working paper
department
of economics

A UNIFIED APPROACH TO ROBUST, REGRESSION-BASED SPECIFICATION TESTS

by

Jeffrey M. Wooldridge

No. 480 January 1988

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

A UNIFIED APPROACH TO ROBUST, REGRESSION-BASED SPECIFICATION TESTS

by

Jeffrey M. Wooldridge

No. 480 January 1988

A UNIFIED APPROACH TO ROBUST, REGRESSION-BASED SPECIFICATION TESTS*

Jeffrey M. Wooldridge
Department of Economics
Massachusetts Institute of Technology, E52-262C
Cambridge, MA 02139
(617) 253-3488

January 1988

*I am grateful to Halbert White for helpful comments and suggestions.

ABSTRACT: This paper develops a new approach to robust specification testing for dynamic econometric models. A novel feature of these tests is that, in addition to the estimation under the null hypothesis, computation requires only a matrix linear least squares regression and then an ordinary least squares regression similar to those employed in popular nonrobust tests. The statistics proposed here are robust to departures from distributional assumptions that are not being tested. Moreover, the statistics may be computed using any $\sqrt{T}$-consistent estimator. Several examples are presented to illustrate the generality of the procedure. Among these are conditional mean tests for models estimated by weighted nonlinear least squares which do not require correct specification of the conditional variance, and tests of conditional means and variances estimated by quasi-maximum likelihood under nonnormality. Also, some new, computationally simple tests for the tobit model are proposed.

1. Introduction

Specification testing has become an integral part of the econometric model building process. The literature is extensive, and model diagnostics are available for most procedures used by applied econometricians. By far the most popular specification tests are those that can be computed using ordinary least squares regressions. Examples are the Lagrange Multiplier (LM) test, versions of Hausman's [8] specification tests, White's [14] information matrix (IM) test, and an LM version of the Davidson-MacKinnon [3] test for nonnested hypotheses. In fact, Newey [10] and White [16] have shown that most of these tests are asymptotically equivalent to one of the conditional moment (CM) tests considered by Newey [10], Tauchen [11], and White [16]. In the maximum likelihood setting with independent observations, Newey [10] has shown how to compute CM tests using auxiliary regressions. White [16] has extended Newey's results to a general dynamic setting.

The simplicity of the regression-based procedures currently used is not without cost. In many cases the validity of these tests relies on certain auxiliary assumptions holding in addition to the relevant null hypothesis. For example, in a nonlinear regression framework where the dynamic regression function is correctly specified under the null hypothesis, the usual LM regression-based statistic is invalid in the presence of conditional heteroskedasticity. The regression form that falls out of Newey [10] or White [16] is also usually invalid. Other examples are the various tests for heteroskedasticity: currently used regression forms require constancy of the conditional fourth moment of the regression errors under the null hypothesis. Finally, LM and other CM tests for jointly parameterized conditional means and variances are inappropriate under nonnormality. All of these situations are characterized by the same feature: validity of the tests requires imposition of more than just the hypotheses of interest under H_0. Furthermore, traditional testing procedures require that the estimators used to compute the statistics are efficient in some sense under the null hypothesis. It is important to stress that this is not merely nitpicking about regularity conditions.

Due primarily to the work of White [12,13,14,15], Domowitz and White [4], Hansen [6], and Newey [10], there now exist general methods of computing robust statistics. Unfortunately, for general classes of specification tests, computing robust versions using currently available methods is burdensome. This is particularly true of LM-like tests where, at least based on currently available formulas, analytically solving for the derivative of the implicit constraint function and computing generalized inverses are needed for computation. Several authors have even claimed that, contrary to the case of the Wald statistic, there are no useful robust forms of the LM statistic.

It is a safe bet that the substantial analytical and computational work required to obtain robust statistics is the reason they appear relatively infrequently in applied econometric work. Evidence of this statement is the growing use of the White

[12] heteroskedasticity-robust t-statistics, which are now computed by many econometrics packages. In the same papers one rarely sees an LM test, a Hausman test, or a nonnested hypothesis test carried out in a manner that is robust to second moment misspecification. This is unfortunate since these tests are inconsistent for the alternative that the conditional mean is correctly specified but the conditional variance has been misspecified. In other words, the standard forms of well known tests can result in inference with the wrong asymptotic size while having no systematic power for testing the auxiliary assumptions that are imposed in addition to H_0.

This paper develops a unified approach to calculating robust statistics which I believe is easily accessible to applied econometricians. It is shown that a general class of tests can be obtained using only linear least squares regressions. These tests maintain only the hypotheses of interest under the null, and are applicable to specification testing of dynamic multivariate conditional means and/or conditional variances without imposing further assumptions on the conditional distribution (except regularity conditions). In classical situations, these tests are asymptotically equivalent to their traditional counterparts under the additional assumptions needed to make the standard tests valid. Moreover, because the statistics may be computed using any $\sqrt{T}$-consistent estimator, the methodology leads to some interesting new tests in cases where the computational burden based on previous approaches is prohibitive.

The remainder of the paper is organized as follows. Section 2 discusses the setup and the general results, Section 3 illustrates the scope of the methodology with several examples, and Section 4 contains concluding remarks. Regularity conditions and proofs are contained in an appendix.

2. General Results

Let $\{(Y_t,Z_t): t=1,2,...\}$ be a sequence of observable random vectors with Y_t $1\times J$, Z_t $1\times K$. Y_t is the vector of endogenous variables. Interest lies in explaining Y_t in terms of the explanatory variables Z_t and (in a time series context) past values of Y_t and Z_t. For time series applications, let $X_t \equiv (Z_t,Y_{t-1},Z_{t-1},...,Y_1,Z_1)$ denote the predetermined variables and let $\mathcal{S}_t \subset \mathbb{R}^{K+(J+K)(t-1)}$ denote the support of X_t. For cross section applications, set $X_t \equiv Z_t$.

The conditional distribution of Y_t given $X_t = x_t$ always exists and is denoted $D_t(\cdot|x_t)$. Assume that the researcher is interested in testing hypotheses about a certain aspect of D_t, for example the conditional expectation and/or the conditional variance. Note that, because at time t the conditioning set contains $\{(Y_{t-1},Z_{t-1}),...,(Y_1,Z_1)\}$, the assumption is that interest lies in getting the dynamics of the relevant aspects of D_t correctly specified. For cross section applications, this point is of course irrelevant.

Many specification tests, including those for conditional means and variances, have asymptotically equivalent versions that can be derived as follows. Let $\eta_t(Y_t,X_t,\theta)$ be an $L\times 1$ random function defined on a parameter set $\Theta \subset \mathbb{R}^P$, and let $\varphi_t(X_t,\theta)$ be an $L\times 1$

function also defined on Θ. Note that η_t depends on Y_t whereas φ_t depends only on the predetermined variables. The null hypothesis of interest is expressed as

$$H_0: \quad E[\eta_t(Y_t,X_t,\theta_o)|X_t] = \varphi_t(X_t,\theta_o), \text{ for some } \theta_o \in \Theta, \qquad (2.1)$$
$$t=1,2,\ldots.$$

The leading case, and the one emphasized in this paper, is when $\{\varphi_t(x_t,\theta): x_t \in \mathscr{S}_t, \theta \in \Theta\}$ is a parameterized family for the conditional mean and/or conditional variance of Y_t given $X_t = x_t$.

