ROBUST GMM TESTS FOR MODEL SPECIFICATION, WITH APPLICATIONS TO CONDITIONAL MOMENTS TESTING AND STRUCTURAL INSTABILITY TESTING

Seung Chan Ahn

Department of Economics
Arizona State University
Tempe, AZ 85287-3806

November 1994

Revised, February 1995

Revised, September 1995

I gratefully acknowledge the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business. I also would like to thank seminar participants at Texas A&M University, Rice University and Econometric Society 7th World Congress, for their helpful comments. The title of earlier versions of this paper was "Model Specification Testing Based on Root-T Consistent Estimators." All errors contained in this paper are my own.

ROBUST GMM TESTS FOR MODEL SPECIFICATION, WITH APPLICATION TO CONDITIONAL MOMENTS TESTING AND STRUCTURAL INSTABILITY TESTING

Seung Chan AHN
Arizona State University

ABSTRACT

Tests for model specification based on the generalized method of moments have been widely used in the literature. Most of the popular tests typically require estimators which are efficient under the hypotheses that models chosen for estimation are correctly specified. This paper develops alternative tests which can be obtained using any consistent estimator. In particular, the test statistics are designed to have the same asymptotic power properties as statistics computed with efficient estimators. Several applications of the alternative testing methods are also considered.

Key Words: Specification tests, generalized method of moments, consistent estimators, conditional moments, structural stability.

JEL Classification Number: C30, C52

Address for correspondence: Professor Seung C. Ahn, Department of Economics, Arizona State University, Tempe, AZ 85287-3806, USA; Tel) 602-965-6574; Fax) 602-965-0748.

1. INTRODUCTION

This paper is concerned with model specification testing based on the generalized method of moments (GMM). Most econometric models imply moment conditions, that is, zero-expectation restrictions on certain functions of data and a vector of unknown parameters. Researchers can estimate the parameter vector by exploiting information about it contained in the sample analogue of the moment conditions. A GMM estimator is obtained by minimizing a quadratic function of sample moments. The class of GMM estimators is wide enough to include conventional estimators, such as maximum likelihood (ML), generalized least squares and instrumental-variables estimators. The consistency of an estimator crucially depends on the legitimacy of the moment conditions involved, so it is important to test them. In the GMM framework, violation of moment conditions can be detected by examining how far sample moments evaluated at an estimator diverge from zero.

The purpose of this paper is to develop statistics whose asymptotic distributions are asymptotically independent of those of the estimators used to compute them. In general, the distribution of a GMM statistic depends on what estimator is used. The distribution of the estimator should be considered in order to formulate a statistic which is asymptotically chi-squared under the null hypothesis that the moment conditions to be tested are legitimate. Newey (1985a, 1985b) provides a general approach to be used to derive relevant statistics when the distributions of the estimators are known. Computation of statistics following his approach is however complicated if the estimator is inefficient under the null hypothesis. This paper provides a unified approach by which tests robust to the distribution of the estimators used can be constructed. All that is required for the tests is $\sqrt{T}$-consistency of the estimators. The distribution of the estimators need not be known. Wooldridge (1990, 1991) has previously considered such tests in the context of conditional-moment (CM) testing. The link between his method and that developed in this paper is also considered.

Popular GMM tests, such as the overidentifying-restriction tests of Hansen (1982), CM tests of Newey (1985b) and Tauchen (1985), and the subsets-of-moment-conditions tests of Eichenbaum, Hansen and Singleton (1988), typically require estimators which are efficient under the null hypothesis. These statistics are easy to compute once the efficient estimators become available. The approach developed in this paper provides convenient alternative tests. That is, researchers can test for model specification using any initial consistent estimators. There is no need to compute efficient estimators before models are tested. Further, the alternative tests preserve the same asymptotic power properties as those based on efficient estimators. The tests would be particularly useful for cases in which efficient estimators are hard to compute, e.g., nonlinear simultaneous-equations models with different instruments for different equations [see Amemiya (1977)], or the asset pricing model recently considered by Heaton (1995).

This paper is organized as follows. Sections 2 and 3 set out the basic approach. In section 2, we consider a class of tests which are asymptotically identical to the Hansen (1982) test under local alternatives. An important characteristic of the tests in this class is that the statistics share the same functional form even if they are computed with different $\sqrt{T}$-consistent estimators. Section 3 studies a similar type of tests in cases in which only a subset of moment conditions is to be tested. For such cases, Eichenbaum, Hansen and Singleton (1988) -- hereafter denoted by EHS -- proposed a statistic which is an analogue of the likelihood ratio (LR) for ML models. Section 3 derives a class of tests which are asymptotically identical to the EHS test. It is also shown that the Wald-type tests of Newey (1985a) for testing subsets of moment conditions belong to the same class.

Sections 4 and 5 discuss several applications. Section 4 develops the regression-based CM tests which can apply to a broader setting than those of Newey (1985b), Tauchen (1985) and Wooldridge (1990, 1991). The link between their tests and the new tests is also discussed in their setting. Section 5 considers structural stability testing developed by Andrews and Fair (1988), Hoffman and Pagan (1989), Ghysels and Hall (1990a, 1990b). Alternative tests based on $\sqrt{T}$-consistent estimators are discussed. Some concluding remarks follow in section 6.

2. TESTING MOMENT CONDITIONS

Suppose that an econometric model implies the null hypothesis,

(2.1) $$H_o: E[g(z_t,\theta_o)] = 0 ,$$

here z_t is a vector of random variables, θ_o is a $p \times 1$ vector of unknown parameters, and $g(z_t,\theta)$ is a $q \times 1$ $(q \geq p)$ vector of functions of z_t and θ. We assume that θ_o is an interior point of a compact set, $\Theta \subset R^p$, and that $g(z_t,\theta)$ is differentiable and bounded almost everywhere on Θ. The stochastic process, $\{z_t: t = ... , -1, 0, 1, ...\}$, is stationary. The function $g(z_t,\theta)$ may include the score vector of a log-likelihood function. In typical cases, $g(z_t,\theta)$ denotes the product of instrumental variables and the error term in a model, and H_o implies the orthogonality between them. Such cases are discussed extensively in section 4.

We define the sample average of $g(z_t,\theta)$ by:

(2.2)
$$g_T(\theta) \equiv g_T(Z_T, \theta) = \frac{1}{T}\sum_{t=1}^{T} g(z_t, \theta) \ .$$

where $Z_T = \{z_1, \ldots, z_T\}$ denotes a data set. When H_o is correct, a consistent GMM estimator of θ_o can be obtained by minimizing $Tg_T(\theta)'\Xi g_T(\theta)$ with respect to θ, where Ξ is a $q \times q$ positive semi-definite nonstochastic weighting matrix. Let V denote the asymptotic covariance matrix of $\sqrt{T}g_T(\theta_o)$; and let $\tilde{\theta}$ be the GMM estimator obtained by using V^{-1} as a weighting matrix. Hansen (1982) shows that $\tilde{\theta}$ is the optimal estimator which has the smallest asymptotic covariance matrix among GMM estimators based on H_o. In practice, a consistent estimator of V, say V_T, should be used to compute $\tilde{\theta}$. For the consistent estimators of V, see Newey and West (1987b) or Andrews and Monahan (1992). We hereafter do not distinguish between V and V_T for notational convenience. It is also important to note that throughout this paper, all the tests are assumed to use the same estimator V_T. When different estimators are used for individual tests, the numerical equivalence results for some tests discussed below may not hold, even though their asymptotic equivalence remains unaffected.

H_o should be tested in order to check the consistency of GMM estimators. Before discussing the tests, we consider the alternatives to H_o under which the distributions of the relevant statistics can be analyzed. We define a sequence of local alternative hypotheses by $H_\ell = \{H_T\}_{T=1}^{\infty}$ where,

(2.3)
$$H_T\colon \sqrt{T}E[g_T(\theta_o)] = \omega + o(1) \ .$$

Newey (1985a) provides sufficient conditions under which (2.3) holds.[1]

Define $G_T(\theta) = \partial g_T(\theta)/\partial\theta'$, and let $G_o = \text{plim } G_T(\theta_o)$. We also define a matrix operator:

$$Q(\Xi,\Gamma) = \Xi^{-1} - \Xi^{-1}\Gamma(\Gamma'\Xi^{-1}\Gamma)^{-1}\Gamma'\Xi^{-1} , \tag{2.4}$$

where Ξ and Γ are conformable matrices and Ξ is invertible. With these notations and definitions, it can be shown that under plausible regularity conditions,

$$\sqrt{T}g_T(\theta_o) \rightarrow N(\omega,V) , \tag{2.5}$$

$$plim\ G_T(\hat{\theta}) = G_o , \tag{2.6}$$

$$\sqrt{T}\tilde{g}_T = VQ(V,G_o)\sqrt{T}g_T(\theta_o) + o_p(1) , \tag{2.7}$$

where "$\rightarrow$" means "converges in distribution," $\hat{\theta}$ is any $\sqrt{T}$-consistent estimator of θ_o,[2] and $\tilde{g}_T = g_T(\tilde{\theta})$. The proof and the detailed conditions required for (2.5)-(2.7) can be found in

---

[1] Suppose that the data Z_T has a well-defined density function $p(Z_T,\eta_T)$, where η_T is a vector of parameters depending on the sample size T. Let $E(\bullet|\eta)$ be the expectation operator when $\eta_T = \eta \neq \eta_o$; and let $E(\bullet)$ be the expectation operator when $\eta_T = \eta_o$. We assume that H_o holds if $\eta_T = \eta_o$; that is, $E[g_T(\theta_o)] \equiv \int g_T(Z_T,\theta_o)p(Z_T,\eta_o)dZ_T = 0$. If we specify $\eta_T = \eta_o + \rho/\sqrt{T}$, Taylor's approximation of $E[g_T(\theta_o)|\eta_T]$ around η_o yields (2.3) with $\omega = K\rho$, where $K = E[g_T(\theta_o)\partial \ln\{p(Z_T,\eta_o)\}/\partial\eta']$.

[2] That is, $\sqrt{T}(\hat{\theta} - \theta_o) = O_p(1)$. Note that any GMM estimator based on all or some of the moment conditions in (2.1) is $\sqrt{T}$-consistent under H_ℓ.

Newey (1985a).[3]

Once $\tilde{\theta}$ is computed, H_o can be easily tested by the Hansen (1982) statistic, $J_T = T\tilde{g}_T V^{-1}\tilde{g}_T$. The conditions (2.5) and (2.7) imply that under H_ℓ, J_T has a noncentral chi-square distribution with (q - p) degrees of freedom and noncentrality parameter $\lambda = \omega' Q(V,G_o)\omega$. There are numerous GMM tests which are asymptotically identical to J_T. In particular, Newey (1985a) shows that all the GMM tests with (q - p) degrees of freedom are equivalent to the Hansen test whatever GMM estimators based on the moment conditions in (2.1) are used. However, each of the statistics with (q - p) degrees of freedom should be constructed with knowledge of the distribution of the estimator used. We below derive a class of statistics alternative to J_T which are robust to the distributions of estimators.

