INTERNATIONAL ECONOMIC REVIEW
Vol. 28, No. 3, October, 1987

HYPOTHESIS TESTING WITH EFFICIENT METHOD OF MOMENTS ESTIMATION*

BY WHITNEY K. NEWEY AND KENNETH D. WEST[1]

1. INTRODUCTION

It is well known that a variety of prodedures are available to test a hypothesis in a model estimated by maximum likelihood. In maximum likelihood models one of the trio of Wald, likelihood ratio, and Lagrange multiplier test statistics is usually used to test a hypothesis (see Engle [1984] for a survey). In some cases a minimum chi-square hypothesis test (see Rothenberg [1973]) may also be convenient. All these statistics are asymptotically equivalent, so that insofar as asymptotic properties are concerned, the researcher has the freedom to choose a test statistic on the basis of computational convenience. Of course, the statistics may behave differently in small samples (Berndt and Savin [1977], Evans and Savin [1982]). The existence of a variety of test statistics then also gives the researcher the freedom to choose a test statistic that appears, from theoretical or Monte Carlo studies, to have better properties.

Less well known is the extent to which a similar sort of flexibility is available in models not estimated by maximum likelihood. The availability of a variety of procedures in the context of other estimation methods is of potential interest to many researchers. In many rational expectations models, for example, the computational burden of maximum likelihood is enormous, since disturbances are serially correlated and heteroskedastic in complicated ways. Instrumental variables procedures are therefore often used to estimate these models (e.g., West [1987a, 1987b, 1987c]).

Our purpose here is to present the form of a variety of test statistics for hypothesis tests in models that are estimated by using the efficient generalized method of moments (GMM; see Hansen [1982]) estimator. These estimators include many of those encountered in practice: ordinary least squares, two stage least squares, three stage least squares, quasi maximum likelihood, and versions of these for nonlinear and non i.i.d. environments. Others have presented versions of test statistics that use the efficient GMM estimator in special cases (Gallant and Jorgenson [1979]) and for special hypothesis tests (Eichenbaum, Hansen, and Singleton [1984]). Our extension to the general non i.i.d. case has a wide variety of uses in testing parametric hypotheses in rational expectations models (e.g., the West papers cited above) and in cross section and panel data

* Manuscript received December, 1985; revised September, 1986.
[1] We thank the National Science Foundation for financial support, and the referees for helpful comments.

models where heteroskedasticity may be present (e.g. Holtz-Eakin, Newey, and Rosen [1987]).

In section 2 we present analogues of the Wald, likelihood ratio, Lagrange multiplier, and minimum chi-square test statistics. We show that they are mutually asymptotically equivalent in an environment that allows for disturbances that are autocorrelated and heteroskedastic, with the form of autocorrelation and and heteroskedasticity possibly unknown, and for instruments that are not strictly econometrically exogenous. Section 3 discusses some interesting special cases, including quasi-maximum likelihood and the case where the econometric model is linear in the parameters and/or constraints. The working paper version of this paper, Newey and West [1985], contains proofs of our theorems and has an empirical example.

2. THE TEST STATISTICS AND THEIR ASYMPTOTIC DISTRIBUTION

Many recent models have been estimated by the techniques developed by Hansen [1982], Hansen and Singleton [1982], and Cumby, Huizinga, and Obstfeld [1983]. Examples in our own work were given above. The models in these and other papers imply an orthogonality condition

$$E[g(z_t, b_0)] = 0, \tag{2.1}$$

where b_0 is a $q \times 1$ vector of parameters, z_t is a $p \times 1$ data vector, and $g(z, b)$ is a $r \times 1$ vector of functions of the data and parameters. A common example is $g(z_t, b) = x_t \cdot u(z_t, b)$. where x_t is a $r \times 1$ vector of instrumental variables and $u(z_t, b)$ is a residual. In this case equation (2.1) represents the usual orthogonality condition for instrumental variables and a residual.

