribute significantly to the explanatory power of the equation.

An implicit assumption behind this test is that $Z_{it}$ contains variables other than the predetermined variables in $X_{1it}$. If, say, the optimal predictor of $X_{2it+j}$ were solely a function of $X_{2it}$ and $X_{2it-1}$, then the above test would not be appropriate. In this case the traditional approach is consistent with the RE hypothesis, and there is nothing to test. The assumption that $Z_{it}$ contains many variables is consistent with the specification of most macroeconometric models, where the implicit reduced form equations for the endogenous variables contain a large number of variables. This assumption is in effect maintained throughout this book. The tests in Chapters 5 and 6 have nothing to say about cases in which there is a very small number of variables in $Z_{it}$.

As an example of the test, consider the wage variable $W$ in the consumption equation 1.4 in Chapter 1. Assume that $W_t$ is known, where $t$ is period 1. The wage variables in equation 1.4 are then $W_t$, $E_{t-1}W_{t+1}$, $E_{t-1}W_{t+2}$, etc. If

[9]Remember that $X_{2it}$ is assumed to be known at the time the expectations are made, which is the reason for treating it as predetermined. In practice, a variable like $X_{2it}$ is sometimes taken to be endogenous, in which case it is not part of the $Z_{it}$ vector. When a variable like $X_{2it}$ is taken to be endogenous, an interesting question is whether one can test the hypothesis that agents know it at the time they make their expectations as opposed to having only a rational expectation of it. It is not possible to test this if the $Z_{it}$ vector used for the 2SLS method is the same vector used by the agents in forming their expectations. It would, however, be possible to test this hypothesis if there were some contemporaneous exogenous variables in the model that agents forming rational expectations do not know at the time they make their forecasts. These variables are appropriate first stage regressors for 2SLS (since they are exogenous), but they are not used by agents. In practice, however, this is likely to be a small difference upon which to base a test, and no attempt is made here to do so. The focus here is on values dated $t + 1$ and beyond.

agents use only current and lagged values of $W$ in forming expectations of future values of $W$, then candidates for explanatory variables are $W_t$, $W_{t-1}$, $W_{t-2}$, etc. Under the RE hypothesis, on the other hand, agents use $Z_{it}$ in forming their expectations for periods $t+1$ and beyond, and candidates for explanatory variables are $W_{t+1}$, $W_{t+2}$, etc., with Hansen's method used for the estimation. The test is to test for the joint significance of the led values.

The test proposed here is quite different from Hendry's (1988) test of expectational mechanisms. Hendry's test requires one to postulate the expectation generation process, which is then examined for its constancy across time. If the structural equation that contains the expectations is constant but the expectations equations are not, this refutes the expectations equations. As noted above, for the test proposed here $Z_{it}$ need not contain all the variables used by agents in forming their expectations, and so the test does not require a complete specification of the expectations generation process. The two main requirements are only that $Z_{it}$ be correlated with $X_{2it+j}$ but not with ${}_{t-1}\epsilon_{it+j}$.

## 4.6 Stability Tests

One of the most important issues to examine about an equation is whether its coefficients change over time, i.e., whether the structure is stable over time. A common test of structural stability is to pick a date at which the structure is hypothesized to have changed and then test the hypothesis that a change occurred at this date. In the standard linear regression model this is an F test, usually called the Chow test. More general settings are considered in Andrews and Fair (1988).