PROPOSITION 1. Define a modified version of the Hansen statistic by:

$$MJ_T(\hat{\theta}) = T\hat{g}_T' Q(V,\hat{G}_T)\hat{g}_T = T\hat{g}_T' V^{-1}\hat{g}_T - \hat{g}_T' V^{-1}\hat{G}_T[\hat{G}_T' V^{-1}\hat{G}_T]^{-1}\hat{G}_T' V^{-1}\hat{g}_T \ , \tag{2.8}$$

where $\hat{g}_T = g_T(\hat{\theta})$ and $\hat{G}_T = G_T(\hat{\theta})$. Then, $MJ_T(\hat{\theta}) = J_T + o_p(1)$, for any $\sqrt{T}$-consistent estimator, $\hat{\theta}$.

PROOF. Taylor's expansion of $\hat{g}_T$ around θ_o yields:

$$\sqrt{T}\hat{g}_T = \sqrt{T}g_T(\theta_o) + G_T(\theta_L)\sqrt{T}(\hat{\theta} - \theta_o) \ , \tag{2.9}$$

where θ_L is a vector between $\hat{\theta}$ and θ_o. Since plim θ_L = plim $\hat{\theta} = \theta_o$, (2.6) should imply that

[3] For the proof of (2.5), Newey assumes that $\{z_t: t = ... , -1, 0, 1, ...\}$ is strictly stationary. However, we may relax this assumption, following Gallant and White (1988).

plim $G_T(\theta_L) = G_o$. Then, premultiplying both sides of (2.9) by $VQ(V,\hat{G}_T)$, and using the fact that $Q(V,\hat{G}_T)G_T(\theta_L) = Q(V,G_o)G_o + o_p(1) = o_p(1)$, we obtain:

(2.10)
$$VQ(V,\hat{G}_T)\sqrt{T}\hat{g}_T = VQ(V,G_o)\sqrt{T}g_T(\theta_o) + o_p(1) .$$

Hence, (2.7), (2.10) and the fact that $Q(V,\hat{G}_T)VQ(V,\hat{G}_T) = Q(V,\hat{G}_T)$ imply the result. ■

There exist numerous different MJ_T statistics depending on the choice of $\hat{\theta}$. While each of the statistics may have different finite-sample properties, any of them should be at least asymptotically as powerful as J_T. It is also important to note that $\hat{\theta}$ need not be a GMM estimator based on the moment conditions in (2.1). All that is required is its $\sqrt{T}$- consistency, and furthermore, its asymptotic distribution does not affect that of the corresponding MJ_T. This result might be used when a researcher wants to compare his/her estimates to those of other studies for the same model. A statistic of the form of MJ_T can be constructed with the estimates from other studies.

When a subset of parameters in θ appears in only a subset of moment functions, the J_T and MJ_T statistics have some interesting properties. For example, assume that $g(z_t,\theta_o)$ takes the form:

(2.11)
$$g(z_t,\theta_o) = \begin{bmatrix} b(z_t,\theta_o) \\ c(z_t,\theta_o) \end{bmatrix} = \begin{bmatrix} b(z_t,\theta_{o,1}) \\ c(z_t,\theta_{o,1},\theta_{o.2}) \end{bmatrix} ,$$

where $\theta_o = [\theta_{o,1}',\theta_{o,2}']'$. Here $b(z_t,\theta)$ and $c(z_t,\theta)$ are $q_b \times 1$ and $q_c \times 1$ vectors, respectively, and $q_b + q_c = q$. The $p_1 \times 1$ parameter vector θ_1 includes a subset of parameters in θ, while the $p_2 \times 1$ ($p_2 = p - p_1$) vector θ_2 includes the parameters which appear only in $c(z_t,\theta)$. Note that $\theta_1 = \theta$ if $p_2 = 0$. We assume that $q_b \geq p_1$, so that θ_1 is identified with the moment conditions on $b(z_t,\theta_1)$ only. Some examples under which $g(z_t,\theta)$ may take the form of (2.11) are found in EHS, Ahn and Schmidt (1995) and Heaton (1995).

Define $b_T(\theta_1)$ and $c_T(\theta)$ similarly to $g_T(\theta)$, and partition $G_T(\theta)$ such that:

$$G_T(\theta) = \frac{\partial[b_T(\theta_1)', c_T(\theta)']'}{\partial(\theta_1', \theta_2')} = \begin{bmatrix} B_{T,1}(\theta_1) & B_{T,2}(\theta_1) \\ C_{T,1}(\theta) & C_{T,2}(\theta) \end{bmatrix}, \tag{2.12}$$

where $B_{T,2}(\theta_1) = 0$. Corresponding to $b_T(\theta_1)$ and $c_T(\theta)$, we also divide V into $[V_{ij}]$, where i, j = b, c. Then, we obtain the following result.

PROPOSITION 2. Let $\tilde{\theta} = [\tilde{\theta}_1', \tilde{\theta}_2']'$ and $\tilde{g}_T = [\tilde{b}_T', \tilde{c}_T']'$. Let $\check{\theta}_1$ be the GMM estimator which minimizes $Tb_T(\theta_1)'(V_{bb})^{-1}b_T(\theta_1)$. Let $\hat{\theta} = [\hat{\theta}_1', \hat{\theta}_2']'$. If $p_2 = q_c$, the following numerical equalities hold:

$$\tilde{\theta}_1 = \check{\theta}_1 \; ; \; J_T = T\check{b}_T'(V_{bb})^{-1}\check{b}_T \; ; \; MJ_T(\hat{\theta}) = \hat{b}_T' Q(V_{bb}, \hat{B}_{T,1})\hat{b}_T , \tag{2.13}$$

where $\check{b}_T = b_T(\check{\theta}_1)$, $\hat{b}_T = b_T(\hat{\theta}_1)$ and $\hat{B}_{T,1} = B_{T,1}(\hat{\theta}_1)$.

The proof is omitted because the first and second equalities are the "separability" result of Ahn and Schmidt (1995), and the last equality can be also shown by tedious but straightforward algebra. The novel finding here is that the "separability" result also applies to MJ_T statistics. If $p_2 = q_c$, both J_T and MJ_T depend on neither the moment function $c_T(\theta)$ nor the estimators of $\theta_{o,2}$. This implies that $c_T(\theta)$ can be used to identify $\theta_{o,2}$, but it does not contain any useful information about the possible violation of moment conditions on $c(z_t,\theta_o)$.

An important example for which Proposition 2 is particularly relevant is the nonlinear simultaneous-equations models. In linear models, the three-stage least squares estimates remain unaffected whether the exactly identified equations are removed or not. Proposition 2 implies that the same result applies to nonlinear models. Furthermore, J_T or MJ_T computed for the entire system of a model (numerically) equal J_T or MJ_T for the sub-system with overidentified equations only. Therefore, some caution is required when J_T or MJ_T statistics are used for the specification of simultaneous-equations models. They have no power to detect any possible misspecification of exactly identified equations.

3. TESTING SUBSETS OF MOMENT CONDITIONS

In some cases, researchers may have prior information about the directions of misspecification which restrict violation of H_o to a certain subset of moment conditions. In such cases, statistics designed to focus their power on the moment conditions to be tested may have better power properties than the Hansen statistic for testing the entire set of conditions under H_o. Such statistics have been considered by EHS and Newey (1985a). In this section, we derive a class of tests which share the same asymptotic local power with theirs.

We consider cases in which the moment function $g(z_t,\theta)$ is of the form (2.11). Suppose that a researcher wishes to test H_o against,

$$\text{(3.1)} \qquad H_A^*:\ E[b(z_t,\theta_{o,1})] = 0 \text{ and } E[c(z_t,\theta_{o,1},\theta_{o,2})] \neq 0 \ .$$

By the nature of the model, the parameter vector $\theta_{o,2}$ is identified only under H_o while $\theta_{o,1}$ is under both H_o and H_A^*. Note that $\check{\theta}_1$ is the optimal GMM estimator of $\theta_{o,1}$ under H_A^*. When $p_2 = 0$, $\check{\theta}_1$ denotes the optimal GMM estimator of θ_o under H_A^*.

The EHS statistic is given by:

$$D_T = T[\tilde{g}_T'V^{-1}\tilde{g}_T - \check{b}_T'(V_{bb})^{-1}\check{b}_T] \;, \tag{3.2}$$

which is asymptotically chi-squared with (q_c - p_2) degrees of freedom under H_o. Newey and West (1987a) [also Gallant and Jorgenson (1979)] have considered a similar type of GMM statistics in the context of parametric-restriction testing. In fact, we may also regard D_T as a statistic for testing the parametric restrictions imposed by the model underlying H_o. To see why, consider cases in which $p_2 = 0$. Define an auxiliary parameter vector $\delta_o = E[c(z_t,\theta_o)]$; and let $\gamma_o = [\theta_o',\delta_o']'$. Let $h_T(z_t,\gamma) = [b_T(\theta)',(c_T(\theta) - \delta)']'$. Note that under both H_o and H_A^*, $E[h_T(z_t,\gamma_o)] = 0$. Define the restricted GMM estimator by $\tilde{\gamma}$, which solves the problem: $\min_\gamma h_T(\gamma)'V^{-1}h_T(\gamma)$ subject to $\delta = 0$. Corresponding to $\tilde{\gamma}$, we define the unrestricted estimator of γ_o by $\dot{\gamma}$. Then, the LR-type statistic of Newey and West (1987a) for testing the hypothesis $\delta_o = 0$ is given by:

$$T[h_T(\tilde{\gamma})'V^{-1}h_T(\tilde{\gamma}) - h_T(\dot{\gamma})'V^{-1}h_T(\dot{\gamma})] \;. \tag{3.3}$$

Obviously, $h_T(\tilde{\gamma})V^{-1}h_T(\tilde{\gamma}) = \tilde{g}_TV^{-1}\tilde{g}_T$ because $\tilde{\gamma} = [\tilde{\theta}',0]'$. Furthermore, we can show that $h_T(\dot{\gamma})'V^{-1}h_T(\dot{\gamma}) = \check{b}_T(V_{bb})^{-1}\check{b}_T$, since $h_T(\gamma)$ satisfies the conditions for Proposition 2 to hold. Substituting these results into (3.3) yields D_T.