When equation (2.1) is correct, the sample moment

$$g_T(b) = \textstyle\sum_{t=1}^{T} g(z_t, b)/T \tag{2.2}$$

should be close to zero when evaluated at $b = b_0$. It is therefore reasonable to estimate b_0 by choosing $\hat{b}_T$ so that $g_T(b)$ is close to zero, i.e., by solving

$$\min_b g_T(b)' W_T g_T(b), \tag{2.3}$$

where W_T is a positive semi-definite matrix. The resulting estimator is a member of the class of generalized method of moments (GMM) estimators that has been considered by Amemiya [1974], Hansen [1982], and Burguete, Gallant, and Souza [1982].

2.1. *The Test Statistics.* It is often the case that it is desired to test hypotheses concerning the vector of parameters b_0. In this paper, we will consider a general null hypothesis of the form

$$H_0\colon a(b) = 0, \tag{2.4}$$

where $a(b)$ is a $s \times 1$ vector. We will consider test statistics of H_0 that use three estimators of b_0. One of these estimators is the optimal unrestricted GMM estimator. As shown by Hansen [1982], the estimator that is optimal in the sense of having an asymptotic covariance matrix that is as small as possible in the class of GMM estimators that do not impose equation (2.4), is obtained from (2.3) when $W_T = V_T^{-1}$; V_T is a consistent estimator of the asymptotic covariance matrix V of $\sqrt{T} g_T(b_0)$. With this choice of W_T, the unrestricted estimator $\hat{b}_T$ will be the solution to

$$\min_b J_T(b), \tag{2.5}$$

where $J_T(b) \equiv g_T(b)' V_T^{-1} g_T(b)$.

To form $J_T(b)$ it is necessary to have a consistent estimator of the matrix V. One estimator of V that is simple, positive semi-definite, and consistent under quite general conditions has recently been presented by Newey and West [1987]. This estimator is

$$V_T^* = \hat{\Omega}_0 + \textstyle\sum_{j=1}^{m} w(j, m)[\hat{\Omega}_j + \hat{\Omega}_j'], \tag{2.6}$$

where $w(j, m) \equiv 1 - [j/(m+1)]$, $\hat{\Omega}_j \equiv \sum_{t=j+1}^{T} g(z_t, b_T^*) g(z_{t-j}, b_T^*)'/T$ is the jth sample autocovariance of $g(z_t, b_T^*)$ and b_T^* is an initial consistent estimator of b_0, which can be obtained by solving equation (2.3) for a choice of W_T other than V_T^{-1}. The number of autocovariances m is chosen as a suitable function of T. See Newey and West [1987] for details and for discussion of other appropriate estimators of V.

The second estimator that is useful for forming test statistics is the optimal restricted GMM estimator, obtained by minimizing $J_T(b)$ subject to the restrictions on b that are implied by the null hypothesis. Let the restricted estimator $\tilde{b}_T$ solve

$$\min_b J_T(b) \text{ subject to } a(b) = 0. \tag{2.7}$$

The third estimator that will be used in our test statistics is the minimum chi-square estimator. Let

$$G = E[\partial g_T(b_0)/\partial b], \quad \hat{G}_T = \partial g_T(\hat{b}_T)/\partial b,$$
$$Q = G' V^{-1} G, \quad \hat{Q}_T = \hat{G}_T' V_T^{-1} \hat{G}_T.$$

As shown by Hansen [1982], Q^{-1} is the asymptotic covariance matrix of the optimal unrestricted GMM estimator $\hat{b}_T$, so that $\hat{Q}_T^{-1}$ will be a consistent estimator of the asymptotic convariance matrix of $\hat{b}_T$. Then the minimum chi-square estimator (e.g., Rothenberg [1973]) based on $\hat{b}_T$, which we will denote by $\bar{b}_T$, is given by the solution to

$$\min_b (\hat{b}_T - b)' \hat{Q}_T (\hat{b}_T - b) \text{ subject to } a(b) = 0. \tag{2.8}$$