The validity of (2.1) can be tested by choosing functions of the predetermined variables X_t and checking whether the sample covariances between these functions and

$$\phi_t(Y_t,X_t,\theta_o) \equiv \eta_t(Y_t,X_t,\theta_o) - \varphi_t(X_t,\theta_o)$$

are significantly different from zero. It is useful to allow the indicators to depend on θ and some nuisance parameters. Let $\pi \in \Pi$ denote a Nx1 vector of nuisance parameters, and let $\delta \equiv (\theta',\pi')'$ be the Mx1 vector of all parameters where $M \equiv P+N$. Let $\Lambda_t(X_t,\delta)$ be an LxQ matrix and let $C_t(X_t,\delta)$ be an LxL, symmetric and positive semi-definite matrix. Assume the availability of an estimator $\hat{\theta}_T$ such that $T^{1/2}(\hat{\theta}_T - \theta_o) = O_p(1)$ under H_0. Also assume that the nuisance parameter estimator $\hat{\pi}_T$ is such that $T^{1/2}(\hat{\pi}_T - \pi_T^o) = O_p(1)$, where $\{\pi_T^o: T=1,2,\ldots\}$ is a nonstochastic sequence in Π. Then a computable test statistic is the Qx1 vector

$$T^{-1}\sum_{t=1}^{T} \hat{\Lambda}_t'\hat{C}_t\hat{\phi}_t = T^{-1}\sum_{t=1}^{T} \hat{\Lambda}_t'\hat{C}_t(\hat{\eta}_t - \hat{\varphi}_t) \qquad (2.2)$$

where "^" denotes that each function is evaluated at $\hat{\theta}_T$ or $\hat{\delta}_T \equiv$

$(\hat{\theta}_T', \hat{\pi}_T')'$ (note that the dependence of the summands in (2.2) on the sample size T is suppressed). From a theoretical standpoint, the p.s.d. matrix C_t could of course be absorbed into Λ_t, and φ_t could be absorbed into η_t, but the structure in (2.2) is exploited below to generate regression-based tests with the additional property that they are asymptotically equivalent to standard tests under classical circumstances.

To use (2.2) as a basis for a test of (2.1), the limiting distribution of

$$\hat{\xi}_T \equiv T^{-1/2} \sum_{t=1}^{T} \hat{\Lambda}_t' \hat{C}_t \hat{\phi}_t \tag{2.3}$$

under H_0 is needed. In general, finding the asymptotic distribution of $\hat{\xi}_T$ entails finding the limiting distribution of

$$\xi_T^o \equiv T^{-1/2} \sum_{t=1}^{T} \Lambda_t^{o\prime} C_t^o \phi_t^o \tag{2.4}$$

(values with "o" superscripts are evaluated at θ_o or $\delta_T^o \equiv (\theta_o', \pi_T^{o\prime})'$) and the limiting distribution of $T^{1/2}(\hat{\theta}_T - \theta_o)$ (the limiting distribution of $T^{1/2}(\hat{\pi}_T - \pi_T^o)$ does not affect the limiting distribution of $\hat{\xi}_T$ under H_0). Because ξ_T^o is the standardized sum of a vector martingale difference sequence under H_0, its limiting distribution is frequently derivable from a central limit theorem. In standard cases $T^{1/2}(\hat{\theta}_T - \theta_o)$ will also be asymptotically normal. Given the asymptotic covariance matrices of ξ_T^o and $T^{1/2}(\hat{\theta}_T - \theta_o)$ and differentiability assumptions on Λ_t, C_t, and ϕ_t, it is possible to derive the asymptotic covariance matrix of $\hat{\xi}_T$ by the usual mean value expansion. In principle, deriving a quadratic form in $\hat{\xi}_T$ which has an asymptotic χ^2 distribution is straightforward. But nothing guarantees that the resulting test statistic is easy to compute.

In specific instances test statistics based on $\hat{\xi}_T$ can be computed from simple OLS regressions. For example, Newey [10] and White [16] have shown how statistics based on covariances of the form (2.2) can be computed from simple auxiliary regressions when $\hat{\theta}_T$ is the maximum likelihood estimator and the conditional density is correctly specified under H_0.

In general, the regression-based statistics appearing in the literature have the drawback that they are not robust to certain departures from distributional assumptions. For example, suppose interest lies in testing hypotheses about the conditional expectation of Y_t (taken to be a scalar for simplicity) given X_t. The parametric model is

$$\{m_t(x_t,\theta): x_t \in \mathscr{X}_t, \ \theta \in \Theta\}, \tag{2.5}$$

where $\Theta \subset \mathbb{R}^P$, and the null hypothesis is

$$H_0: E(Y_t|X_t) = m_t(X_t,\theta_o), \text{ some } \theta_o \in \Theta, \ t=1,2,\ldots. \tag{2.6}$$

Setting $L \equiv 1$, $C_t(X_t,\delta) \equiv 1$, $\eta_t(Y_t,X_t,\theta) \equiv Y_t$, and $\varphi_t(X_t,\theta) \equiv m_t(X_t,\theta)$ in (2.1) yields a class of tests based on

$$T^{-1} \sum_{t=1}^{T} \lambda_t(X_t,\hat{\delta}_T)'\hat{U}_t \tag{2.7}$$

where $\hat{U}_t \equiv Y_t - m_t(X_t,\hat{\theta}_T)$, $\hat{\theta}_T$ is the nonlinear least squares (NLLS) estimator, $\lambda_t(X_t,\delta)$ is a 1xQ vector function of misspecification indicators, and δ is a vector containing θ and possibly other nuisance parameters. The standard LM approach leads to a test based on the (uncentered) R^2 from the regression

$$\hat{U}_t \quad \text{on} \quad \nabla_\theta \hat{m}_t, \ \hat{\lambda}_t \qquad t=1,\dots,T. \tag{2.8}$$

Under H_0 <u>and</u> conditional homoskedasticity, TR^2 is asymptotically χ^2_Q. Thus, the LM approach effectively takes the null hypothesis to be

$$H_0': H_0 \text{ holds and } V(Y_t|X_t) = \sigma_o^2 \text{ for some } \sigma_o^2 > 0, \ t=1,2,\dots \tag{2.9}$$

but it is of course an inconsistent test for the alternative

H_1': H_0 holds but H_0' does not.