The EHS test may have poor power properties against certain alternatives if $p_2 \neq 0$. For this case, let $c_T(\theta) = [c_{1,T}(\theta)',c_{2,T}(\theta)']'$ where $\dim[c_{1,T}(\theta)] = p_2$. Define $b_m(\theta) = [b_T(\theta_1)',c_{1,T}(\theta_1,\theta_2)']'$ and $c_m(\theta) = c_{2,T}(\theta)$, suppressing subscript "T" for notational convenience. The partition of $c_T(\theta)$ is arbitrary, except that it is chosen such that $\mathrm{Rank}[\partial b_m(\theta)/\partial\theta'] = p$ uniformly for $\theta \in \Theta$. According to $b_m(\theta)$ and $c_m(\theta)$, we partition V as $[V_{ij}^m]$, where i, j =

b, c. Let $\check{\theta}_m$ be the GMM estimator which is the solution of the problem: $\min_\theta T b_m(\theta)'(V_{bb}^m)^{-1} b_m(\theta)$. Since $b_m(\theta)$ and $\check{\theta}_m$ satisfy all the conditions of Proposition 2, we have $T\check{b}_T'(V_{bb})^{-1}\check{b}_T = T\check{b}_m'(V_{bb}^m)^{-1}\check{b}_m$, where $\check{b}_m = b_m(\check{\theta}_m)$. Therefore, we obtain the equality:

(3.4)
$$D_T = T[\tilde{g}_T' V^{-1}\tilde{g}_T - \check{b}_m'(V_{bb}^m)^{-1}\check{b}_m] \ .$$

Note that D_T takes the form of D_T for testing the null hypothesis $E[c_{2,T}(\theta_o)] = 0$ against the alternative hypothesis $E[c_{2,T}(\theta_o)] \neq 0$ while the condition $E[c_{1,T}(\theta_o)] = 0$ is falsely assumed to be legitimate under both H_o and H_A^*. This implies that the test may not have power against $E[c_{1,T}(\theta_o)] \neq 0$ when $p_2 \neq 0$.

A practical disadvantage of D_T is that it requires computation of both $\tilde{\theta}$ and $\check{\theta}_1$. However, similarly to the MJ_T statistics, we can derive a class of statistics, each of which is asymptotically equivalent to D_T and requires one estimator only.

PROPOSITION 3. Let $\hat{\theta} = [\hat{\theta}_1',\hat{\theta}_2']'$ be any $\sqrt{T}$-consistent estimator of θ_o. Define a modified version of D_T by:

(3.5)
$$MD_T(\hat{\theta}) = T[\hat{g}_T' Q(V,\hat{G}_T)\hat{g}_T - \hat{b}_T' Q(V_{bb},\hat{B}_{T,1})\hat{b}_T] \ ,$$

where $\hat{b}_T = b_T(\hat{\theta}_1)$ and $\hat{B}_{T,1} = B_{T,1}(\hat{\theta}_1)$. Then, $MD_T(\hat{\theta}) = D_T + o_p(1)$ under H_ℓ. Further, $MD_T(\hat{\theta}) \geq 0$.

PROOF. See the Appendix. □

Proposition 3 indicates that for any $\sqrt{T}$-consistent $\hat{\theta}$, $MD_T(\hat{\theta})$ has the same local power as D_T. This result is not affected even if the two terms in MD_T are computed using different $\sqrt{T}$-consistent estimators. However, it is desirable to use the same estimator, because otherwise the modified statistics may have negative values.

Incorporating H_A^*, we may specify H_ℓ as $H_\ell^* = \{H_T^*\}_{T=1}^{\infty}$ where,

$$H_T^*: \sqrt{T}E[b_T(\theta_{o,1})] = 0 \textit{ and } \sqrt{T}E[c_T(\theta_o)] = \omega_c + o(1) \; . \tag{3.6}$$

Newey (1985a) derives the Wald-type GMM statistics which are designed to have maximum power against H_ℓ^* when $p_2 = 0$. We here introduce his tests allowing $p_2 > 0$. Let $B_T(\theta) = [B_{T,1}(\theta_1),0]$ and $C_T(\theta) = [C_{T,1}(\theta_1),C_{T,2}(\theta)]$. Define:

$$r_T(\theta) = c_T(\theta) - V_{cb}(V_{bb})^{-1}b_T(\theta_1) \; ,$$

$$R_T(\theta) = C_T(\theta) - V_{cb}(V_{bb})^{-1}B_T(\theta) \; ,$$

$$\Omega = V_{cc} - V_{cb}(V_{bb})^{-1}V_{bc} \; ,$$

$$\Psi_T(\theta) = \Omega - R_T(\theta)[B_T(\theta)'(V_{bb})^{-1}B_T(\theta) + R_T(\theta)'\Omega^{-1}R_T(\theta)]^{-1}R_T(\theta)' \; .$$

Then, one of Newey's statistics is obtained by:

$$N_T = T\tilde{r}_T'\Omega^{-1}(\Omega^{-1}\tilde{\Psi}_T\Omega^{-1})^{+}\Omega^{-1}\tilde{r}_T \; , \tag{3.7}$$

where $\tilde{r}_T = r_T(\tilde{\theta})$, $\tilde{\Psi}_T = \Psi_T(\tilde{\theta})$ and $(\bullet)^+$ is the Moore-Penrose g-inverse. Following the proof of Proposition 3 of Newey (1985a), we can show that under H_ℓ^*, N_T has a noncentral chi-square distribution with $(q_c - p_2)$ degrees of freedom and noncentrality parameter $\lambda = \omega_c'\Omega^{-1}\Psi_o\Omega^{-1}\omega_c$, where $\Psi_o = \text{plim } \Psi_T(\theta_o)$.

When $p_2 = 0$, the Moore-Penrose g-inverse in N_T can be replaced by the usual inverse.

Furthermore, for the same case, Newey (1985a) derives an alternative statistic which is constructed with $\check{\theta}_1$ but asymptotically identical to N_T:

$$AN_T = T\check{r}_T'\{\Omega + \check{R}_T[\check{B}_T'(V_{bb})^{-1}\check{B}_T]^{-1}\check{R}_T'\}^{-1}\check{r}_T = T\check{r}_T'\Omega^{-1}\check{\Psi}_T\Omega^{-1}\check{r}_T \,, \tag{3.8}$$

where $\check{r}_T = r_T(\check{\theta}_1)$, $\check{R}_T = R_T(\check{\theta}_1)$, $\check{B}_T = B_T(\check{\theta}_1)$ and $\check{\Psi}_T = \Psi_T(\check{\theta}_1)$. Ahn and Schmidt (1995) show that AN_T equals the Wald statistic for testing the hypothesis $\delta_o = E[c(z_t,\theta_o)] = 0$ which is constructed with the unrestricted GMM estimator $\dot{\gamma}$ defined above. Therefore, in cases where $p_2 = 0$, the asymptotic equivalence between D_T and AN_T (and N_T) follows from Theorem 2 of Newey and West (1987a).

When $p_2 \neq 0$, it can be shown that similarly to D_T, N_T takes the form of N_T for testing $E[c_{2,T}(\theta_o)] = 0$. [See Lemma 1 of Appendix.] This remarkable similarity between D_T and N_T suggests that both statistics should be related for any p_2. In fact, we obtain the following result.

PROPOSITION 4. (i) $MD_T(\tilde{\theta}) = N_T$. (ii) When $p_2 = 0$, $MD_T(\check{\theta}_1) = AN_T$.

PROOF. See the Appendix. □

Proposition 4 implies that both N_T and AN_T belong to the class of MD_T statistics. Therefore, Propositions 3 and 4 formally establish the asymptotic equivalence between D_T and N_T (and AN_T when $p_2 = 0$). Whenever N_T and AN_T have desirable power properties (e.g., under H_ℓ^* with $p_2 = 0$), so do D_T and any MD_T statistic.

4. CONDITIONAL MOMENT TESTS

Econometric models with exogenous variables usually imply conditional moment (CM) restrictions. The specification of the models can be tested by checking the legitimacy of such restrictions. Regression-based CM tests have been popular in the literature. For example, Newey (1985b) and Tauchen (1985) examine CM tests of ML models with independently distributed data. Wooldridge (1990, 1991) considers CM tests of conditional mean or variance specifications.[4] In particular, his CM statistics are computed with any $\sqrt{T}$-consistent estimator, similarly to the tests considered in previous sections. In this section, extending the approach developed in previous sections, we derive alternative regression-based CM tests, which may apply to a broader setting than those of previous studies.

Suppose that an economic model implies the CM hypothesis:

$$\textbf{(4.1)} \qquad H_o^c:\ E[u_j(z_t,\theta_o)\,|\,x_{jt}] \equiv E[u_{jt}(\theta_o)\,|\,x_{jt}] = 0\ ;\ j = 1, 2, \dots , k\ ,$$

where u_{jt} is a scalar function of z_t and θ_o, and x_{jt} is a vector of variables exogenous or predetermined with respect to u_{jt}. Each of the x_{jt} ($j = 1, 2, \dots , k$) includes a subset of variables in z_t. In multivariate time-series settings, some or all of the past values of z_t could be included in the x_{jt}. Define the vector of all the distinct variables in at least one of the x_{jt} by x_t^*; and let $u_t(\theta_o) = [u_{1t}(\theta_o), \dots , u_{kt}(\theta_o)]'$. As in the previous studies mentioned above, we also assume "No-Autocorrelation" among the $u_t(\theta_o)$:

$$\textbf{(NA)} \qquad E[u_t(\theta_o)u_{t+i}(\theta_o)'\,|\,x_t^*, x_{t+i}^*] = 0.$$

Assumption NA is appropriate for models with cross-sectional data. It may be also valid for time-series dynamic models which impose martingale-difference restrictions on both first and

---

[4] An excellent survey on CM tests can be found in White (1994, Ch. 9 - 10).

second order moments. [For such cases, see Wooldridge (1991).]

Each of the x_{jt} may be or may not be the same for j = 1, 2, ... , k. In order to distinguish these two possible cases, we can make two alternative assumptions about exogeneity, which we refer to as "Strong Exogeneity" (SE) and "Weak Exogeneity" (WE) assumptions, respectively. Stated formally:

(SE) For any j = 1, 2, ... , k, $x_{jt} = x_t^*$.

(WE) For some j, $x_{jt} \neq x_t^*$.

Assumption SE, which has been adopted in previous studies on CM tests, implies that if a variable is exogenous to one of the $u_{jt}(\theta_o)$, it is also exogenous to the others. There are many models which may imply H_o^c with SE. Examples include nonlinear regression models and quasi-ML models using densities in the linear exponential family. [See White (1980) and Gourieroux, Monfort and Trognon (1984).] For these models, the $u_{jt}(\theta_o)$ denote (generalized) residuals, each of which has zero expectation conditional on a common set of exogenous variables, x_t^*.