To present the test statistics some additional notation is required. Let

$$A = \partial a(b_0)/\partial b, \qquad \hat{A} = \partial a(\hat{b}_T)/\partial b,$$
$$\tilde{G}_T = \partial g_T(\tilde{b}_T)/\partial b, \quad \tilde{Q}_T = \tilde{G}_T' V_T^{-1} \tilde{G}_T.$$

The test statistics we consider are

$$(2.9)\qquad \begin{aligned} W &= Ta(\hat{b}_T)'[\hat{A}\hat{Q}_T^{-1}\hat{A}']^{-1}a(\hat{b}_T), \\ D &= T[J_T(\tilde{b}_T) - J_T(\hat{b}_T)] \\ LM &= T[g_T(\tilde{b}_T)'V_T^{-1}\tilde{G}_T]\tilde{Q}_T^{-1}[\tilde{G}_T'V_T^{-1}g_T(\tilde{b}_T)], \\ MC &= T[\hat{b}_T - \bar{b}_T]'\hat{Q}_T[\hat{b}_T - \bar{b}_T]. \end{aligned}$$

The statistic W is the usual Wald statistic. Its principal advantage is that it only requires the unconstrained estimator to compute it. Its principal disadvantages are that it is not invariant to reparameterization and in some contexts (see Hausman and McFadden [1984] for a maximum likelihood example) its finite sample distribution is not as well approximated by the asymptotic approximation as are the distributions of other statistics.

The D statistic is made up of an appropriately normalized difference of the restricted and unrestricted objective functions for the efficient GMM estimator. The principal disadvantage of this test statistic is that it requires two minimizations to compute. On the other hand, once these minimizations have been accomplished it is often trivial to calculate, since $J_T(\hat{b}_T)$ and $J_T(\tilde{b}_T)$ are often computed during the course of the minimization. A statistic like D has previously been presented in the context of instrumental variables estimation in an i.i.d. environment by Gallant and Jorgenson [1979]. Since D is the difference of restricted and unrestricted optimal objective functions this statistic is analogous to the likelihood ratio test and, in the linear model, to the test based on the difference of restricted and unrestricted sum of squared residuals.

The LM statistic is a Lagrange multiplier, or score, statistic. It uses the gradient of $J_T(b)$ evaluated at the restricted parameter estimator, like its maximum likelihood counterpart. Its principal advantage is that it only requires the restricted estimator.

The MC statistic is the minimum chi-square statistic for imposing the constraints on the unrestricted estimator. Like the D statistic, it requires two minimizations to compute.

It is important to note that all of the test statistics use the same estimator of V. The properties of the test statistics may be affected if different estimates of V are used to form them. If different estimators of V are used to form the restricted and unrestricted optimal GMM estimators then the D statistic need not be positive. Also, the numerical equality results for the linear case discussed below depend on using the same estimator of V to form each test statistic.

2.2. *Asymptotic Properties of the Test Statistics.* We begin with the list of assumptions we make. Our first is one of strict stationarity, ruling out the kind of moment drift that has been considered by White and Domowitz [1984]. The stationarity assumption is commonly maintained in rational expectations models. Our assumption 1 also rules out fixed regressors, which essentially means that the resulting inference is unconditional, rather than conditional on a set of exogenous variables.

ASSUMPTION 1. *The data* z_t, $(t=1,\ldots,T)$, *are random vectors that are the first* T *elements of a strictly stationary stochastic process* $\{z_t\}_{t=1}^{\infty}$, *and has a measurable joint density function* $f(z_1,\ldots,z_T,b_T)$ *with respect to a measure* $\prod_{t=1}^{T} v$, *where* v *is a* σ*-finite measure on* R^p.

Note that the data generating process is allowed to depend on the sample size through the parameter vector b_T, and that an extra T subscript on z_t has been suppressed for notational convenience.

A problem in deriving asymptotic approximations to the distribution of test statistics that allow for power calculations is that if the null hypothesis is false then with probability one the test will reject as the sample size grows. The classical method of making an approximate power calculation is to subject the data to Pitman [1949] drift, and in assumption 2 we adopt this approach.