The regression form from Newey [10] and White [16] is

$$1 \quad \text{on} \quad \hat{U}_t \nabla_\theta \hat{m}_t, \ \hat{U}_t \hat{\lambda}_t \qquad t=1,\dots,T. \tag{2.10}$$

In general, H_0' is also required for TR^2 from this regression to be asmptotically χ^2_Q. [1]

There are many other examples where the goal is to test hypotheses about certain aspects of a distribution but auxiliary assumptions are maintained under the null hypothesis in order to obtain a simple regression-based test. Because the limiting distributions of test statistics can be sensitive to violations of the auxiliary assumptions, it is important to use robust forms of tests for which H_0 includes only the hypotheses of interest. But as mentioned above, applying the standard mean-value approach to the general statistic $\hat{\xi}_T$ results in a statistic for which computation can be prohibitively burdensome.

A relatively simple statistic is available if $\hat{\xi}_T$ is appropriately modified. Assume that $\theta_o \in \text{int}(\Theta)$ and that φ_t is differentiable on $\text{int}(\Theta)$. Then, instead of using the indicator $\hat{\Lambda}_t'\hat{C}_t$, the idea is to first purge from $\hat{C}_t^{1/2}\hat{\Lambda}_t$ its linear projection onto $\hat{C}_t^{1/2}\nabla_\theta\hat{\varphi}_t$. That is, consider the modified statistic

$$\ddot{\xi}_T \equiv T^{-1/2} \sum_{t=1}^{T} [\hat{C}_t^{1/2}\hat{\Lambda}_t - \hat{C}_t^{1/2}\nabla_\theta\hat{\varphi}_t\hat{B}_T]'\hat{C}_t^{1/2}\hat{\phi}_t \qquad (2.11)$$

where

$$\hat{B}_T \equiv \left(\sum_{t=1}^{T} \nabla_\theta\hat{\varphi}_t'\hat{C}_t\nabla_\theta\hat{\varphi}_t\right)^{-1} \sum_{t=1}^{T} \nabla_\theta\hat{\varphi}_t'\hat{C}_t\hat{\Lambda}_t \qquad (2.12)$$

is the PxQ matrix of regression coefficients from the regression

$$\hat{C}_t^{1/2}\hat{\Lambda}_t \quad \text{on} \quad \hat{C}_t^{1/2}\nabla_\theta\hat{\varphi}_t \qquad t=1,\ldots,T. \qquad (2.13)$$

Equation (2.11) can be written more succinctly as

$$\ddot{\xi}_T \equiv T^{-1/2} \sum_{t=1}^{T} \ddot{\Lambda}_t'\tilde{\phi}_t \qquad (2.14)$$

where $\{\ddot{\Lambda}_t: t=1,\ldots,T\}$ are the residuals from the regression in (2.13) and $\tilde{\phi}_t \equiv \hat{C}_t^{1/2}\hat{\phi}_t$. It is important to note that $\hat{\xi}_T$ and $\ddot{\xi}_T$ are <u>not</u> always asymptotically equivalent in the sense that $\hat{\xi}_T - \ddot{\xi}_T \overset{p}{\to} 0$ under H_0. In general, the indicators $\hat{\Lambda}_t'\hat{C}_t$ and $[\hat{\Lambda}_t - \nabla_\delta\hat{\varphi}_t\hat{B}_T]'\hat{C}_t$ are useful for checking different departures from (2.1). I return to this issue below.

Even when $\hat{\xi}_T$ and $\ddot{\xi}_T$ are not asympotically equivalent, $\ddot{\xi}_T$ can be used as the basis for a useful specification test. The computational simplicity of a limiting χ^2 quadratic form in $\ddot{\xi}_T$ is a consequence of the following theorem.

<u>Theorem 2.1</u>: Assume that the following conditions hold under H_0:

(i) Regularity conditions A.1 in the appendix;

(ii) For some $\theta_o \in \text{int}(\Theta)$,

(a) $E[\eta_t(Y_t,X_t,\theta_o)|X_t] = \varphi_t(X_t,\theta_o)$, t=1,2,...;

(b) $E[\nabla_\theta\eta_t(Y_t,X_t,\theta_o)|X_t] = 0$, t=1,2,...;

(c) $T^{1/2}(\hat{\delta}_T - \delta_T^o) = O_p(1)$.

Then

$$\ddot{\xi}_T = T^{-1/2} \sum_{t=1}^{T} [\Lambda_t^o - \nabla_\theta \varphi_t^o B_T^o]' C_t^o \phi_t^o + o_p(1)$$

where

$$B_T^o \equiv \left[\sum_{t=1}^{T} E[\nabla_\theta \varphi_t^{o\prime} C_t^o \nabla_\theta \varphi_t^o] \right]^{-1} \sum_{t=1}^{T} E[\nabla_\theta \varphi_t^{o\prime} C_t^o \Lambda_t^o].$$

In addition,

$$TR^2 \xrightarrow{d} \chi_Q^2,$$

where R^2 is the uncentered r-squared from the regression

$$1 \quad \text{on} \quad \hat{\phi}_t' \hat{C}_t [\hat{\Lambda}_t - \nabla_\theta \hat{\varphi}_t \hat{B}_T] \qquad t=1,\ldots,T \tag{2.15}$$

and $\hat{B}_T$ is given by (2.12). ■

Theorem (2.1) can be applied as follows:

(1) Given Λ_t, C_t, η_t, φ_t, and $\hat{\delta}_T$, compute $\hat{\Lambda}_t$, $\hat{C}_t$, $\hat{\eta}_t$, $\hat{\varphi}_t$, and $\nabla_\theta \hat{\varphi}_t$. Define $\tilde{\Lambda}_t \equiv \hat{C}_t^{1/2} \hat{\Lambda}_t$, $\nabla_\theta \tilde{\varphi}_t \equiv \hat{C}_t^{1/2} \nabla_\theta \hat{\varphi}_t$, and $\tilde{\phi}_t \equiv \hat{C}_t^{1/2} \hat{\phi}_t$;

(2) Run the matrix regression

$$\tilde{\Lambda}_t \quad \text{on} \quad \nabla_\theta \tilde{\varphi}_t \qquad t=1,\ldots,T \tag{2.16}$$

and save the residuals, say $\ddot{\Lambda}_t$;

(3) Run the regression

$$1 \quad \text{on} \quad \tilde{\phi}_t' \ddot{\Lambda}_t \qquad t=1,\ldots,T$$

and use TR^2 as asymptotically χ_Q^2 under H_0, assuming that $\ddot{\Lambda}_t$ does not contain redundant indicators. ■

Note that condition (ii.b) is an additional restriction on η_t that must be satisfied in order for (1)-(3) to be a valid procedure under H_0. This assumption rules out certain specification tests, but is applicable to the leading case of diagnostics for conditional means (hence conditional probabilities) and/or conditional variances. These are usually the cases where one would like to be robust against other distributional departures.[2]

Assumption (ii.c) is perhaps more properly listed as a regularity condition, but it is placed in the text to emphasize the generality of Theorem 2.1. Having a $\sqrt{T}$-consistent estimator of δ_T^o is a fairly weak requirement, and allows relatively simple specification tests when θ_o (as well as π_T^o) has been estimated by an inefficient procedure. An application to the tobit model is given in Section 3.