Assumption WE means that there exist some variables which are exogenous to some but not all of the $u_{jt}(\theta_o)$. One leading example of the models that imply H_o^c with WE is the nonlinear (or linear) simultaneous-equations model with a different set of instrumental variables for each equation. [See Amemiya (1977) and Schmidt (1990).] For this model, u_{jt} denotes the residual of the j'th equation, and x_{jt} the vector of predetermined variables in the same equation. Another example for WE is the panel data model with sequential moment restrictions, which is considered by Chamberlain (1992). For his model, the subscripts "t" and "j" denote cross-sectional unit and time, respectively, and $x_{jt} \subset x_{j+1,t}$. Allowing WE, the analysis adopted in this section is general enough to encompass those of previous studies.

If H_o^c is correct, functions of x_{jt} should be uncorrelated with $u_{jt}(\theta_o)$ for any j. Any evidence against these orthogonality conditions implies violation of H_o^c. Based on this observation, we can derive appropriate tests of H_o^c. For each j = 1, 2, ... , k, let $\lambda_{jt} = \lambda(x_{jt})$ be an arbitrary $1 \times q_j$ vector of functions of x_{jt}, where $\Sigma_{j=1}^{k} q_j = q > p$; and let $\Lambda_t = \Lambda(x_t^*) = \text{diag}(\lambda_{1t}, \ldots , \lambda_{kt})$. When assumption SE holds, Λ need not be a block-diagonal. Specification of the model underlying H_o^c can be checked by testing the orthogonality between Λ_t and $u_t(\theta_o)$:

$$E[\Lambda_t' u_t(\theta_o)] = 0 \; . \tag{4.2}$$

Here, the "criterion" matrix Λ_t may be chosen such that it also depends on θ_o or (possibly) a vector of nuisance parameters (say, π_o). Such a case may arise if λ_{jt} is a function of $E[\partial u_{jt}(\theta_o)/\partial\theta' | x_{jt}]$ and/or $E[u_{jt}(\theta_o)u_{jt}(\theta_o)' | x_{jt}]$, as we discuss below. In this case, a GMM statistic for testing (4.2) can be obtained with consistent estimates of the Λ_t [e.g., Λ_t evaluated at $\sqrt{T}$-consistent estimates of θ_o and π_o]. Of course, tests with the estimated Λ_t are asymptotically equivalent to those with the "true" Λ_t.

Hansen methods can be used to test the orthogonality condition of form (4.2). In particular, we can compute a Hansen statistic via an auxiliary regression. Using the same notation as we defined above, let $g_T(\theta) = T^{-1}\Sigma_{t=1}^{T}\Lambda_t' u_t(\theta)$ and $G_T(\theta) = T^{-1}\Sigma_{t=1}^{T}\Lambda_t' U_t(\theta)$, where $U_t(\theta) = \partial u_t(\theta)/\partial\theta'$. We assume that the Λ_t are chosen such that $\text{Rank}[G_T(\theta)] = p$ uniformly for $\theta \in \Theta$. This assumption ensures that none of the orthogonality conditions in (4.2) are redundant. Let $\tilde{\theta}$ be the optimal GMM estimator based on (4.2). Define:

Since we assume NA, $V_T(\hat{\theta})$ is a consistent estimator of the asymptotic covariance matrix of

$$V_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \Lambda_t' u_t(\theta) u_t(\theta)' \Lambda_t \ . \tag{4.3}$$

$\sqrt{T}\bar{g}_T(\theta_o)$, for any $\sqrt{T}$-consistent $\hat{\theta}$. When $V_T(\tilde{\theta})$ is chosen, it can be easily shown that a Hansen statistic is obtained by T times the uncentered R^2 (R_u^2) from the regression of one on $u_t(\tilde{\theta})'\Lambda_t$.

Alternatively, we can use modified Hansen statistics which are computed with any $\hat{\theta}$. Let $OMJ_T(\hat{\theta})$ denote $MJ_T(\hat{\theta})$ for testing the orthogonality conditions in (4.2) computed with $V_T(\hat{\theta})$. Interestingly, this statistic can be also obtained via regression-based procedures, which are quite similar to Wooldridge (1990). To show how, we partition Λ_t into $[\Lambda_{b,t},\Lambda_{c,t}]$ such that the number of columns of $\Lambda_{b,t}$ equals p, the number of parameters in θ. Define $\hat{u}_t = u_t(\hat{\theta})$ and $\hat{U}_t = U_t(\hat{\theta})$. Let $\Lambda_{c,t}^r$ be the t'th residual matrix [or row vector if $u_t(\theta)$ is scalar] from the two-stage least squares (2SLS) matrix regression of $\Lambda_{c,t}$ on $\Lambda_{b,t}$ with the instrument $\hat{U}_t$:

$$\Lambda_{c,t}^r = \Lambda_{c,t} - \Lambda_{b,t}\Big[\sum_{t=1}^{T} \hat{U}_t'\Lambda_{b,t}\Big]^{-1}\sum_{t=1}^{T} \hat{U}_t'\Lambda_{c,t} \ , \quad t = 1, \ldots , T \ . \tag{4.4}$$

Then, we obtain the following result.

PROPOSITION 5. For any $\hat{\theta}$, $OMJ_T(\hat{\theta})$ is computed by TR_u^2 from the regression of

$$1 = \hat{u}_t'\Lambda_{c,t}^r\psi + error \ . \tag{4.5}$$

PROOF. See the Appendix. □

Since OMJ_T statistics do not require $\tilde{\theta}$, they are easier to compute than the Hansen statistic in many cases. For example, consider the specification test of the nonlinear (or linear) simultaneous-equations model:

$$f_j(z_t,\alpha_j) = u_{jt} \; ; \; j = 1, 2, \ldots , k \; , \tag{4.6}$$

where α_j is a $p_j \times 1$ vector of parameters in the j'th equation, and u_{jt} is the residual of the j'th equation. Let λ_{jt} ($j = 1, 2, \ldots , k$) be the row vector of instruments for the j'th equation; and partition it into $[\lambda_{b,jt},\lambda_{c,jt}]$ such that $\lambda_{b,jt}$ includes only p_j variables. For each equation, the parameter vector (α_j) can be consistently estimated by the nonlinear two-stage least squares method of Amemiya (1974). For example, let $\hat{\alpha}_j$ be the nonlinear 2SLS estimator with instrument $\lambda_{b,jt}$. Let $\lambda^r_{c,jt}$ be the residual vectors from the 2SLS regression of $\lambda_{c,jt}$ on $\lambda_{b,jt}$ with instruments $\partial f_j(z_t,\hat{\alpha}_j)/\partial\alpha_j'$. Then, the specification of the model given in (4.6) can be tested by TR^2_u from the regression of one on $[f_1(z_t,\hat{\alpha}_1)\lambda^r_{c,1t}, \ldots ,f_k(z_t,\hat{\alpha}_k)\lambda^r_{c,kt}]$.[5]

We now examine the relationship of OMJ_T tests with the CM tests of previous studies, mainly under Wooldridge's (1990, 1991) framework. When assumption SE holds, we may define $\Delta_t(\theta) = E[U_t(\theta)|x^*_t]$. Let $W_t(\theta,\pi_o)$, $t = 1, 2, \ldots , T$, be a $k \times k$ positive definite matrix which may be a function of x^*_t, θ and a nuisance parameter vector π_o. For analytical convenience, we assume that π_o is known.[6] Let θ^* be the GMM estimator which satisfies:

$$\sum_{t=1}^{T} \Delta_t(\theta^*)'[W_t(\theta^*,\pi_o)]^{-1}u_t(\theta^*) = 0 \; . \tag{4.7}$$

When data are independently distributed, θ^* is efficient if H^c_o holds and $W^o_t \equiv W_t(\theta_o,\pi_o) =$

---

5 This simplification is possible because $\Lambda_{b,t}$, $\Lambda_{c,t}$, $\hat{U}_t$ and $\Lambda^r_{c,t}$ are all block-diagonal.

6 When π_o is unknown, we may replace it by a $\sqrt{T}$-consistent estimator, without changing the result obtained below.

$E[u_t(\theta_o)u_t(\theta_o)'|x_t^*]$. [See Chamberlain (1987).] For ML models, $\Sigma_{t=1}^{T}\Delta_t(\theta)[W_t(\theta,\pi)]^{-1}u_t(\theta)$ may represent the score vector while the $u_t(\theta)$ are generalized residuals. In this case, θ^* is the ML estimator of θ_o.

Let Φ_t be a k × s "indicator" matrix of functions of x_t^* (and possibly θ_o or π_o).[7] Define:

(4.8)
$$\zeta_T = \frac{1}{T}\sum_{t=1}^{T} \Phi_t'[W_t(\theta)]^{-1}u_t(\theta) \ ,$$

where $W_t(\theta) = W_t(\theta,\pi_o)$. Note that if H_o^c is correct, $\zeta_T(\hat{\theta})$ must be close to zero for any $\sqrt{T}$-consistent $\hat{\theta}$. Therefore, Wald-type tests of H_o^c can be constructed based on $\zeta_T(\hat{\theta})$. In cases where θ^* is the ML estimator, Newey (1985b) and Tauchen (1985) show that a Wald statistic based on $\zeta_T(\theta^*)$ is obtained by TR_u^2 from the regression of one on $u_t(\theta^*)'[W_t(\theta^*)]^{-1}\Delta_t(\theta^*)$ and $u_t(\theta^*)'[\hat{W}_t(\theta^*)]^{-1}\Phi_t$.