ASSUMPTION 2. $b_T=b_0+\beta/\sqrt{T}$.

Our assumption 3 allows us to ignore complications that are introduced when the model is possibly misspecified. The effect of model misspecification on test statistics of parametric hypotheses is considered in Burguete, Gallant, and Souza [1982]. Thus, we will assume that the moment conditions used to obtain the GMM estimators are satisfied in the population.

ASSUMPTION 3. *For each* b *in* B, *a subset of* R^q, *the elements of* $g(z,b)$ *are measurable in* z *and*

$$\int g(z,b)f(z,b)dv=0.$$

To obtain the asymptotic distribution of the test statistics under a sequence of local alternatives it is useful to impose further regularity conditions and restrict the dependence across observations of the data generating process. The set of regularity conditions in assumptions 4–6 is by no means the weakest possible set of sufficient conditions for the asymptotic distribution theory, but it is relatively straightforward to check in parctice.

ASSUMPTION 4. *The vector* $g(z,b)$ *is continuously differentiable on* B, *almost everywhere* v, *and* $a(b)$ *is continuously differentiable on* B. *For each positive integer* $m\geq 2$ *the joint density* $f(z_1,z_m,b)$ *is continuous in* b *almost everywhere* $v\times v$. *Also,* b_0 *is an element of the interior of* B, *which is compact.*

For a matrix $C=[c_{ij}]$, let $|C|\equiv\max_{i,j}|c_{ij}|$.

ASSUMPTION 5. *There exist measurable functions* $\gamma_1(z)$ *and* $\gamma_2(z)$, *and* $d>1$, *such that almost everywhere* v, *and for all* b *in* B *and* $m\geq 2$,

$$|g(z,b)|^4\leq\gamma_1(z),\quad |\partial g(z,b)/\partial b|^2\leq\gamma_1(z)$$

$$f(z, b) \le \gamma_2(z), \quad f(z_1, z_m, b) \le \gamma_2(z_1)\,\gamma_2(z_m)$$

$$\int [\gamma_1(z)]^d \gamma_2(z) dv < +\infty, \quad \int \gamma_2(z) dv < +\infty.$$

To restrict the dependence across observations of the data it is useful to employ mixing conditions. Mixing conditions have recently been discussed in some detail by White and Domowitz [1984], where definitions and notation for the following assumption can be found.

ASSUMPTION 6. *There exist constants D, $\lambda > 0$ such that either*
(i) *For all b in B, $\{z_t\}_{t=1}^{\infty}$ is uniform mixing with*

$$\phi(m) \le Dm^{-\lambda}, \quad \lambda \ge \max\{2, d/(d-1)\};$$

(ii) *For all b in B, $\{z_t\}_{t=1}^{\infty}$ is strong mixing with*

$$\alpha(m) \le Dm^{-\lambda}, \quad \lambda \ge 2d/(d-1).$$

The next assumption is an identification condition that guarantees that the minimization problem (2.3) has unique solution asymptotically. Throughout $E[\cdot]$ denotes the expectation taken at $f(z, b_0)$.

ASSUMPTION 7. *For any b in B, $E[g(z_t, b)] = 0$ only if $b = b_0$. Also, G has rank q, V is nonsingular, and A has rank s.*

Under these assumptions we can give an explicit formula for V, which is

$$V = \Omega_0 + \textstyle\sum_{j=1}^{\infty} [\Omega_j + \Omega_j'],$$

where $\Omega_j = E[g(z_t, b_0) g(z_{t+j}, b_0)']$ (see Hansen [1982]). Also these assumptions are sufficient to guarantee consistency of the equation (2.6) estimator V_T^* of V, under the sequence of local alternatives.

THEOREM 1. *If assumptions 1–6 are satisfied, and $\sqrt{T}(b_T^* - b_0)$ is bounded in probability, then plim $V_T^* = V$.*

All our results are proved in the working paper version of this paper.