One test in the more general setting is simply the $\chi^2$ test discussed in the previous section, where $S_i^{**}$ is the value of the minimand under the assumption of no structural change and $S_i^*$ is the value of the minimand under the assumption that the change occurred at the specified date. Assume, for example, that the estimation period is from 1 through $T$ and that the specified date of the structural change is $T^*$. Assume also that the equation is estimated by 2SLS. Computing the $\chi^2$ value in this case requires estimating the equation for three periods: 1 through $T^*$, $T^*+1$ through $T$, and 1 through $T$. Let $S_i^{(1)}$ be the value of the minimand in 4.3 for the first estimation period, and let $S_i^{(2)}$ be the value for the second estimation period. Then $S_i^* = S_i^{(1)} + S_i^{(2)}$. $S_i^{**}$ is the value of the minimand in 4.3 that is obtained when the equation is estimated over the full estimation period. When estimating over the full period, the $Z_i$ matrix used for the full period must be the union of the matrices used for the two subperiods in order to make $S_i^{**}$ comparable to $S_i^*$. This means that for each first stage regressor $Q_{it}$, two variables must be used in $Z_i$ for the full estimation period, one that is equal to $Q_{it}$ for the first subperiod and zero otherwise and one that is equal to $Q_{it}$ for the second subperiod and zero otherwise. If this is done, then the $\chi^2$ value is $(S_i^{**} - S_i^*)/\hat{\sigma}_{ii}$, where $\hat{\sigma}_{ii}$ is equal to the sum of the sums of squared residuals from the first and second estimation periods divided by $T - 2k_i$, where $k_i$ is the number of estimated coefficients in the equation.

Recently, Andrews and Ploberger (1994) have proposed a class of tests that does not require that the date of the structural change be chosen *a priori*. Let the overall sample period be 1 through $T$. The hypothesis tested is that a structural change occurred between observations $T_1$ and $T_2$, where $T_1$ is an observation close to 1 and $T_2$ is an observation close to $T$. If the time of the change (if there is one) is completely unknown, Andrews and Ploberger suggest taking $T_1$ very close to 1 and $T_2$ very close to $T$. This puts little restriction on the time of the change. If, on the other hand, the time of the change is known to lie in a narrower interval, the narrower interval should be used to maximize power. One of the main advantages of the Andrews-Ploberger tests is that they have nontrivial power asymptotically and have been designed to have certain optimality properties.

The particular Andrews-Ploberger test used here is easy to compute. The test is carried out as follows:

1. Compute the $\chi^2$ value for the hypothesis that the change occurred at observation $T_1$. This requires estimating the equation three times—once each for the estimation periods 1 through $T_1 - 1$, $T_1$ through $T$, and 1 through $T$. Denote this value as $\chi^{2(1)}$.[10]

2. Repeat step 1 for the hypothesis that the change occurred at observation $T_1 + 1$. Denote this $\chi^2$ value as $\chi^{2(2)}$. Keep doing this through the hypothesis that the change occurred at observation $T_2$. This results in $N = T_2 - T_1 + 1$ $\chi^2$ values being computed—$\chi^{2(1)}, \ldots, \chi^{2(N)}$.

3. The Andrews-Ploberger test statistic (denoted $AP$) is

$$AP = \log[(e^{\frac{1}{2}\chi^{2(1)}} + \ldots + e^{\frac{1}{2}\chi^{2(N)}})/N] \tag{4.30}$$

   In words, the $AP$ statistic is a weighted average of the $\chi^2$ values, where there is one $\chi^2$ value for each possible split in the sample period between observations $T_1$ and $T_2$.

[10] This $\chi^2$ value is computed in the regular way as discussed above.

**Table 4.1**
**Critical Values for the AP Statistic**
**for $\lambda = 2.75$**

| No.of coefs. | 5% | 1% | No.of coefs. | 5% | 1% |
|---|---|---|---|---|---|
| 1 | 2.01 | 3.36 | 8 | 8.22 | 10.23 |
| 2 | 3.07 | 4.69 | 9 | 9.01 | 11.20 |
| 3 | 4.00 | 5.62 | 10 | 9.55 | 12.14 |
| 4 | 4.95 | 7.00 | 11 | 10.33 | 12.73 |
| 5 | 5.80 | 7.65 | 12 | 11.03 | 13.43 |
| 6 | 6.59 | 8.72 | 13 | 11.62 | 14.47 |
| 7 | 7.31 | 9.50 | 14 | 12.37 | 15.20 |

Asymptotic critical values for $AP$ are presented in Tables I and II in Andrews and Ploberger (1994). The critical values depend on the number of coefficients in the equation and on a parameter $\lambda$, where in the present context $\lambda = [\pi_2(1 - \pi_1)]/[\pi_1(1 - \pi_2)]$ , where $\pi_1 = (T_1 - .5)/T$ and $\pi_2 = (T_2 - .5)/T$.