An important issue, mentioned earlier, is the relationship between $\ddot{\xi}_T$ and $\hat{\xi}_T$. There is a simple characterization of their asymptotic equivalence.

Lemma 2.2: Let the conditions of Theorem 2.1 hold. If, in addition,

$$\text{(iii)} \quad T^{-1/2} \sum_{t=1}^{T} \nabla_\theta \hat{\varphi}_t' \hat{C}_t (\hat{\eta}_t - \hat{\varphi}_t) = o_p(1),$$

then

$$\ddot{\xi}_T - \hat{\xi}_T = o_p(1). \quad \blacksquare \tag{2.17}$$

The importance of this lemma is that if (iii) holds then the modified indicator is testing for departures from H_0 in the same directions as the originally chosen indicator. When (2.17) holds, the statistics based on quadratic forms in $\ddot{\xi}_T$ and $\hat{\xi}_T$ are asymptotically equivalent. This is useful when comparing tests derived from Theorem 2.1 to more traditional forms of tests.

Condition (iii) is usefully interpreted as the sample covariance between $\{\hat{C}_t^{1/2}\nabla_\theta\hat{\varphi}_t$: t=1,...,T} and $\{\hat{C}_t^{1/2}\hat{\phi}_t$: t=1,...,T} being zero. It is trivially satisfied if

$$\sum_{t=1}^{T} \nabla_\theta\varphi_t(\theta)'C_t(\theta,\hat{\pi}_T)[\eta_t(\theta) - \varphi_t(\theta)] \equiv 0 \qquad (2.18)$$

is the defining first-order condition for $\hat{\theta}_T$. This is frequently the case, particularly when $\hat{\theta}_T$ is a quasi-maximum likelihood estimator (QMLE) of the parameters of a conditional mean (see Wooldridge [19]) or of the conditional mean and conditional variance (see Example 3.3 below). Note that in these cases (2.17) holds (trivially) for local alternatives. Therefore, the difference between the test based on $\ddot{\xi}_T$ and a more traditional nonrobust test based on $\hat{\xi}_T$ (e.g. an LM test) is simply that different estimators have been used for the moment matrix appearing in the quadratic form. Consequently, under the conditions required for the classical test to be valid, the two procedures are asymptotically equivalent under local alternatives. The robust test has the advantage of having a limiting noncentral χ^2 distribution even when the auxiliary assumptions are violated under local alternatives (e.g. heteroskedasticity is present in a dynamic regression model).

3. Examples of Regression-Based, Robust Tests

Example 3.1: Let Y_t be a scalar and let $\{m_t(x_t,\beta)$: $x_t \in \mathscr{X}_t$, $\beta \in B\}$, $B \subset \mathbb{R}^P$, be a parametric family for the conditional expectation of Y_t given X_t. The null hypothesis is

$$H_0: E(Y_t|X_t) = m_t(X_t,\beta_o), \text{ some } \beta_o \in B, \; t=1,2,.... \qquad (3.1)$$

Let $\{c_t(x_t,\alpha)$: $x_t \in \mathscr{X}_t$, $\alpha \in A\}$ be a sequence of weighting functions such that $c_t(x_t,\alpha) > 0$, and suppose that $\hat{\alpha}_T$ is an estimator such that $T^{1/2}(\hat{\alpha}_T - \alpha_T^o) = O_p(1)$, where $\{\alpha_T^o\} \subset A$. It is not assumed that $\{c_t(x_t,\alpha): x_t \in \mathscr{S}_t, \alpha \in A\}$ contains a version of $V(Y_t|X_t)$. The researcher merely chooses the set of weights $\{c_t(x_t,\hat{\alpha}_T)\}$ and performs weighted NLLS (WNLLS). The WNLLS estimator $\hat{\beta}_T$ solves

$$\sum_{t=1}^{T} \nabla_\beta m_t(\beta)'[Y_t - m_t(\beta)]/c_t(\hat{\alpha}_T) \equiv 0. \tag{3.2}$$

A general class of diagnostics is based on

$$\sum_{t=1}^{T} \lambda_t(\hat{\delta}_T)'[c_t(\hat{\alpha}_T)]^{-1}[Y_t - m_t(\hat{\beta}_T)] \tag{3.3}$$

where δ can contain β, α and other nuisance parameters. Letting $\theta \equiv \beta$, $C_t(\delta) \equiv [c_t(\alpha)]^{-1}$, $\eta_t(\theta) \equiv Y_t$, and $\varphi_t(\theta) \equiv m_t(\beta)$, it is easy to see that conditions (ii.a) and (ii.b) of Theorem 2.1 hold under H_0. Condition (ii.c) will also usually be satisfied. Because (iii) of Lemma 2.2 holds, the statistic obtained from Theorem (2.1) is asymptotically equivalent to the statistic based on (3.3). The following procedure is valid under H_0, without any assumptions about $V(Y_t|X_t)$ (except, of course, regularity conditions):

(i) Estimate β_o by WNLLS. Compute the residuals $\hat{U}_t$, the gradient $\nabla_\beta \hat{m}_t(\hat{\beta}_T)$, and the indicator $\lambda_t(\hat{\delta}_T)$. Define $\tilde{U}_t \equiv \hat{c}_t^{-1/2}\hat{U}_t$, $\nabla_\beta \tilde{m}_t \equiv \hat{c}_t^{-1/2}\nabla_\beta \hat{m}_t$, and $\tilde{\lambda}_t \equiv \hat{c}_t^{-1/2}\hat{\lambda}_t$;

(ii) Regress $\tilde{\lambda}_t$ on $\nabla_\beta \tilde{m}_t$ and keep the residuals, say $\ddot{\lambda}_t$;

(iii) Regress 1 on $\tilde{U}_t\ddot{\lambda}_t$ and use TR^2 from this regression as asymptotically χ^2_Q under H_0.

The indicator $\hat{\lambda}_t$ can be chosen to yield heteroskedasticity-robust LM tests, Hausman tests based on two WNLLS regressions which do not assume that either estimator is relatively efficient, and tests of nonnested hypotheses, such as the Davidson-MacKinnon [3] test, which are valid in the presence of heteroskedasticity. These tests are considered in more detail in Wooldridge [19].