The following result gives the asymptotic distribution of the test statistics we have presented. The distribution will remain unchanged as long as any consistent estimator of V is used.

THEOREM 2. *If assumptions 1–7 are satisfied, plim $V_T = V$, and $a(b_0) = 0$, then W converges in distribution to a noncentral chi-squared distribution with s degrees of freedom and noncentrality parameter $\beta' A'(AQ^{-1}A')^{-1} A\beta$. In addition*

$$W - D = o_p(1), \quad W - LM = o_p(1), \quad W - MC = o_p(1), \tag{2.10}$$

where $o_p(1)$ denotes a random variable that converges in probability to zero.

Equation (2.10) says that all of the test statistics we have presented are asymptotically equivalent and distributed in large samples as a chi-squared random variable. Under the null hypothesis we have $\beta=0$, so that the noncentrality parameter is zero, and an approximate size for a hypothesis test that rejects the null hypothesis if a test statistic is too large can be computed in the usual way from the chi-squared distribution with s degrees of freedom.

It is important to note that the asymptotic equivalence above includes the D statistic. The difference between this result and the type of result found by Burguete, Gallant, and Souza [1982], where the "likelihood ratio" test is not asymptotically chi-squared, is that we have based the test statistics on the optimal GMM estimator, for which $W_T = V_T^{-1}$. For another choice of W_T it will not be the case that a test statistic based on the difference of unrestricted and restricted objective functions is asymptotically chi-squared. Also, the form of the LM statistic would be more complicated, and would in general involve the Jacobian matrix A of the constraints.

In the next section, we examine some interesting special cases, some of which result in numerical equivalence of the test statistics.

3. SPECIAL CASES

3.1. *The Exactly Identified Case.* We will refer to the case when the number of orthogonality conditions r is equal to the number of parameters q as the exactly identified case. We use this terminology since in instrumental variables estimation the number of orthogonality conditions is equal to the number of instrumental variables. So the exactly identified case occurs when the number of orthogonality conditions is equal to the number of parameters.

One thing that is special about the exactly identified case is that D and LM are equal. When $r=q$, note that $\tilde{G}_T$ will be nonsingular and $g_T(\hat{b}_T)$ will be zero with probability approaching one. (For expositional ease we will omit the qualifier "with probability approaching one" in the rest of this paper, although this qualifier applies to all our various results on numerical equivalence of test statistics.) From inspection of the formulae in equation (2.9) we see that $LM = D$.

PROPOSITION 1. *If the hypotheses of Theorem* 1 *are satisfied and* $r=q$, *then,* $LM = D$.

One interesting example of the exactly identified case occurs in the context of quasi-maximum likelihood estimation. Suppose that $\hat{b}_T$ is obtained by maximizing $S_T(b)$, where

$$S_T(b) \equiv \textstyle\sum_{t=1}^{T} s(z_t, b)/T, \tag{3.1}$$

and $s(z, b)$ is some function of the data and parameters that is not necessarily a log-likelihood. The general properties of such estimators in the independent observations case have been considered by White [1982b] and Burguete, Gallant, and Souza [1982], among others. When $\hat{b}_T$ satisfies the first order conditions to the problem we see that $\hat{b}_T$ is a GMM estimator with $g(z, b) = \partial s(z, b)/\partial b$. Thus, for this special case the restricted GMM estimator $\tilde{b}_T$ would be obtained by solving

$$\text{(3.2)} \qquad \min_b \left[\textstyle\sum_{t=1}^{T} \partial s(z_t, b)/\partial b\right]' V_T^{-1} \left[\textstyle\sum_{t=1}^{T} \partial s(z_t, b)/\partial b\right] \quad \text{s.t.} \quad a(b) = 0.$$

In the context of quasi-maximum likelihood estimation there is another natural way of imposing restrictions that gives an estimator that is different than $\tilde{b}_T$. One can obtain an estimator that imposes the restrictions by choosing $\dot{b}_T$ to solve

$$\text{(3.3)} \qquad \max_b S_T(b) \text{ subject to } a(b) = 0.$$

$\dot{b}_T$ is different than $\tilde{b}_T$, and consequently the Lagrange multiplier statistics obtained by White [1982b] and Burguete, Gallant, and Souza [1982] are different than LM. When one compares $LM = D$ with these other Lagrange multiplier statistics, it is apparent that LM is considerably simpler than the other statistics, since the latter involve the Jacobian of the constraints. There is an interesting explanation for this difference that involves the properties of the respective constrained estimators.