If the $AP$ value is significant, it may be of interest to examine the individual $\chi^2$ values to see where the maximum value occurred. This is likely to give one a general idea of where the structural change occurred even though the $AP$ test does not reveal this in any rigorous way.

Since the $AP$ test is used in the next two chapters, it will be useful to give a few critical values. For the work in the next chapter the basic sample period is 1954:1–1993:2, and for the stability tests $T_1$ was taken to be 1970:1 and $T_2$ was taken to be 1979:4. This choice yields a value of $\lambda$ of 2.75. Table 4.1 presents the 5 percent and 1 percent asymptotic critical values for this value of $\lambda$ and various values of the number of estimated coefficients in the equations. These values are interpolated from Table I in Andrews and Ploberger (1994).

Although the values in Table 4.1 are for just one particular value of $\lambda$ (namely, 2.75), Andrews and Ploberger's Table I shows that the critical values are not very sensitive to different values of $\lambda$. The above critical values are thus approximately correct for different choices of $T_1$ and $T_2$ than the one made here.

## 4.7 Tests of Age Distribution Effects[11]

A striking feature of post war U.S. society has been the baby boom of the late 1940s and the 1950s and the subsequent falling off of the birth rate in the 1960s. The number of births in the United States rose from 2.5 million in 1945 to 4.2 million in 1961 and then fell back to 3.1 million in 1974. This birth pattern implies large changes in the percentage of prime age (25–54) people in the working age (16+) population. In 1952 this percentage was 57.9, whereas by 1977 it had fallen to 49.5. Since 1980 the percentage of prime aged workers has risen sharply as the baby boomers have begun to pass the age of 25.

As noted in Chapter 1, an important issue in macroeconomics is whether the coefficients of macroeconomic equations change over time as other things change. The Lucas critique focuses on policy changes, but other possible changes are changes in the age distribution of the population. This section discusses a procedure for examining the effects of the changes in the U.S. population age distribution on macroeconomic equations. The procedure is as follows.

Divide the population into $J$ age groups. Let $D1_{ht}$ be 1 if individual $h$ is in age group 1 in period $t$ and 0 otherwise; let $D2_{ht}$ be 1 if individual $h$ is in age group 2 in period $t$ and 0 otherwise; and so on through $DJ_{ht}$. Consider equation $i$ in 4.2, an equation that is linear in coefficients. Let equation $i$ for individual $h$ be:

$$y_{hit} = X_{hit}\alpha_i + \beta_{0i} + \beta_{1i}D1_{ht} + \ldots + \beta_{Ji}DJ_{ht} + u_{hit}$$

$$(h = 1, \ldots, N_t), \quad (t = 1, \ldots, T) \tag{4.31}$$

where $y_{hit}$ is the value of variable $i$ in period $t$ for individual $h$ (e.g., consumption of individual $h$ in period $t$), $X_{hit}$ is a vector of explanatory variables excluding the constant, $\alpha_i$ is a vector of coefficients, and $u_{hit}$ is the error term. The constant term in the equation is $\beta_{0i} + \beta_{ji}$ for an individual in age group $j$ in period $t$. $N_t$ is the total number of people in the population in period $t$.

Equation 4.31 is restrictive because it assumes that $\alpha_i$ is the same across all individuals, but it is less restrictive than a typical macroeconomic equation, which also assumes that the constant term is the same across individuals. Given $X_{hit}$, $y_{hit}$ is allowed to vary across age groups in equation 4.31. Because most macroeconomic variables are not disaggregated by age groups, one cannot test for age sensitive $\alpha_i$'s. For example, suppose that one of the variables in $X_{hit}$

[11]The material in this section is taken from Fair and Dominguez (1991).

is $Y_{ht}$, the income of individual $h$ in period $t$. If the coefficient of $Y_{ht}$ is the same across individuals, say $\gamma_{1i}$, then $\gamma_{1i} Y_{ht}$ enters the equation, and it can be summed in the manner discussed in the next paragraph. If, on the other hand, the coefficient differs across age groups, then the term entering the equation is $\gamma_{11i} D1_{ht} Y_{ht} + \ldots + \gamma_{1Ji} DJ_{ht} Y_{ht}$. The sum of a variable like $D1_{ht} Y_{ht}$ across individuals is the total income of individuals in age group 1, for which data are not generally available. One is thus restricted to assuming that age group differences are reflected in different constant terms in equation 4.31.