Example 3.2: Suppose now, in the context of Example 3.1, $c_t(\alpha)$ is set to 1 and the goal is to test the assumption of homoskedasticity (actually, the goal is to test the joint assumption of correctness of the conditional mean and homoskedasticity). In particular, the null hypothesis is

$$H_0: \ E(Y_t|X_t) = m_t(X_t,\beta_o), \ V(Y_t|X_t) = \sigma_o^2 \quad \text{some } \beta_o \in B, \quad \text{some } \sigma_o^2 > 0, \ t=1,2,\ldots. \tag{3.4}$$

In the notation of Theorem 2.1, $\theta \equiv (\beta',\sigma^2)'$. Let $\hat{\beta}_T$ be the NLLS estimator, and let $\hat{U}_t$ be the NLLS residuals. Let $\lambda_t(X_t,\delta)$ be a $1\times Q$ vector of indicators. Most tests for heteroskedasticity are based on a statistic of the form

$$T^{-1}\sum_{t=1}^{T}\hat{\lambda}_t'[\hat{U}_t^2 - \hat{\sigma}_T^2] \tag{3.5}$$

where $\hat{\sigma}_T^2$ is the usual estimator $T^{-1}\sum_{t=1}^{T}\hat{U}_t^2$. Choosing $\lambda(X_t,\delta)$ to be the nonconstant, nonredundant elements of vech $[\nabla_\beta m_t(\beta)'\nabla_\beta m_t(\beta)]$ leads to the White [12] test for heteroskedasticity. Choosing $\lambda_t(X_t,\delta) \equiv X_{t1}$, where X_{t1} is a $1\times Q$ subvector of X_t, leads to the Lagrange Multiplier test for a general form of heteroskedasticity (see Breusch and Pagan [1]). Setting $\lambda_t(X_t,\delta) \equiv (U_{t-1}^2(\beta),\ldots,U_{t-Q}(\beta))$ gives Engle's [5] test for ARCH.

The correspondences for Theorem 2.1 are $L \equiv 1$, $C_t(\delta) \equiv 1$, $\eta_t(\theta)$

$\equiv U_t^2(\beta)$, and $\varphi_t(\Theta) \equiv \sigma^2$. Under H_0, $E[U_t^2(\beta_o)|X_t] = \sigma_o^2$ so that (ii.a) of Theorem 2.1 is satisfied. Also, $\nabla_\beta U_t^2(\beta) = -2\nabla_\beta m_t(\beta)U_t(\beta)$. Under H_0, $E[U_t(\beta_o)|X_t] = 0$ so that $E[\nabla_\beta U_t^2(\beta_o)|X_t] = 0$ and (ii.b) is holds. In this case, the relevant element of $\nabla_\Theta \varphi_t$ is simply 1. Thus, the auxiliary regression in the second step of the robust procedure simply demeans the indicators. Given $\hat{U}_t^2$, $\hat{\sigma}_T^2$, and a choice of $\hat{\lambda}_t$, the χ^2 statistic is obtained as TR^2 from the regression

$$1 \text{ on } (\hat{U}_t^2 - \hat{\sigma}_T^2)(\hat{\lambda}_t - \bar{\lambda}_T) \qquad t=1,\ldots,T \tag{3.6}$$

where $\bar{\lambda}_T \equiv T^{-1}\sum_{t=1}^{T}\hat{\lambda}_t$. This procedure is asymptotically equivalent to the corresponding more traditional forms of the tests under the additional assumption that $E[U_t^4(\beta_o)|X_t]$ is constant (note that (iii) of Lemma 2.2 is satisfied). Interestingly, the slight modification in (3.6) (which is the demeaning of the indicators $\hat{\lambda}_t$) yields an asymptotically χ^2 distributed statistic without the additional assumption of constant fourth moment for U_t^o. In the case of the White test in a linear time series model, the demeaning of the indicators yields a statistic which is asymptotically equivalent to Hsieh's [9] suggestion for a robust form of the White test, but the above statistic is significantly easier to compute. Rarely does one care to assume anything about the fourth moment of Y_t, so that the robust regression form in (3.6) seems to be a useful modification.

In the case of the ARCH test, TR^2 from the regression in (3.6) is asymptotically equivalent to TR^2 from the regression

$$1 \text{ on } (\hat{U}_t^2-\hat{\sigma}_T^2)(\hat{U}_{t-1}^2-\hat{\sigma}_T^2),\ldots,(\hat{U}_t^2-\hat{\sigma}_T^2)(\hat{U}_{t-Q}^2-\hat{\sigma}_T^2) \quad t=Q+1,\ldots T. \tag{3.7}$$

The regression based form in (3.7) is robust to departures from the conditional normality assumption, and from any other auxiliary assumptions, such as constant conditional fourth moment for U_t^o. Contrast this to the usual method of computing tests for ARCH. ■

<u>Example 3.3</u>: Theorem 2.1 can also be applied to models that jointly parameterize the conditional mean and conditional variance. The general setup is as follows. For simplicity, let Y_t be a scalar, and consider LM tests which do not assume conditional normality. The unconstrained conditional mean and variance functions are

$$\{\mu_t(x_t,\gamma),\ \omega_t(x_t,\gamma): \gamma \in \Gamma\} \tag{3.8}$$

where $\Gamma \subset \mathbb{R}^M$. It is assumed that

$$E(Y_t|X_t) = \mu_t(X_t,\gamma_o),\quad V(Y_t|X_t) = \omega_t(X_t,\gamma_o),\ \text{some } \gamma_o \in \Gamma. \tag{3.9}$$

Take the null hypothesis to be

$$H_0:\quad \gamma_o = r(\theta_o) \qquad \text{for some } \theta_o \in \Theta \subset \mathbb{R}^P \tag{3.10}$$

where $P < M$ and r is continuously differentiable on $\text{int}(\Theta)$. Let $m_t(\theta) \equiv \mu_t(r(\theta))$ and $w_t(\theta) \equiv \omega_t(r(\theta))$ be the constrained mean and variance functions. QMLE is carried out under the null hypothesis. Let $\hat{\theta}_T$ be the estimator of θ_o under H_0, and let $\hat{\gamma}_T \equiv r(\hat{\theta}_T)$ be the constrained estimator of γ_o. $\nabla_\theta \hat{m}_t$ and $\nabla_\theta \hat{w}_t$ are the 1xP gradients of m_t and w_t under H_0. Note that $\hat{\omega}_t = \hat{w}_t$ and $\hat{\mu}_t = \hat{m}_t$ by definition. The LM test of (3.10) is based on the unrestricted score of the quasi-log likelihood evaluated at $\hat{\gamma}_T$. The transpose of the score is

$$s_t(\gamma)' = \nabla_\gamma\mu_t(\gamma)'U_t(\gamma)/\omega_t(\gamma) + \nabla_\gamma\omega_t(\gamma)'[U_t^2(\gamma) - \omega_t(\gamma)]/2\omega_t^2(\gamma) \tag{3.11}$$

$$= \begin{bmatrix}\nabla_\gamma\mu_t(\gamma)\\ \nabla_\gamma\omega_t(\gamma)\end{bmatrix}' \begin{bmatrix}1/\omega_t(\gamma) & 0\\ 0 & 1/2\omega_t(\gamma)^2\end{bmatrix} \begin{bmatrix}U_t(\gamma)\\ U_t^2(\gamma) - \omega_t(\gamma)\end{bmatrix}. \tag{3.12}$$