For the cases where θ^* is not the ML estimator, we may construct alternative Wald statistics, following Wooldridge's (1990) approach. Let $\Delta_t^o = \Delta_t(\theta_o)$, and define:

$$M_1 = \frac{1}{T}\sum_{t=1}^{T} \Delta_t^{o\prime}[W_t^o]^{-1}\Delta_t^o \ ; \quad M_2 = \frac{1}{T}\sum_{t=1}^{T} \Phi_t'[W_t^o]^{-1}\Delta_t^o \ .$$

Consider the statistic:

[7] When Φ_t depends on θ_o and π_o, we replace Φ_t by its consistent estimator.

$$\textbf{(4.9)} \qquad \xi_T(\theta,\hat{M}_1,\hat{M}_2) \;=\; \frac{1}{T}\sum_{t=1}^{T} \Phi_t'[W_t(\theta)]^{-1}u_t(\theta) \;-\; \hat{M}_2(\hat{M}_1)^{-1}\frac{1}{T}\sum_{t=1}^{T} \Delta_t(\theta)'[W_t(\theta)]^{-1}u_t(\theta) \;,$$

where $\hat{M}_1$ and $\hat{M}_2$ are consistent estimators of M_1 and M_2, respectively. Expanding $\zeta_T(\hat{\theta})$ and $\xi_T(\hat{\theta},\hat{M}_1,\hat{M}_2)$ around θ_o, we can show that:

$$\textbf{(4.10)} \qquad \sqrt{T}\zeta_T(\hat{\theta}) \;=\; \sqrt{T}\xi_T(\hat{\theta},\hat{M}_1,\hat{M}_2) \;+\; o_p(1) \;=\; \sqrt{T}\xi_T(\theta_o,M_1,M_2) \;+\; o_p(1) \;,$$

for any $\sqrt{T}$-consistent $\hat{\theta}$. Therefore, a Wald test based on $\zeta_T(\hat{\theta})$ is asymptotically equivalent to that based on $\xi_T(\hat{\theta},\hat{M}_1,\hat{M}_2)$. Define:

$$D(\theta) \;=\; \frac{1}{T}\sum_{t=1}^{T} [\Delta_t(\theta),\Phi_t]'[W_t(\theta)]^{-1}u_t(\theta)u_t(\theta)'[W_t(\theta)]^{-1}[\Delta_t(\theta),\Phi_t] \;,$$

and $D = \text{plim}\, D(\theta_o)$. Corresponding to Φ_t and Δ_t^o, we partition D into $[D_{ij}]$, where i, j = 1, 2. Then, it is straightforward to show that the asymptotic covariance matrix of $\sqrt{T}\xi_T(\theta_o)$ is given by:

$$\Pi \;=\; D_{22} \;-\; M_2(M_1)^{-1}D_{12} \;-\; D_{21}(M_1)^{-1}M_2' \;+\; M_2(M_1)^{-1}D_{11}(M_1)^{-1}M_2' \;.$$

Letting $\hat{D} = [\hat{D}_{ij}]$ be a consistent estimator of D, define a consistent estimator of Π by:

$$\Pi_T \;=\; \hat{D}_{22} \;-\; \hat{M}_2(\hat{M}_1)^{-1}\hat{D}_{12} \;-\; \hat{D}_{21}(\hat{M}_1)^{-1}\hat{M}_2 \;+\; \hat{M}_2(\hat{M}_1)^{-1}\hat{D}_{11}(\hat{M}_1)^{-1}\hat{M}_2 \;.$$

Then, we can define a Wald statistic by[8]:

$$\textbf{(4.11)} \qquad W_T(\hat{\theta},\hat{M}_1,\hat{M}_2,\hat{D}) \;=\; T\xi_T(\hat{\theta},\hat{M}_2,\hat{M}_2)'(\Pi_T)^{-1}\xi_T(\hat{\theta},\hat{M}_1,\hat{M}_2) \;.$$

Depending on the choices of $\hat{\theta}$, $\hat{M}_1$, $\hat{M}_2$ and $\hat{D}$, we can have many different Wald test statistics. An interesting point here is that all the Wald statistics of form (4.11) are

---

8 See Corollary 9.10 of White (1994).

asymptotically identical because of (4.10).

Suppose that $W_t^o = E[u_t(\theta_o)u_t(\theta_o)'|x_t^*]$. Then, $M_1 = D_{11}$ and $M_2 = D_{21}$, so that we have $\Pi = D_{22} - D_{21}(D_{11})^{-1}D_{12}$. Based on this observation, we define $\hat{D}^* = [(\hat{D}^*)_{ij}] = D(\theta^*)$, $\hat{M}_1^* = (\hat{D}^*)_{11}$ and $\hat{M}_2^* = (\hat{D}^*)_{21}$. Then, we can show that $W_T(\theta^*,\hat{M}_1^*, \hat{M}_2^*,\hat{D}^*)$ is obtained by the procedure proposed by Newey (1985a) and Tauchen (1985). This result implies that their method can be used even if θ^* is not the ML estimator. It requires only that $W_t^o = E[u_t(\theta_o)u_t(\theta_o)'|x_t^*]$.

When $W_t^o \neq E[u_t(\theta_o)u_t(\theta_o)'|x_t^*]$, the regression-based tests of Newey and Tauchen may have the wrong size under H_o^c. In order to obtain the robust statistic of Wooldridge, define $\hat{M}_{W,1} = T^{-1}\Sigma_{t=1}^T\hat{\Delta}_t'(\hat{W}_t)^{-1}\hat{\Delta}_t$, $\hat{M}_{W,2} = T^{-1}\Sigma_{t=1}^T\Phi_t'(\hat{W}_t)^{-1}\hat{\Delta}_t$ and $\hat{D} = D(\hat{\theta})$, where $\hat{\Delta}_t = \Delta_t(\hat{\theta})$, $\hat{W}_t = W_t(\hat{\theta})$. The Wooldridge statistic is given by $W_T(\hat{\theta},\hat{M}_{W,1},\hat{M}_{W,2},\hat{D})$, which can be obtained by an auxiliary regression. Let $\tilde{\Phi}_t^r$ (t = 1, 2, ... , T) be the residual matrix from the regression of $\tilde{\Phi}_t = (\hat{W}_t)^{-1/2}\Phi_t$ on $\tilde{\Delta}_t = (\hat{W}_t)^{-1/2}\hat{\Delta}_t$; that is,

(4.12) $$\tilde{\Phi}_t^r = \tilde{\Phi}_t - \tilde{\Delta}_t[\Sigma_t\tilde{\Delta}_t'\tilde{\Delta}_t]^{-1}\Sigma_t\tilde{\Delta}_t'\tilde{\Phi}_t ,$$

where "Σ_t" means "$\Sigma_{t=1}^T$." Then, $W_T(\hat{\theta},\hat{M}_{W,1},\hat{M}_{W,2},\hat{D})$ is computed by TR_u^2 from a regression of:

(4.13) $$1 = \tilde{u}_t(\hat{\theta})'\tilde{\Phi}_t^r\psi + error ,$$

where $\tilde{u}_t(\hat{\theta}) = (\hat{W}_t)^{-1/2}u_t(\hat{\theta})$. This statistic is asymptotically equivalent to that of Newey and Tauchen if $W_t^o = E[u_t(\theta_o)u_t(\theta_o)'|x_t^*]$.

Since the asymptotic distribution of $W_T(\hat{\theta},\hat{M}_1,\hat{M}_2,\hat{D})$ is robust to its components, we may use different $\hat{M}_1$ and $\hat{M}_2$. Let us alternatively choose $\hat{M}_{O,1} = T^{-1}\Sigma_{t=1}^T\hat{\Delta}_t'(\hat{W}_t)^{-1}U_t(\hat{\theta})$ and $\hat{M}_{O,2} = T^{-1}\Sigma_{t=1}^T\Phi_t'(\hat{W}_t)^{-1}U_t(\hat{\theta})$. Then, it can be shown that the Wald statistic $W_T(\hat{\theta},\hat{M}_{O,1},\hat{M}_{O,2},\hat{D})$

numerically equals $OMJ_T(\hat{\theta})$ with $\Lambda_{b,t} = (\hat{W}_t)^{-1}\hat{\Delta}_t$ and $\Lambda_{c,t} = (\hat{W}_t)^{-1}\Phi_t$. That is, OMJ_T statistics (and also the Hansen statistic) for testing the hypothesis, $E\{[\Delta_t^o,\Phi_t]'(W_t^o)^{-1}u_t(\theta_o)\} = 0$, are asymptotically identical to those of Wooldridge's. This remarkable equivalence indicates that OMJ_T statistics are a generalization of the CM statistics of Newey, Tauchen and Wooldridge.

One important advantage of OMJ_T statistics over Wooldridge's and others is that they can be computed with any $\Lambda_{b,t}$. As Wooldridge (1990) noted, extensions of his approach to nonlinear simultaneous-equations models or rational expectation models might be limited, because for such models analytical computation of $\Delta_t(\theta)$ is complicated.[9] Furthermore, the Wald-type statistics of form (4.11) do not have particular power properties superior to OMJ_T statistics. Newey (1985a, 1985b) shows that an optimal CM test whose power dominates that of any other tests can be derived if information about the density function under alternative specifications is available. Otherwise, any two CM tests with the same degrees of freedom may dominate each other depending on the direction of local alternatives.

5. TESTING STRUCTURAL CHANGES

GMM tests of structural change in nonlinear models have been studied by Andrews and Fair (1988), Hoffman and Pagan (1989), and Ghysels and Hall (1990a, 1990b). These studies assume that a structural breakpoint is known where the sample is split into two

---

9 Nonparametric estimates of $\Delta_t(\theta_o)$ may be used when assumption SE holds. However, Wooldridge's method may be inappropriate for the cases where only assumption WE holds.

subsamples, and that both sizes of subsamples grow with that of the entire sample.[10] Andrews and Fair (1988) -- hereafter denoted by AF -- consider Wald, LR and Lagrangean multiplier (LM) tests for parameter stability over the entire sample. Hoffman and Pagan (1989) and Ghysels and Hall (1990a) -- hereafter HP and GH (1990a), respectively -- independently propose post-sample predictive tests which are based on out-of-sample moments evaluated at in-sample GMM estimates. Ghysels and Hall (1990b) -- hereafter GH (1990b) -- examine a LR-type statistic which is constructed similarly to D_T considered in section 3. This section provides a unified approach to these tests of structural stability, based on the results in sections 2 and 3.

Following the studies mentioned above, we assume that the sample is split into two subperiods:

(5.1)
$$\Upsilon_1 = \{t = 1, \dots, T_1\} \; ; \; \Upsilon_2 = \{t = T_1 + 1, \dots., T\} \; .$$

Let $T_2 = T - T_1$. Both T_1 and T_2 grow with T; that is, letting $\tau_1 = T_1/T$ and $\tau_2 = T_2/T$, we assume that $n_1 = \lim \tau_1 \neq 0$, $n_2 = \lim \tau_2 \neq 0$ and $n_1 + n_2 = 1$. Let $g_j(z_t,\theta)$ be the $q_j \times 1$ moment functions for $t \in \Upsilon_j$, where j = 1, 2. While q_1 and q_2 are equal in usual cases, we may allow them to be different. Consider the three individual hypotheses:

$$H_o^1 \colon E[g_1(z_t,\theta_o)] = 0 \;, \;\; \textit{for any } t \in \Upsilon_1 \; .$$

$$H_o^2 \colon E[g_2(z_t,\bar{\theta}_o)] = 0 \;, \;\; \textit{for any } t \in \Upsilon_2 \; .$$

---

[10] See Andrews (1993) and Dufour, Ghysels and Hall (1994), for tests in cases where the breakpoint is unknown.

$$H_o^3:\ \bar{\theta}_o = \theta_o\ .$$

We define the "Stability" hypothesis by:

(5.2) $$H_o^s:\ \textit{All of } H_o^1,\ H_o^2 \textit{ and } H_o^3 \textit{ hold}\ .$$

We assume that $q_1 \geq p$ so that θ_o can be consistently estimated with the first subsample only.