Let $N_{jt}$ be the total number of people in age group $j$ in period $t$, let $y_{it}$ be the sum of $y_{hit}$, let $X_{it}$ be the vector whose elements are the sums of the corresponding elements in $X_{hit}$, and let $u_{it}$ be the sum of $u_{hit}$. (All sums are for $h = 1, \ldots, N_t$.) Given this notation, summing equation 4.31 yields:

$$y_{it} = X_{it}\alpha_i + \beta_{0i} N_t + \beta_{1i} N_{1t} + \ldots + \beta_{Ji} N_{Jt} + u_{it}, \quad (t = 1, \ldots, T) \quad (4.32)$$

If equation 4.32 is divided through by $N_t$, it is converted into an equation in per capita terms. Let $p_{jt} = N_{jt}/N_t$, and reinterpret $y_{it}$, the variables in $X_{it}$, and $u_{it}$ as being the original values divided by $N_t$. Equation 4.32 in per capita terms can then be written:

$$y_{it} = X_{it}\alpha_i + \beta_{0i} + \beta_{1i} p_{1t} + \ldots + \beta_{Ji} p_{Jt} + u_{it}, \quad (t = 1, \ldots, T) \quad (4.33)$$

A test of whether age distribution matters is simply a test of whether the $\beta_{1i}, \ldots, \beta_{Ji}$ coefficients in equation 4.33 are significantly different from zero.[12] If the coefficients are zero, one is back to a standard macroeconomic equation. Otherwise, given $X_{it}$, $y_{it}$ varies as the age distribution varies. Since the sum of $p_{jt}$ across $j$ is one and there is a constant in the equation, a restriction on the $\beta_{ji}$ coefficients must be imposed for estimation. In the estimation work below, the age group coefficients are restricted to sum to zero: $\sum_{j=1}^{J} \beta_{ji} = 0$. This means that if the distributional variables do not matter, then adding them to the equation will not affect the constant term.

### The Age Distribution Data

The age distribution data that are used in the next chapter are from the U.S. Bureau of the Census, *Current Population Reports,* Series P-25. The data from

[12] Stoker (1986) characterizes this test (that all proportion coefficients are zero) as a test of microeconomic linearity or homogeneity (that all marginal reactions of individual agents are identical). He shows that individual differences or more general behavioral nonlinearities will coincide with the presence of distributional effects in macroeconomic equations.

the census surveys, which are taken every ten years, are updated yearly using data provided by the National Center for Health Statistics, the Department of Defense, and the Immigration and Naturalization Service. The data are estimates of the total population of the United States, including armed forces overseas, in each of 86 age groups. Age group 1 consists of individuals less than 1 year old, age group 2 consists of individuals between 1 and 2 years of age, and so on through age group 86, which consists of individuals 85 years old and over. The published data are annual (July 1 of each year). Because the equations estimated below are quarterly, quarterly population data have been constructed by linearly interpolating between the yearly points.

Fifty five age groups are considered: ages 16, 17, ..., 69, and 70+. The "total" population, $N_t$, is taken to be the population 16+. In terms of the above notation, 55 $p_{jt}$ variables ($j = 1, \ldots, 55$) have been constructed, where the 55 variables sum to one for a given $t$.

### Constraints on the Age Coefficients

Since there are 55 $\beta_{ji}$ coefficients to estimate, some constraints must be imposed on them if there is any hope of obtaining sensible estimates. One constraint is that the coefficients sum to zero. Another constraint, which was used in Fair and Dominguez (1991), is that the coefficients lie on a second degree polynomial. The second degree polynomial constraint allows enough flexibility to see if the prime age groups behave differently from the young and old groups while keeping the number of unconstrained coefficients small. A second degree polynomial in which the coefficients sum to zero is determined by two coefficients, and so there are two unconstrained coefficients to estimate per equation. The two variables that are associated with two unconstrained coefficients will be denoted $AGE_{1t}$ and $AGE_{2t}$.