It is well-known that in many contexts (e.g., see Ferguson [1958] for the maximum likelihood case) minimum chi-square estimation is an asymptotically optimal method of imposing constraints on a vector of estimated parameters. It turns out that the restricted efficient GMM estimator $\tilde{b}_T$ is asymptotically equivalent to the minimum chi-square estimator $\bar{b}_T$, which was defined in equation (2.8), while the restricted quasi-maximum likelihood estimator is inefficient relative to the other two estimators.

PROPOSITION 2. *If the hypotheses of Theorem* 1 *are satisfied and* $a(b_0) = 0$, *then* $\sqrt{T}(\tilde{b}_T - \bar{b}_T) = o_p(1)$, *while* $\sqrt{T}(\dot{b}_T - \bar{b}_T) = o_p(1)$ *if and only if* $VG^{-1}A' = A'H$ *for some nonsingular matrix* H. *Otherwise* $\dot{b}_T$ *is asymptotically inefficient relative to* $\tilde{b}_T$.

One important case in which $\tilde{b}_T$ and $\dot{b}_T$ are asymptotically efficient is when $G = -V$. This occurs when $S_T(b)$ is the actual log likelihood of the data (so that $G = E\partial^2 s/\partial b \partial b'$ and $V = E(\partial s/\partial b)(\partial s/\partial b)'$). Under many circumstances likely to be encountered in practice, however, $\tilde{b}_T$ will be more efficient. A simple example is OLS with disturbances conditionally heteroskedastic. Chamberlain [1982] discusses this example in the context of minimum chi-square estimation. Thus, as happens in other contexts (see, e.g., the discussion of tests of overidentifying restrictions in Hansen [1982]), a simpler form of test statistic results when a relatively efficient estimator is used to form the statistic. Also, if the restricted estimator is of intrinsic interest, then the efficient GMM estimator and the minimum chi-square estimator are the estimators of choice, in terms of asymptotic efficiency. Chamberlain [1982] has made this point for minimum chi-square estimators.

3.2. *The Linear Case.* Interesting and useful simplifications occur when the orthogonality conditions are linear in the parameters. Suppose that

$$g(z, b) = g_1(z) - g_2(z)b. \tag{3.4}$$

The most common example of the linear case arises with linear instrumental variables estimators. For single equation and an instrument vector X_t a linear instrumental variables estimator would have orthogonality conditions $g(z_t, b) = X_t \cdot (y_t - W_t'b)$, where W_t is a $q \times 1$ vector of right hand side variables, so that $g_1(z) = X_t y_t$ and $g_2(z) = X_t W_t'$.

Note that among the statistics we have presented, only the Wald statistic involves the Jacobian of the constraints, and that in the linear case $\hat{G}_T = \tilde{G}_T$. The following result on the equality of the other test statistics follows from these facts.

PROPOSITION 3. *If the hypotheses of Theorem* 1 *are satisfied and equation* (3.4) *is satisfied, then* $D = LM = MC$.

When the constraint is also linear, all of the test statistics will be numerically equal.

PROPOSITION 4. *If the hypotheses of Theorem* 1 *are satisfied, equation* (3.4) *is satisfied, and* $a(b) = Ab + a$, *where* A *and* a *are a matrix and a vector of constants, respectively, then* $W = D = LM = MC$.