Evaluating s_t at $r(\theta)$ gives

$$s_t(r(\theta))' \equiv \Lambda_t(\theta)'C_t(\theta)[\eta_t(\theta) - \varphi_t(\theta)] \qquad (3.13)$$

where $\Lambda_t(\theta)' \equiv [\nabla_\gamma \mu_t(r(\theta))' \;\vdots\; \nabla_\gamma \omega_t(r(\theta))']$, $C_t(\theta)$ is the diagonal matrix in the middle of (3.11) evaluated at $r(\theta)$, $\eta_t(\theta)' \equiv [U_t(r(\theta)), U_t^2(r(\theta))]$, and $\varphi_t(\theta)' \equiv [m_t(\theta), w_t(\theta)]$. The standardized score evaluated at $\hat{\gamma}_T$ is

$$T^{-1/2} \sum_{t=1}^{T} s_t(\hat{\gamma}_T). \qquad (3.14)$$

Under H_0 <u>and</u> the assumption of conditional normality, TR^2 from the regression

$$1 \quad \text{on} \quad \hat{s}_t \qquad t=1,\ldots,T \qquad (3.15)$$

is asymptotically χ_Q^2, where $Q \equiv M - P$ is the number of restrictions under H_0. Unfortunately, this procedure is invalid under nonnormality. Theorem (2.1) suggests a robust form of the test. In this case,

$$\underset{2\times P}{\nabla_\theta \hat{\varphi}_t} \equiv \begin{pmatrix} \nabla_\theta \hat{m}_t \\ \nabla_\theta \hat{w}_t \end{pmatrix}, \quad \underset{2\times M}{\hat{\Lambda}_t} \equiv \begin{pmatrix} \nabla_\gamma \hat{\mu}_t \\ \nabla_\gamma \hat{\omega}_t \end{pmatrix}$$

$$\underset{2\times 2}{\hat{C}_t} \equiv \begin{pmatrix} 1/\hat{w}_t & 0 \\ 0 & 1/2\hat{w}_t^2 \end{pmatrix}, \quad \underset{2\times 1}{\hat{\phi}_t} \equiv \begin{pmatrix} \hat{U}_t \\ \hat{U}_t^2 - \hat{w}_t \end{pmatrix}$$

where $\hat{U}_t \equiv Y_t - m_t(\hat{\theta}_T)$. The transformed quantities are

$$\nabla_\theta \tilde{\varphi}_t \equiv \begin{bmatrix} \nabla_\theta \hat{m}_t/\sqrt{\hat{w}_t} \\ \nabla_\theta \hat{w}_t/\hat{w}_t\sqrt{2} \end{bmatrix}, \quad \tilde{\Lambda}_t \equiv \begin{bmatrix} \nabla_\gamma \hat{\mu}_t/\sqrt{\hat{w}_t} \\ \nabla_\gamma \hat{\omega}_t/\hat{w}_t\sqrt{2} \end{bmatrix}$$

$$(3.16)$$

$$\tilde{\phi}_t \equiv \begin{bmatrix} \hat{U}_t/\sqrt{\hat{w}_t} \\ [\hat{U}_t^2 - \hat{w}_t]/\hat{w}_t\sqrt{2} \end{bmatrix}.$$

The robust test statistic is obtained by first running the regression

$$\tilde{\Lambda}_t \quad \text{on} \quad \nabla_\theta \tilde{\varphi}_t \qquad t=1,\ldots,T \tag{3.17}$$

and saving the matrix residuals $\{\ddot{\Lambda}_t: t=1,\ldots,T\}$. Then run the regression

$$1 \quad \text{on} \quad \tilde{\phi}_t' \ddot{\Lambda}_t \qquad t=1,\ldots,T \tag{3.18}$$

and use TR^2 as asymptotically χ^2_Q under H_0. Note that the regression in (3.18) contains perfect multicollinearity since $\ddot{\Lambda}_t \nabla_\theta r(\hat{\theta}_T) \equiv 0$, where $\nabla_\theta r(\theta)$ is the MxP gradient of r. Many regression packages nevertheless compute an R^2; for those that do not, P regressors can be omitted from (3.18).

Note that the first order condition for $\hat{\theta}_T$ is simply

$$\sum_{t=1}^{T} \nabla_\theta \varphi_t(\hat{\theta}_T)' C_t(\hat{\theta}_T)[\eta_t(\hat{\theta}_T) - \varphi_t(\hat{\theta}_T)] \equiv 0, \tag{3.19}$$

so that the robust indicator is asymptotically equivalent to the usual LM indicator. The matrix regression in (3.17) is the cost to the researcher in guarding against nonnormality. ∎

<u>Example</u> <u>3.4</u>: Suppose that Y_t is a random scalar censored below zero, and let X_{t1} be a 1xP vector of predetermined variables from X_t. A popular model for Y_t is the tobit model. The tobit model implies that

$$E(Y_t | Y_t > 0, X_t) = X_{t1}\beta_o + \sigma_o \nu(X_{t1}\beta_o/\sigma_o) \tag{3.20}$$

where $\nu(\cdot)$ is the Mills ratio, σ_o^2 is the conditional variance usually associated with the "latent" variable, and $X_{t1}\beta_o$ is conditional mean of the latent variable. From a statistical point of view, the tobit model is no more sensible than

$$\log Y_t | Y_t > 0, X_t \sim N(X_{t1}\alpha_o, \omega_o^2) \qquad (3.21)$$

((3.21) also seems reasonable for many economic applications). If (3.21) is valid, α_o and ω_o^2 can be estimated by OLS of

$$\log Y_t \quad \text{on} \quad X_{t1} \qquad t=1,\ldots,T$$

using only the positive values of Y_t. Recall that (3.21) implies

$$E(Y_t | Y_t > 0, X_t) = \exp[\omega_o^2/2 + X_{t1}\alpha_o]. \qquad (3.22)$$

Let $\hat{\lambda}_t \equiv \exp[\hat{\omega}_T^2/2 + X_{t1}\hat{\alpha}_T]$ be the fitted values in (3.22). Then, if the tobit model is true, $\hat{\lambda}_t$ should be statistically insignificant as a regressor in equation (3.20). Let $\hat{\beta}_T$, $\hat{\sigma}_T^2$ be any $\sqrt{T}$-consistent estimators of β_o, σ_o^2 under H_0. These include Heckman's [7] two-step estimators. Let

$$\hat{U}_t \equiv Y_t - X_{t1}\hat{\beta}_T - \hat{\sigma}_T\nu(X_{t1}\hat{\beta}_T/\hat{\sigma}_T).$$