The asymptotic power of GMM tests may depend on the directions of possible violation of H_o^s. We may consider three types of alternative hypothesis:

$$H_A^A:\ \textit{At least one of } H_o^1,\ H_o^2 \textit{ and } H_o^3 \textit{ does not hold}\ .$$

$$H_A^B:\ H_0^1 \textit{ holds, but at least one of } H_o^2 \textit{ and } H_o^3 \textit{ does not}\ .$$

$$H_A^C:\ H_o^1 \textit{ and } H_o^2 \textit{ hold, but } H_o^3 \textit{ does not hold}\ .$$

Which of these three is the appropriate one to be tested against H_o^s should be determined on the basis of prior information about possible misspecifications. The alternative hypothesis H_A^C implies that the structural break has changed the true values of the parameters in θ. On the contrary, H_A^A or H_A^B may be appropriate in cases where possible sources of model misspecification other than parameter instability are also suspected. For example, consider a dynamic regression model with moving-average errors, which can be estimated by GMM using lagged dependent variables and other exogenous regressors as instrumental variables. Suppose that θ is the vector of coefficients of regressors, and that the $g_j(z_t,\theta)$ are the products of instrumental variables and residuals. There are several cases in which GMM estimators may appear to be unstable across two subsample periods. The first possible case is when misspecifications of the order of moving average lead to the use of contaminated instrumental variables for both subsamples. For this case, H_A^A would be the one that researchers should consider. The second case is when a structural break increases the order of moving average in the second period. Fewer instrumental variables remain legitimate for the second subsample-period, because some lagged dependent variables that are predetermined in the first subsample may become contaminated in the second period. Therefore, the GMM estimate of θ_o with the second subsample could be biased if the same instruments are used for both periods. That is, even if θ_o may remain constant over time, estimates of θ_o may vary significantly over subsamples. If this possibility is of concern, researchers may have to consider H_A^B.

Incorporating the three possible alternative hypotheses, we may consider three types of local alternatives under which asymptotic distributions of GMM statistics can be analyzed:

$$H_T^1\colon\ \sqrt{T_1}E[g_1(\theta_o)] \ = \ \omega_1 \ + \ o(1) \ ;$$

$$H_T^2\colon\ \sqrt{T_2}E[g_2(\theta_o)] \ = \ \omega_2 \ + \ o(1) \ ;$$

$$H_T^3\colon\ E[g_2(\bar{\theta}_T)] \ = \ 0 \ \textit{and} \ \sqrt{T_2}(\bar{\theta}_T - \theta_o) \ = \ \rho_2 \ ,$$

where $g_1(\theta)$ and $g_2(\theta)$ are subsample means of $g_1(z_t,\theta)$ and $g_2(z_t,\theta)$, respectively. We here drop the subscript "T" for notational convenience. Define $H_\ell^A = \{(H_T^1,H_T^2)\}_{T=1}^\infty$, $H_\ell^B = \{(H_o^1,H_T^2)\}_{T=1}^\infty$ and $H_\ell^C = \{(H_o^1,H_T^3)\}_{T=1}^\infty$. Note that H_ℓ^A encompasses H_ℓ^B while the latter subsumes H_ℓ^C.

We first consider the tests of H_o^s against H_A^A. The usual Hansen statistics (J_T) or its modified versions (MJ_T) may be used to test H_o^s. However, as GH (1990b) have shown, J_T has only a low power to detect violation of H_o^s when local alternatives are of the form H_ℓ^C. A standard response to this problem is to use the Hansen statistic constructed to test two sets of subsample orthogonality conditions jointly.[11] Define a dummy variable d_t such that $d_t = 1$ if $t \in \underline{\Upsilon}_1$ and $d_t = 0$ otherwise. Let $f_t(\theta) = [d_t g_1(z_t,\theta)', (1-d_t)g_2(z_t,\theta)']'$ and,

(5.3)
$$f(\theta) = \frac{1}{T}\sum_{t=1}^{T} f_t(\theta) = \begin{bmatrix} \tau_1 g_1(\theta) \\ \tau_2 g_2(\theta) \end{bmatrix} .$$

Denote the asymptotic covariance matrices of $\sqrt{T_1}\bar{g}_1(\theta_o)$ and $\sqrt{T_2}\bar{g}_2(\theta_o)$ by V_1 and V_2, respectively. Note that $\sqrt{T_1}\bar{g}_1(\theta_o)$ and $\sqrt{T_2}\bar{g}_2(\theta_o)$ are uncorrelated because $d_t(1-d_t) = 0$. Therefore, the asymptotic covariance of $\sqrt{T}\bar{f}(\theta_o)$ equals $V_f = \mathrm{diag}(n_1V_1,n_2V_2)$. Both V_1 and V_2 can be separately estimated within subsamples. Then, a consistent estimator of V_f can be computed based on the estimators of V_1 and V_2. For notational convenience, we denote the consistent estimator of V_f by $\hat{V}_f = \mathrm{diag}(\tau_1V_1,\tau_2V_2)$.

Let $\tilde{\theta}_f$ be the estimator which solves the problem: $\min_\theta Tf(\theta)'(\hat{V}_f)^{-1}f(\theta)$. Then, a Hansen statistic for testing H_o^s is obtained by:

(5.4)
$$SJ_T = Tf(\tilde{\theta}_f)'(\hat{V}_f)^{-1}f(\tilde{\theta}_f) = T_1 g_1(\tilde{\theta}_f)'(V_1)^{-1}g_1(\tilde{\theta}_f) + T_2 g_2(\tilde{\theta}_f)'(V_2)^{-1}g_2(\tilde{\theta}_f) ,$$

where the second equality results from the fact that $\hat{V}_f$ is block-diagonal. In order to find the asymptotic distribution of SJ_T under H_ℓ^A, we assume that each subsample satisfies the regularity conditions required for (2.5). Define $\omega_f = [\sqrt{n_1}\bar{\omega}_2',\sqrt{n_2}\bar{\omega}_2']'$. Then, it can be shown that:

(5.5)
$$\sqrt{T}f_T(\theta_o) \rightarrow N(\omega_f, V_f) .$$

Therefore, all the results obtained in sections 2 and 3 can apply to the tests of H_o^s. Define

---

[11] See Hamilton (1994) or GH (1990a).

$F(\theta)$ and F_o similarly to $G_T(\theta)$ and G_o, respectively. With these notations and (5.5), we can show that SJ_T has a noncentral chi-square distribution with $(q_a + q_b - p)$ degrees of freedom and noncentrality parameter $\lambda = \omega_f' Q(V_f,F_o)\omega_f$.

The modified Hansen statistic for testing H_o^s, which we denote by MSJ_T, is readily available. For j = 1, 2, define $\hat{g}_j = g_j(\hat{\theta})$ and $\hat{G}_j = G_j(\hat{\theta}) = \partial g_j(\theta)/\partial\theta'$ where $\hat{\theta}$ is any $\sqrt{T}$-consistent estimator. Let $\hat{f} = f(\hat{\theta}) = [\tau_1\hat{g}_1',\tau_2\hat{g}_2']'$ and $\hat{F} = F(\hat{\theta}) = [\tau_1\hat{G}_1',\tau_2\hat{G}_2']'$. For notational convenience, define:

$$\kappa(\hat{\theta}) = [\Sigma_j T_j \hat{g}_j'(V_j)^{-1}\hat{G}_j][\Sigma_j T_j \hat{G}_j'(V_j)^{-1}\hat{G}_j]^{-1}[\Sigma_j T_j \hat{G}_j'(V_j)^{-1}\hat{g}_j] , \tag{5.6}$$

where "Σ_j" means "$\Sigma_{j=1}^2$." Then, a little algebra shows that:

$$MSJ_T(\hat{\theta}) = \hat{f}'Q(\hat{V}_f,\hat{F})\hat{f} = T_1\hat{g}_1'(V_1)^{-1}\hat{g}_1 + T_2\hat{g}_2'(V_2)^{-1}\hat{g}_2 - \kappa(\hat{\theta}) . \tag{5.7}$$

Under H_ℓ^A, any MSJ_T is asymptotically identical to SJ_T whatever $\hat{\theta}$ is used.

When prior information about the first subsample period indicates that the moment condition for the first subsample period is legitimate, a test statistic of form D_T can be used, which we refer to as SD_T. Let $\check{\theta}_1$ be the optimal GMM estimator of θ_o under H_A^B. Then, we obtain:

$$SD_T = T_1 g_1(\tilde{\theta}_f)'(V_1)^{-1}g_1(\tilde{\theta}_f) + T_2 g_2(\tilde{\theta}_f)'(V_2)^{-1}g_2(\tilde{\theta}_f) - T_1 g_1(\check{\theta}_1)'(V_1)^{-1}g_1(\check{\theta}_1) , \tag{5.8}$$

which is proposed by GH (1990b). Alternatively, we may use the modified versions, MSD_T, of SD_T constructed similarly to MD_T statistics:

$$MSD_T(\hat{\theta}) = T_2\hat{g}_2'(V_2)^{-1}\hat{g}_2 + T_1\hat{g}_1'(V_1)^{-1}\hat{G}_1[\hat{G}_1'(V_1)^{-1}\hat{G}_1]^{-1}\hat{G}_1'(V_1)^{-1}\hat{g}_1 - \kappa(\hat{\theta}) . \tag{5.9}$$

When we use $\check{\theta}_1$ and $\tilde{\theta}_f$ to compute MSD_T statistics, we can obtain the statistics AN_T

or N_T. $MSD_T(\check{\theta}_1)$ is an analogue of AN_T. Using the fact that $G_1(\check{\theta}_1)'(V_1)^{-1}g_1(\check{\theta}_1) = 0$, we can show that:

[redacted]

where $\check{g}_2 = g_2(\check{\theta}_1)$, $\check{G}_1 = G_1(\check{\theta}_1)$ and $\check{G}_2 = G_2(\check{\theta}_1)$. This statistic is exactly the post-sample prediction statistic proposed by GH (1990a) and HA. This result implies that all the tests proposed by GH (1990a, 1990b) and HA are asymptotically identical.