The exact numerical equivalence may seem puzzling in light of the inequality that ranks the maximum likelihood Wald, likelihood ratio, and Lagrange multiplier test statistics in the linear model with normal errors (Berndt and Savin [1977], Evans and Savin [1984]). Part of the explanation is that we assume that all statistics use the same estimate V_T of V. For the ordinary least squares estimator using the same V_T means that the same estimate of the disturbance variance is used. Suppose that in such a case we instead calculate W using the unconstrained estimate of the variance, LM using constrained estimate, and D in the obvious way from the constrained and unconstrained estimates. Then we would have $LM = D \leq W$, with LM and W identical to their maximum likelihood counterparts. The reason for equivalence of LM and D is that D is a "likelihood ratio" test that is based on the first order conditions for the least squares estimator rather than the log-likelihood itself, similar to the quasi-maximum likelihood case we discussed earlier.

The numerical equivalence noted in Proposition 4 means that in the linear case a choice of a test statistic depends largely on computational convenience. There is a particularly convenient way to form the D statistic to test a linear hypothesis in a linear equation that has serially uncorrelated instrumental variable and disturbance cross-products and (possibly) conditionally heteroskedastic errors, and which is estimated by instrumental variables. Let a linear equation be given by

$$y_t = W_t'b_0 + u_t, \quad (t=1,\ldots, T), \tag{3.5}$$

where W_t is a $q\times 1$ vector of right side variables and u_t is a serially uncorrelated disturbance term. Let a $r\times 1$ vector of instrumental variables be given by X_t, and suppose that the orthogonality condition $E[X_t u_t]=0$ is satisfied. The moment condition vector in this context is then equal to $g(z_t, b)=X_t(y_t - W_t'b)$, where $z_t=(y_t, W_t, X_t)$.

To calculate D, simply do the following. Get an initial consistent unconstrained estimate $\hat{b}_T^*$, say, by two stage least squares. Let $u_t^* \equiv y_t - X_t\hat{b}_T^*$ be the residual. Define $X_t^* \equiv u_t^* X_t$, $X^* \equiv (X_1^*,\ldots, X_T^*)'$, $X\equiv(X_1,\ldots, X_T)'$, $y_{t*}\equiv y_t/u_t^*$, $y_* \equiv (y_{1*},\ldots, y_{T*})'$, $y\equiv(y_1,\ldots, y_T)'$, $W_{t*}\equiv W_t/u_t^*$, $W_*\equiv(W_{1*},\ldots, W_{T*})'$,

$$W=(W_1,\ldots, W_T)'.$$

Next calculate $\hat{b}_T$, the two stage least squares estimator for a regression of y_* on W_* with instruments X^*. Because $V_T=\sum_{t=1}^T g(z_t, \hat{b}_T^*)g(z_t, \hat{b}_T^*)'/T=\sum_{t=1}^T (u_t^*)^2\cdot X_t X_t'/T=(X^*)'X^*/T$, this gives the efficient GMM estimator, which in this context is equal to White's [1982a] two stage instrumental variables estimator. Define $\hat{u}_{t*}\equiv y_{t*} - W_{t*}'\hat{b}_T$, $\hat{u}_*\equiv(\hat{u}_{1*},\ldots, \hat{u}_{T*})'$. Calculate $\hat{S}$, the sum of squares of the predicted values (i.e., the numerator of the constant unadjusted r-squared) of a regression of $\hat{u}_{t*}$ on X_t^*:

$$\hat{S} \equiv \hat{u}_*'X^*[(X^*)'X^*]^{-1}(X^*)'\hat{u}_* = (y - W\hat{b}_T)'XV_T^{-1}X'(y-W\hat{b}_T)/T = T\cdot J_T(\hat{b}_T).$$

Now substitute out the constraint in 3.5, and repeat the previous paragraph for the restricted estimator. Let $\tilde{S}$ be the result of the last step. Then D is calculated simply as $D=\tilde{S}-\hat{S}$.

The same method of calculating D is also applicable when a nonlinear constraint is being tested except that in general it will not be possible to substitute out for the constraint and use the above two stage least squares calculation to obtain the residuals for the restricted estimates. The method may be simplified somewhat if there is no conditional heteroskedasticity or if the estimation technique is OLS. See Startz [1982] and Domowitz [1983].