A test which should have some power for testing departures from the tobit model can be based on the correlation between $\hat{U}_t$ and $\hat{\lambda}_t$. Unfortunately, the usual LM statistic is invalid for two reasons. First, $V(Y_t | Y_t > 0, X_t)$ is not constant, and second, the estimators $(\hat{\beta}_T, \hat{\sigma}_T^2)$ need not have been obtained from a nonlinear least squares problem. Nevertheless, a statistic is available from Theorem 2.1. Let

$$\varphi_t(\beta,\sigma) \equiv X_{t1}\beta + \sigma\nu(X_{t1}\beta/\sigma)$$

and let $\nabla_\theta\hat{\varphi}_t$ denote the $1 \times (P+1)$ gradient of φ_t with respect to β and σ, evaluated at $(\hat{\beta}_T, \hat{\sigma}_T)$. Then the following procedure is asymptotically valid:

(i) Run the OLS regression

$$\hat{\lambda}_t \quad \text{on} \quad \nabla_\theta\hat{\varphi}_t \qquad t=1,\ldots,T$$

and save the residuals $\ddot{\lambda}_t$.

(ii) Run the regression

$$1 \quad \text{on} \quad \hat{U}_t\ddot{\lambda}_t \qquad t=1,\ldots,T$$

and use TR^2 as asymptotically χ_1^2 under H_0.

Note that weighted least squares could also be used, where the weight corresponds to the inverse of $V(Y_t|Y_t>0,X_t)$ under the tobit model. If $\hat{c}_t$ is an estimate of this variance, replace $\hat{\lambda}_t$ and $\nabla_\theta\hat{\varphi}_t$ by $\hat{\lambda}_t/\sqrt{\hat{c}_t}$ and $\nabla_\theta\hat{\varphi}_t/\sqrt{\hat{c}_t}$, respectively in (i), and replace $\hat{U}_t$ by $\hat{U}_t/\sqrt{\hat{c}_t}$ in (ii). Although it intuitively makes sense to use the weighted version, it is not possible to say one approach is better than the other without more information about the origins of $\hat{\beta}_T$ and $\hat{\sigma}_T^2$.

One can of course change the roles of the models, and test for a significant covariance between $\hat{\lambda}_t \equiv X_{t1}\hat{\beta}_T + \hat{\sigma}_T\nu(X_{t1}\hat{\beta}_T/\hat{\sigma}_T)$ and the residuals based on (3.22). In this case, the purging regression takes the form

$$\hat{\lambda}_t \quad \text{on} \quad \exp[\hat{\omega}_T^2/2 + X_{t1}\hat{\alpha}_T]X_{t1} \qquad t=1,\ldots,T. \tag{3.23}$$

Note that a similar test could be based on competing specifications for $E(Y_t|X_t)$; that is, the zero as well as positive observations for Y_t can be used. This would require specifying $P(Y_t>0|X_t)$ in the competing model (3.21) such as in Cragg [2].

Finally, many other indicators could be included in $\hat{\lambda}_t$, such as the gradient of the competing conditional mean function: $\hat{\lambda}_t \equiv \exp[\hat{\omega}_T^2/2 + X_{t1}\hat{\alpha}_T]X_{t1}$ in the case of (3.21). I do not know the power properties of these tests. They are included here primarily to illustrate the scope of Theorem 2.1. ■

4. Conclusions

This paper has developed a general class of specification tests for dynamic multivariate models which impose under H_0 only the hypotheses being tested (e.g. correctness of the conditional mean and/or correctness of the conditional variance). It is hoped that the computational simplicity of the methods proposed here removes some of the barriers to using robust test statistics in practice.

The possibility of generating simple test statistics when $T^{1/2}(\hat{\theta}_T - \theta_o)$ has a complicated limiting distribution should be useful in several situations. The tobit example in Section 3 is only one case where the conditional mean parameters are estimated using a method other than the efficient WNLLS procedure. Another example is choosing between log-linear and linear-linear specifications. In this case, both models can be estimated by OLS, and then transformed in the manner of the tobit example to obtain estimates of $E(Y_t|X_t)$ for the separate models.

Theorem 2.1 can be extended to certain unit root time series models. The initial purging of $\hat{C}_t^{1/2}\nabla_\theta\hat{\varphi}_t$ from $\hat{C}_t^{1/2}\hat{\Lambda}_t$ in some cases results in indicators that are effectively stationary. This is the case for the LM test in linear time series models where the regressors excluded under the null hypothesis are individually cointegrated with the regressors included under the null. Statistics derived from Theorem 2.1 have the advantage over the usual Wald or LM tests of being robust to conditional heteroskedasticity under H_0. Extending Theorem 2.1 to general nonstationary time series models is left for future research.

Footnotes

1. If $E[U_t(\theta_o)^2 \nabla_\theta m_t(\theta_o)'\lambda_t(\delta_T^o)] = 0$ and $E[\nabla_\theta m_t(\theta_o)'\lambda_t(\delta_T^o)] = 0$ then the regression form in (2.10) is valid in the presence of heteroskedasticity. These orthogonality conditions occur only in limited cases. One example is testing for serial correlation in a static regression model ($E(Y_t|X_t)$ depends only on Z_t under H_0) with static heteroskedasticity ($V(Y_t|X_t)$ depends only on Z_t under H_0). If $E[\nabla_\theta m_t(\theta_o)'\lambda_t(\delta_T^o)] = 0$ under H_0 then a simple test which is robust in the presence of arbitrary heteroskedasticity is TR^2 from the regression

$$1 \text{ on } \hat{U}_t\hat{\lambda}_t \qquad t=1,\ldots,T,$$

that is, $\hat{U}_t\nabla_\theta\hat{m}_t$ can be omitted from the auxiliary regression.

2. Hal White has suggested an interesting extension to Theorem 2.1. First, there is no need to split $\phi_t(Y_t,X_t,\theta)$ into $\eta_t(Y_t,X_t,\theta)$ and $\varphi_t(X_t,\theta)$. Then, instead of imposing (ii.b), use $\Phi_t(X_t,\hat{\theta}_T)$ in the purging regression, where $\Phi_t(X_t,\theta) \equiv E_\theta[\nabla_\theta\phi_t(Y_t,X_t,\theta)|X_t]$. Note that it is now important to index the expectation operator by θ. This expectation is the common expectation of the equivalence class $\mathcal{P}_\theta$ of probability measures defined as follows: $P \in \mathcal{P}(\theta)$ if and only if

$$E_P[\phi_t(Y_t,X_t,\theta)|X_t] = 0$$

and

$$E_P[\nabla_\theta\phi_t(Y_t,X_t,\theta)|X_t] = \Phi_t(X_t,\theta) \qquad t=1,2,\ldots.$$

The need to compute $\Phi_t(X_t,\theta)$ generally imposes additional restrictions under the null hypothesis. However, this more general setup would allow robust tests in certain situations not covered by Theorem 2.1, such as tests for dynamic linear models estimated by two stage least squares.