Since $MSD_T(\tilde{\theta}_f)$ is an analogue of N_T, Proposition 3 of Newey (1985a) immediately applies. That is, $MSD_T(\tilde{\theta}_f)$ is an optimal GMM test which has maximum power toward H^B_ℓ. Therefore, SD_T and any MSD_T statistic should be also optimal. Both SD_T and MSD_T statistics have the same noncentral chi-square distribution with q_2 degrees of freedom and the noncentrality parameter, λ_{MSD}, equal to,

(5.11)
$$\omega_2'(V_2)^{-1}[V_2 - n_2 G_{2,o}\{\Sigma_j n_j G_{j,o}'(V_j)^{-1}G_{j,o}\}^{-1}G_{2,o}'](V_2)^{-1}\omega_2 ,$$

where $G_{j,o} = \text{plim } G_j(\theta_o)$, for j = 1, 2.

As a final step, we consider the tests of H^s_o against H^C_A. Assuming $p \le q_2$, let $\check{\theta}_2$ denote the estimator which solves: $\min_\theta T_2 g_2(\theta)'(V_2)^{-1}g_2(\theta)$. The LR-type statistic of AF is given by:

(5.12)
$$LR_T = \Sigma_j T_j g_j(\tilde{\theta}_f)'(V_j)^{-1}g_j(\tilde{\theta}_f) - \Sigma_j T_j g_j(\check{\theta}_j)'(V_j)^{-1}g_1(\check{\theta}_1) ,$$

which is asymptotically chi-squared with p degrees of freedom under H^s_o. Applying Proposition 1, we can obtain the modified LR statistics of the form:

(5.13)
$$MLR_T(\hat{\theta}) = \Sigma_j T_j \hat{g}_j'(V_j)^{-1}\hat{G}_j[\hat{G}_j'(V_j)^{-1}\hat{G}_j]^{-1}\hat{G}_j'(V_j)^{-1}\hat{g}_j - \kappa(\hat{\theta}) .$$

Under H^C_ℓ, it can be shown that (5.5) holds with $\omega_1 = 0$ and $\omega_2 = G_{2,o}\rho_2$. [See GH (1990b).] Therefore, Proposition 1 implies that for any $\sqrt{T}$-consistent $\hat{\theta}$, $MLR_T(\hat{\theta})$ is asymptotically identical to LR_T.

An interesting feature of the class of MLR_T statistics is that it subsumes all the other tests considered by AF. Note that $MLR_T(\tilde{\theta}_f)$ equals their LM statistic.[12] We can also establish the equivalence between their Wald statistic and $MLR_T(\check{\theta}_1)$. After a little algebra, we can show that:

$$\begin{aligned} MLR_T(\check{\theta}_1) &= T_2 g_2(\check{\theta}_1)'(V_2)^{-1}G_2(\check{\theta}_1)[G_2(\check{\theta}_1)'(V_2)^{-1}G_2(\check{\theta}_1)]^{-1} \\ &\quad \times \left\{[G_2(\check{\theta}_1)'(V_1)^{-1}G_2(\check{\theta}_1)]^{-1} + (T_2/T_1)[G_1(\check{\theta}_1)'(V_1)^{-1}G_1(\check{\theta}_1)]^{-1}\right\}^{-1} \\ &\quad \times [G_2(\check{\theta}_1)'(V_1)^{-1}G_2(\check{\theta}_1)]^{-1}G_2(\check{\theta}_1)'(V_2)^{-1}g_2(\check{\theta}_1) \; . \end{aligned} \tag{5.14}$$

We now define the one-step linearized GMM estimator:

$$\theta_2^L = \check{\theta}_1 - [G_2(\check{\theta}_1)'(V_2)^{-1}G_2(\check{\theta}_1)]^{-1}G_2(\check{\theta}_1)'(V_2)^{-1}g_2(\check{\theta}_1) \; . \tag{5.15}$$

Lemma 4 of Newey (1985a) implies that

$$\sqrt{T_2}(\theta_2^L - \check{\theta}_2) = o_p(1) \; . \tag{5.16}$$

Using the fact that $G_2(\check{\theta}_1)'(V_2)^{-1}G_2(\check{\theta}_1) = G_2(\check{\theta}_2)'(V_2)^{-1}G_2(\check{\theta}_2) + o_p(1)$, and substituting (5.15) and (5.16) into (5.14), we obtain:

$$MLR_T(\check{\theta}_1) = (\check{\theta}_2 - \check{\theta}_1)'[\Sigma_j\{T_j G_j(\check{\theta}_j)'(V_j)^{-1}G_j(\check{\theta}_j)\}^{-1}]^{-1}(\check{\theta}_2 - \check{\theta}_1) + o_p(1) \; , \tag{5.17}$$

where the first term on the right-hand side is exactly the Wald statistic of AF.

Following Proposition 3 of Newey (1985a), it can be shown that $MLR_T(\tilde{\theta}_f)$ is an

---

[12] Observe that $\kappa(\tilde{\theta}_f) = 0$.

optimal GMM statistic which has maximum power toward H_ℓ^C. This result and the equivalence between LR_T and any MLR_T statistic imply that all of the Wald, LR and LM tests of AF are optimal under H_ℓ^C. All of the MLR_T statistics have the same noncentral chi-square distribution with p degrees of freedom and the noncentrality parameter, λ_{MLR}, equal to,

(5.18) $$\omega_2^{*\prime}(V_2)^{-1}[V_2 - n_2 G_{2,o}\{\Sigma_j n_j G_{j,o}'(V_j)^{-1}G_{j,o}\}^{-1}G_{2,o}'](V_2)^{-1}\omega_2^* \ ,$$

where $\omega_2^* = G_{2,0}\rho_2$. If $q_2 = p$, both MSD_T and MLR_T statistics are asymptotically identical because $SD_T = LR_T$.[13] On the contrary, if $q_2 > p$, MSD_T statistics have more degrees of freedom than MLR_T statistics even if both statistics have the same noncentrality parameter under H_ℓ^C. These results imply that when $q_2 > p$, MLR_T (or LR_T) statistics have a better power against H_ℓ^C than MSD_T (or SD_T) statistics. However, it is also important to note that MSD_T statistics can have a better power to detect misspecifications other than parameter instability.

6. CONCLUDING REMARKS

This paper has developed alternative GMM tests which can be obtained using any $\sqrt{T}$-consistent estimator. The tests are robust to the distributions of the estimators used, and share the same asymptotic power properties with the tests based on efficient estimators. The alternative tests are also easy to perform. In particular, the statistics for testing orthogonality conditions can be computed by auxiliary regressions when the sample moments involved are serially uncorrelated. In the context of CM testing, the approach of this paper can be

---

[13] Observe that $G_2(\check{\theta}_2)'(V_2)^{-1}g_2(\check{\theta}_2) = 0$ if $q_2 = p$.

regarded as a generalization of Wooldridge's (1990, 1991). However, the former can apply to a wider range of econometric models than the latter.

This paper has focused on two types of tests: one class is for the tests asymptotically equivalent to the Hansen test, and the other for the tests equivalent to the EHS test. These two types of testing procedures have been extended to the contexts of CM testing and structural stability testing. It has been shown that many existing tests are members of the two classes computed with particular estimators. Therefore, the approach of this paper is general enough to encompass those of many previous studies.

The test procedures developed in this paper could be extended to many other cases. For example, they may apply to tests of structural change in cases where the structural breakpoint is unknown. Recently, Andrews (1993) developed a test which applies to such cases. Using his approach, researchers should first determine all the possible breakpoints and compute the statistics (LR, Wald or LM) of AF corresponding to each point. Then, the maximum value of these statistics serves as the statistic for testing the unknown structural break. Unfortunately, the Andrews method is computationally expensive, because each statistic for possible breakpoints requires a different estimator. The procedures developed in this paper might be used to simplify the Andrews method, because with them all of the statistics required for his method can be computed with the same $\sqrt{T}$-consistent estimator. Further study on this line would be worth pursuing.

APPENDIX

PROOF OF PROPOSITION 3. Similarly to the proof of Proposition 1, we can easily establish the asymptotic equivalence between $T\check{b}_T'(V_{bb})^{-1}\check{b}_T$ and $T\hat{b}_T'Q(V_{bb},\hat{B}_{T,1})\hat{b}_T$. Therefore, the equivalence between D_T and $MD_T(\hat{\theta})$ immediately follows. We now determine the sign of $MD_T(\hat{\theta})$. Define $B_m(\theta) = \partial b_m(\theta)/\partial\theta'$ and $C_m(\theta) = \partial c_m(\theta)/\partial\theta'$; and let $\hat{B}_m = B_m(\hat{\theta})$. Then, Proposition 2 implies that $\hat{b}_T'Q(V_{bb},\hat{B}_T)\hat{b}_T = \hat{b}_m'Q(V^m_{bb},\hat{B}_m)\hat{b}_m$. That is,

$$MD_T = T[\hat{g}_T'Q(V,\hat{G}_T)\hat{g}_T - \hat{b}_m'Q(V^m_{bb},\hat{B}_m)\hat{b}_m] . \tag{A.1}$$

Let $\Omega_m = V^m_{cc} - V^m_{cb}(V^m_{bb})^{-1}V^m_{bc}$. Define $r_m(\theta) = c_m(\theta) - V^m_{cb}(V^m_{bb})^{-1}b_m(\theta)$ and $R_m(\theta) = C_m(\theta) - V^m_{cb}(V^m_{bb})^{-1}B_m(\theta)$; and let $\hat{r}_m = r_m(\hat{\theta})$ and $\hat{R}_m = R_m(\hat{\theta})$. Let SSF_1 be the sum of squared fitted values (SSF) from a regression of $[\hat{b}_m'(V^m_{bb})^{-½},\hat{r}_m'(\Omega_m)^{-½}]'$ on $S_1 = \mathrm{diag}[(V^m_{bb})^{-½}\hat{B}_m,(\Omega_m)^{-½}]$, and let SSF_2 be SSF from a regression of $[\hat{b}_m'(V^m_{bb})^{-½},\hat{r}_m'(\Omega_m)^{-½}]'$ on $S_2 = [\hat{B}_m'(V^m_{bb})^{-½},\hat{R}_m'(\Omega_m)^{-½}]'$. After some algebra, we can show that $MD_T(\hat{\theta})/T$ equals:

$$\begin{aligned} &\hat{r}_m'\Omega_m^{-1}\hat{r}_m + \hat{b}_m'(V^m_{bb})^{-1}\hat{B}_m[\hat{B}_m'(V^m_{bb})^{-1}\hat{B}_m]^{-1}\hat{B}_m'(V^m_{bb})^{-1}\hat{b}_m \\ &- [\hat{b}_m'(V^m_{bb})^{-1}\hat{B}_m+\hat{r}_m'\Omega_m^{-1}\hat{R}_m][\hat{B}_m'(V^m_{bb})^{-1}\hat{B}_m+\hat{R}_m'\Omega_m^{-1}\hat{R}_m]^{-1}[\hat{B}_m'(V_{bb})^{-1}\hat{b}_m+\hat{R}_m'\Omega_m^{-1}\hat{r}_m] \\ &= SSF_1 - SSF_2 . \end{aligned} \tag{A.2}$$

Since all the columns of S_2 lie in the column space of S_1, $SSF_1 \geq SSF_2$. ■

The following lemma is useful to prove Proposition 3.