Princeton University, U.S.A.

REFERENCES

AMEMIYA, T., "The Nonlinear Two-Stage Least-Squares Estimator," *Journal of Econometrics*, 2 (1974), 105–110.

BERNDT, E. K. AND N. E. SAVIN, "Conflict among the Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," *Econometrica*, 45 (1977), 1263–1278.

BURGUETE, J. F., A. R. GALLANT AND G. SOUZA, "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews*, 1 (1982), 151–190.

CHAMBERLAIN, G., "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18 (1982), 5–46.

CUMBY, ROBERT E., JOHN HUIZINGA AND MAURICE OBSTFELD, "Two Step, Two Stage Least

Squares Estimation in Models with Rational Expectations," *Journal of Econometrics*, 21 (1983), 333–335.

DOMOWITZ, IAN, "Comment," *Econometric Reviews*, 2 (1983), 219–222.

EICHENBAUM, MARTIN S., LARS PETER HANSEN AND KENNETH J. SINGLETON, "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," Carnegie Mellon University, manuscript (1984).

ENGLE, ROBERT, "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics," in, Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics, vol. II*, (Amsterdam: North Holland, 1984), 775–826.

EVANS, G. B. A. AND N. E. SAVIN, "Conflict Among the Criteria Revisited: The W, LR and LM Tests," *Econometrica*, 50 (1982), 737–748.

FERGUSON, T. S., "A Method of Generating Best Asymptotically Normal Estimates with An Application to The Estimation of Bacterial Densities," *Annals of Mathematical Statistics*, 29 (1958), 1046–1062.

GALLANT, A. R. AND D. W. JORGENSON, "Statistical Inference for a System of Simultaneous, Nonlinear, Implicit Equations in the Context of Instrumental Variables Estimation," *Journal of Econometrics*, 11 (1979), 275–302.

HANSEN, LARS PETER, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 (1982), 1029–1054.

——— AND THOMAS SARGENT, "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models," *Journal of Monetary Economics*, 9 (1982), 263–296.

——— AND KENNETH SINGLETON, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50 (1982), 1269–1286.

HAUSMAN, JERRY AND DANIEL MCFADDEN, "Specification Tests for the Multinomial Logit Model," *Econometrica*, 52 (1984), 1219–1240.

HOLTZ-EAKIN, DOUGLAS, WHITNEY K. NEWEY AND HARVEY S. ROSEN, "Estimating Vector Autoregressions in Panel Data," manuscript (1987).

NEWEY, WHITNEY K., "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29 (1985), 299–256.

——— AND KENNETH D. WEST, "Hypothesis Testing with Efficient Method of Moments Estimation," Princeton University Woodrow Wilson School Discussion Paper, No. 103 (1985).

———, "A Simple, Positive Definite, Heteroskedasticity and Auto Correlation Consistent Covariance Matrix," forthcoming, *Econometrica* (1987).

PITMAN, E. T. G., "Notes on Nonparametric Statistical Inference," manuscript (1949).

STARTZ, RICHARD A., "Computation of Linear Hypothesis Tests for Two Stage Least Squares," *Economics Letters*, 11 (1982), 129–131.

WEST, KENNETH D., "A Specification Test for Speculative Bubbles," forthcoming, *Quarterly Journal of Economics* (1987a).

———, "A Standard Monetary Model and the Variability of The Deutschemark-Dollar Exchange Rate," forthcoming, *Journal of International Economics* (1987b).

———, "Dividend Innovations and Stock Price Volatility," forthcoming, *Econometrica* (1987c).

WHITE, HALBERT, "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50 (1982a), 483–501.

———, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50 (1982b), 1–26.

———, *Asymptotic Theory for Econometricions* (New York: Academic Press, 1984).

——— AND IAN DOMOWITZ, "Nonlinear Regression with Dependent Observations," *Econometrica*, 52 (1984), 143–162.