The variables $AGE_{1t}$ and $AGE_{2t}$ are as follows. First, the age variables enter equation $i$ as $\sum_{j=1}^{55} \beta_{ji} p_{jt}$, where $\sum_{j=1}^{55} \beta_{ji} = 0$. The polynomial constraint is

$$\beta_{ji} = \gamma_0 + \gamma_1 j + \gamma_2 j^2 \quad , \quad (j = 1, \ldots, 55) \tag{4.34}$$

where $\gamma_0$, $\gamma_1$, and $\gamma_2$ are coefficients to be determined.[13] The zero sum constraint on the $\beta_{ji}$'s implies that

$$\gamma_0 = -\,\gamma_1 \frac{1}{55} \sum_{j=1}^{55} j \;-\; \gamma_2 \frac{1}{55} \sum_{j=1}^{55} j^2 \tag{4.35}$$

[13] For ease of notation, no $i$ subscripts are used for the $\gamma$ coefficients.

The way in which the age variables enter the estimated equation is then

$$\gamma_1 AGE_{1t} + \gamma_2 AGE_{2t}$$

where

$$AGE_{1t} = \sum_{j=1}^{55} j p_{jt} \; - \; \frac{1}{55}(\sum_{j=1}^{55} j)(\sum_{j=1}^{55} p_{jt}) \tag{4.36}$$

and

$$AGE_{2t} = \sum_{j=1}^{55} j^2 p_{jt} \; - \; \frac{1}{55}(\sum_{j=1}^{55} j^2)(\sum_{j=1}^{55} p_{jt}) \tag{4.37}$$

Given the estimates of $\gamma_1$ and $\gamma_2$, the 55 $\beta_{ji}$ coefficients can be computed. This technique is simply Almon's (1965) polynomial distributed lag technique, where the coefficients that are constrained are the coefficients of the $p_{jt}$ variables ($j = 1, \ldots, 55$) rather than coefficients of the lagged values of some variable.

One test of whether age distribution matters is thus to add $AGE_{1t}$ and $AGE_{2t}$ to the equation and test if the two variables are jointly significant.

For the work in the next chapter a different set of constraints was imposed on the $\beta_{ji}$ coefficients. The population 16+ was divided into four groups (16–25, 26–55, 56–65, and 66+) and it was assumed that the coefficients are the same within each group. Given the constraint that the coefficients sum to zero, this leaves three unconstrained coefficients to estimate. Let $P1625$ denote the percent of the 16+ population aged 16–25, and similarly for $P2655$, $P5665$, and $P66+$. Let $\gamma_0$ denote the coefficient of $P1625$ in the estimated equation, $\gamma_1$ the coefficient of $P2655$, $\gamma_2$ the coefficient of $P5665$, and $\gamma_3$ the coefficient of $P66+$, where $\gamma_0+\gamma_1+\gamma_2+\gamma_3 = 0$. The summation constraint can be imposed by entering three variables in the estimated equation: $AG1 = P2655 - P1625$, $AG2 = P5665 - P1625$, and $AG3 = (P66+) - P1625$. $AG1$, $AG2$, and $AG3$ are variables in the US model. The coefficient of $AG1$ in an equation is $\gamma_1 - \gamma_0$, the coefficient of $AG2$ is $\gamma_2 - \gamma_0$, and the coefficient of $AG3$ is $\gamma_3 - \gamma_0$. From the estimated coefficients for $AG1$, $AG2$, and $AG3$ and the summation constraint, one can calculate the four $\gamma$ coefficients.

Imposing the constraints in the manner just described has an advantage over imposing the quadratic constraint of allowing more flexibility in the sense that three unconstrained coefficients are estimated instead of two. Also, I have found that the quadratic constraint sometimes leads to extreme values of $\beta_{ji}$ for the very young and very old ages. The disadvantage of the present approach over the quadratic approach is that the coefficients are not allowed to change within the four age ranges.