LEMMA 1. We assume that $p_2 \neq 0$. Let $\Psi_m(\theta) = \Omega_m - R_m(\theta)[G_T(\theta)'V^{-1}G_T(\theta)]^{-1}R_m(\theta)'$; and define:

(A.3)
$$N_T^m = \tilde{r}_m'(\tilde{\Psi}_m)^{-1}\tilde{r}_m ,$$

where $\tilde{r}_m = r_m(\tilde{\theta})$ and $\tilde{\Psi}_m = \Psi_m(\tilde{\theta})$. Then, $N_T^m = N_T$.

PROOF. Let $L_1 = [0(q_c,q_b),I(q_c,q_c)]$, where $0(q_c,q_b)$ is the $q_c \times q_b$ zero matrix and $I(q_c,q_c)$ is the $q_c \times q_c$ identity matrix; and let $L_2 = [0(q_c\text{-}p_2,q_b+p_2), I(q_c\text{-}p_2,q_c\text{-}p_2)]$. Let $P(M) = M(M'M)^+M'$ denote the projection onto the column space of an arbitrary matrix M. Then, using the fact that $\tilde{g}_T = VQ(V,\tilde{G}_T)\tilde{g}_T$, we can show:

(A.4)
$$N_T = T\tilde{g}_T'V^{-1/2}P(U_1)V^{-1/2}\tilde{g}_T \; ; \; N_T^m = T\tilde{g}_T'V^{-1/2}P(U_2)V^{-1/2}\tilde{g}_T ,$$

where $U_1 = V^{1/2}Q(V,\tilde{G}_T)L_1'$ and $U_2 = V^{1/2}Q(V,\tilde{G}_T)L_2'$. Since the columns of $\tilde{G}_T$ form a base for the null space of $Q(V,\tilde{G}_T)$, Lemma A.5 of Newey (1985a) implies that $\mathrm{Rank}(U_1) = \mathrm{Rank}(\tilde{G}_T,L_1') - p = q_c - p_2$. Similarly, we can show that $\mathrm{Rank}(U_2) = q_c - p_2$. Therefore, both matrices U_1 and U_2 have the same rank. Then since all the columns of U_2 are in U_1, we must have $P(U_2) = P(U_1)$. ■

PROOF OF PROPOSITION 4. The first-order condition for $\tilde{\theta}$ and a little algebra show that:

(A.5)
$$\tilde{G}_T'V^{-1}\tilde{g}_T = \tilde{B}_m'(V_{bb}^m)^{-1}\tilde{b}_m + \tilde{R}_m'\Omega_m^{-1}\tilde{r}_m = 0 .$$

Substituting (A.5) into (A.2) yields:

(A.6)
$$\begin{aligned}\frac{1}{T}MD_T(\tilde{\theta}) &= \tilde{r}_m'\Omega_m^{-1}\tilde{r}_m + \tilde{b}_m'(V_{bb}^m)^{-1}\tilde{B}_m[\tilde{B}_m'(V_{bb}^m)^{-1}\tilde{B}_m]^{-1}\tilde{B}_m'(V_{bb}^m)^{-1}\tilde{b}_m \\ &= \tilde{r}_m'\Omega_m^{-1}[\Omega_m + \tilde{R}_m[\tilde{B}_m'(V_{bb}^m)^{-1}\tilde{B}_m]^{-1}\tilde{R}_m']\Omega_m^{-1}\tilde{r}_m .\end{aligned}$$

Applying the usual matrix inversion rule, we obtain:

$$\text{(A.7)} \qquad \Omega_m^{-1}\tilde{\Psi}_m\Omega_m^{-1} = [\Omega_m + \tilde{R}_m[\tilde{B}_m'(V_{bb}^m)^{-1}\tilde{B}_m]^{-1}\tilde{R}_m]^{-1} .$$

Substituting (A.7) into (A.6) and applying Lemma 1 yield part (i). Similarly, we can show part (ii). ■

PROOF OF PROPOSITION 5. Define $F = [F_b', F_c']'$, where $F_b = [I_{p\times p}, 0_{p\times(q-p)}]$ and $F_c = [-(\Sigma_{t=1}^T\Lambda_{c,t}'\hat{U}_t)(\Sigma_{t=1}^T\Lambda_{b,t}'\hat{U}_t)^{-1}, I_{(q-p)\times(q-p)}]$. Let $g_F = F\hat{g}_T$, $G_F = F\hat{G}_T$ and $V_F = F\hat{V}_TF'$, where $\hat{g}_T = g_T(\hat{\theta})$, $\hat{G}_T = G_T(\hat{\theta})$ and $\hat{V}_T = V_T(\hat{\theta})$. Since F is a nonsingular square matrix, we should have:

$$\text{(A.8)} \qquad OMJ_T(\hat{\theta}) = Tg_F'[(V_F)^{-1} - (V_F)^{-1}G_F(G_F'(V_F)^{-1}G_F)^{-1}G_F'(V_F)^{-1}]g_F .$$

Note that

$$\text{(A.9)} \qquad g_F = \begin{bmatrix} F_b\hat{g}_T \\ F_c\hat{g}_T \end{bmatrix};\ G_F = \begin{bmatrix} F_b\hat{G}_T \\ 0_{(q-p)\times p} \end{bmatrix};\ V_F = \begin{bmatrix} F_b\hat{V}_TF_b' & F_b\hat{V}_TF_c' \\ F_c\hat{V}_TF_b' & F_c\hat{V}_TF_c' \end{bmatrix} .$$

Substituting (A.9) into (A.8), and using the fact that $F_b\hat{G}_T$ is invertible, we can show that

$$\text{(A.10)} \qquad OMJ(\hat{\theta}) = T(F_c\hat{g}_T)'(F_c\hat{V}_TF_c')^{-1}(F_c\hat{g}_T) .$$

Then since $F_c\hat{g}_T = T^{-1}\Sigma_{t=1}^T\Lambda_{b,t}^{r}{}'\hat{u}_t$ and $F_cV_TF_c' = T^{-1}\Sigma_{t=1}^T\Lambda_{c,t}^{r}{}'\hat{u}_t\hat{u}_t'\Lambda_{c,t}^r$, the conclusion of Proposition 5 results. ■

REFERENCES

Ahn, S.C. and P. Schmidt, 1995, A separability result for GMM estimation, with applications to GLS prediction and conditional moment tests, Econometric Reviews, 14, 19-34.

Amemiya, T., 1974, The nonlinear two-stage least-squares estimator, Journal of Econometrics, 2, 105-110.

Amemiya, T., 1977, The maximum likelihood and the nonlinear three stage least squares estimator in the general nonlinear simultaneous equation models, Econometrica, 45, 955-968.

Andrews, D. W. K., 1993, Tests for parameter instability and structural change with unknown change point, Econometrica, 61, 821-856.

Andrews, D. W. K. and R.C. Fair, 1988, Inference in nonlinear econometric models with structural change, Review of Economic Studies, 55, 615-639.

Andrews, D. W. K. and J. C. Monahan, 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, Econometrica, 60, 953-966.

Chamberlain, G., 1987, Asymptotic efficiency in estimation with conditional moment restrictions, Journal of Econometrics, 34, 305-334.

Chamberlain, G. 1992, Comment: sequential moment restrictions in panel data, Journal of Business & Economic Statistics, 10, 20-26.

Dufour, J. M., E. Ghysels and A. Hall, 1994, Generalized predictive tests and structural change analysis in econometrics, International Economic Review, 35, 199-229.

Eichenbaum, M. S., L. P. Hansen and K. J. Singleton, 1988, A time series analysis of representative agent models of consumption and leisure choice under uncertainty, The Quarterly Journal of Economics, 103, pp. 51 - 78.

Gallant, A. R. and D. W. Jorgenson, 1979, Statistical inference for a system of simultaneous, non-linear, implicit equations in the context of instrumental variable estimation, Journal of Econometrics, 11, 275 - 302.

Gallant, A. R. and H. White, 1988, A unified theory of estimation and inference for nonlinear dynamic models (Basil Blackwell, Oxford).

Gourieroux, C., A. Monfort, and A. Trognon, 1984, Pseudo-maximum likelihood methods: theory, Econometrica, 52, 681-700.

Ghysels, E. and A. Hall, 1990a, A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator, International Economic Review, 31, 355-364.

Ghysels, E. and A. Hall, 1990b, Are consumption-based intertemporal capital asset pricing models structural?, Journal of Econometrics, 45, 121-139.

Hamilton, J. D., 1994, Time series analysis (Princeton University press, Princeton, NJ).

Hansen, L. P., 1982, Large sample properties of generalized method of moments estimators, Econometrica, 50, 1029-1054.

Heaton, J., 1995, An empirical investigation of asset pricing with temporally dependent preference specifications, Econometrica, 63, 681-717.

Hoffman, D. and A. Pagan, 1989, Post-sample prediction tests for generalized method of moments estimators, Oxford Bulletin of Economics and Statistics, 51, 333-343.

Newey, W. K., 1985a, Generalized method of moments specification testing, Journal of Econometrics, 29, 229 - 256.

Newey, W. K., 1985b, Maximum likelihood specification testing and conditional moment tests, Econometrica, 53, 1047-1070.

Newey, W. K. and K. D. West, 1987a, Hypothesis testing with efficient method of moments estimation, International Economic Review, 28, 777 - 787.

Newey, W. K. and K. D. West, 1987b, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica, 55, 703 - 708.

Rao, C. R. and S. K. Mitra, 1971, Generalized inverse of matrices and its applications (Wiley, New York, NY).

Schmidt, P., 1990, Three-stage least squares with different instruments for different equations, Journal of Econometrics, 43, 389-394.

Tauchen, G., 1985, Diagnostic testing and evaluation of maximum likelihood models, Journal of Econometrics, 30, 415-443.

White, H., 1980, Nonlinear regression on cross section data, Econometrica, 48, 721-746.

White, H., 1994, Estimation, inference and specification analysis (Cambridge University Press, New York, NY).

Wooldridge, J. M., 1990, A unified approach to robust, regression-based specification tests, Econometric Theory, 6, 17-43.

Wooldridge, J. M., 1991, On the application of robust, regression-based diagnostics to models of conditional means and conditional variances, Journal of Econometrics, 47, 5-46.