References

1. Breusch, T.S. and A.R. Pagan. A simple test for heteroskedasticity and random coefficient variation. *Econometrica* 47 (1979): 1287-1294.

2. Cragg, J.G. Some statistical models for limited dependent variables with applications to the demand for durable goods. *Econometrica* 39 (1971): 829-844.

3. Davidson, R. and J.G. MacKinnon. Several tests of model specification in the presence of alternative hypotheses. *Econometrica* 49 (1981): 781-793.

4. Domowitz I. and H. White. Misspecified models with dependent observations. *Journal of Econometrics* 20 (1982): 35-58.

5. Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of United Kingdom inflation. *Econometrica* 50 (1982): 987-1008.

6. Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* 50 (1982): 1029-1054.

7. Heckman, J.J. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5 (1976): 475-492.

8. Hausman, J.A. Specification tests in econometrics. *Econometrica* 46 (1978): 1251-1271.

9. Hsieh, D.A. A heteroskedasticity-consistent covariance matrix estimator for time series regressions. *Journal of Econometrics* 22 (1983): 281-290.

10. Newey, W.K. Maximum likelihood specification testing and conditional moment tests," *Econometrica* 53 (1985): 1047-1070.

11. Tauchen, G. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30 (1985): 415-443.

12. White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (1980a): 817-838.

13. White, H. Nonlinear regression on cross section data. *Econometrica* 48 (1980b): 721-746.

14. White, H. Maximum likelihood estimation of misspecified models. *Econometrica* 50 (1982): 1-26.

15. White, H. *Asymptotic Theory for Econometricians*. New York: Acamedic Press, 1984.

16. White, H. Specification testing in dynamic models. Invited Paper at the Fifth World Congress of the Econometric Society, Cambridge, Massachusetts, August 1985.

17. White, H. and I. Domowitz. Nonlinear regression with dependent observations. *Econometrica* 52 (1984): 143-162.

18. Wooldridge, J.M. *Asymptotic properties of econometric estimators*. UCSD Ph.D. dissertation, 1986.

19. Wooldridge, J.M. Specification testing and quasi-maximum likelihood estimation. Mimeo, MIT, 1987.

Mathematical Appendix

For convenience, I include a lemma which is used repeatedly in the proof of Theorem 2.1.

Lemma A.1: Assume that the sequence of random functions $\{Q_T(W_T,\theta): \theta \in \Theta,\ T=1,2,\ldots\}$, where $Q_T(W_T,\cdot)$ is continuous on Θ and Θ is a compact subset of $\mathbb{R}^P$, and the sequence of nonrandom functions $\{\bar{Q}_T(\theta): \theta \in \Theta,\ T=1,2,\ldots\}$, satisfy the following conditions:

(i) $\sup_{\theta\in\Theta} |Q_T(W_T,\theta) - \bar{Q}_T(\theta)]| \overset{p}{\to} 0$;

(ii) $\{\bar{Q}_T(\theta): \theta \in \Theta,\ T=1,2,\ldots\}$ is continuous on Θ uniformly in T.

Let $\ddot{\theta}_T$ be a sequence of random vectors such that $\ddot{\theta}_T - \theta_T^o \overset{p}{\to} 0$ where $\{\theta_T^o\} \subset \Theta$. Then

$$Q_T(W_T,\ddot{\theta}_T) - \bar{Q}_T(\theta_T^o) \overset{p}{\to} 0.$$

Proof: see Wooldridge [18, Lemma A.1, p.229]. ■

A definition simplifies the statement of the conditions.

Definition A.1: A sequence of random functions $\{q_t(Y_t,X_t,\theta): \theta \in \Theta,\ t=1,2,\ldots\}$, where $q_t(Y_t,X_t,\cdot)$ is continuous on Θ and Θ is a compact subset of $\mathbb{R}^P$, is said to satisfy the Uniform Weak Law of Large Numbers (UWLLN) and Uniform Continuity (UC) conditions provided that

$$\text{(i)}\ \sup_{\theta\in\Theta} |T^{-1}\sum_{t=1}^{T} q_t(Y_t,X_t,\theta) - E[q_t(Y_t,X_t,\theta)]| \overset{p}{\to} 0$$

and

$$\text{(ii)}\ \{T^{-1}\sum_{t=1}^{T} E[q_t(Y_t,X_t,\theta)]: \theta \in \Theta,\ T=1,2,\ldots\} \text{ is } O(1) \text{ and}$$

continuous on Θ uniformly in T. ■

In the statement of the conditions, the dependence of functions on the variables Y_t and X_t is frequently suppressed for notational convenience. If $a(\Theta)$ is a $1\times L$ function of the $P\times 1$ vector Θ then, by convention, $\nabla_\Theta a(\Theta)$ is the $L\times P$ matrix $\nabla_\Theta[a(\Theta)']$. If $A(\Theta)$ is a $Q\times L$ matrix then the matrix $\nabla_\Theta A(\Theta)$ is the $LQ\times P$ matrix defined as

$$\nabla_\Theta A(\Theta)' \equiv [\nabla_\Theta A_1(\Theta)' \mid \dots \mid \nabla_\Theta A_Q(\Theta)']$$

where $A_j(\Theta)$ is the jth row of $A(\Theta)$ and $\nabla_\Theta A_j(\Theta)$ is the $L\times P$ gradient of $A_j(\Theta)$ as defined as above. For simplicity, for any $L\times 1$ vector function φ, define the second derivative of φ to be the $LP\times P$ matrix

$$\nabla^2_\Theta \varphi(\Theta) \equiv \nabla_\Theta[\nabla_\Theta \varphi(\Theta)'].$$

<u>Conditions</u> <u>A.1</u>:

(i) $\Theta \subset \mathbb{R}^P$ and $\Pi \subset \mathbb{R}^N$ are compact and have nonempty interiors;

(ii) $\Theta_o \in \text{int}(\Theta)$, $\{\pi_T^o\colon T=1,2,\dots\} \subset \text{int}(\Pi)$ uniformly in T;

(iii) (a) $\{\eta_t(y_t,x_t,\Theta)\colon \Theta \in \Theta\}$ is a sequence of $L\times 1$ functions such that $\eta_t(\cdot,\Theta)$ is Borel measurable for each $\Theta \in \Theta$ and $\eta_t(y_t,x_t,\cdot)$ is continuously differentiable on the interior of Θ for all y_t,x_t, $t=1,2,\dots$;

(b) $\{\varphi_t(x_t,\Theta)\colon \Theta \in \Theta\}$ is a sequence of $L\times 1$ functions such that $\varphi_t(\cdot,\Theta)$ is Borel measurable for each $\Theta \in \Theta$ and $\varphi_t(x_t,\cdot)$ is twice continuously differentiable on the interior of Θ for all x_t, $t=1,2,\dots$;

(c) $\{C_t(x_t,\delta)\colon \delta \in \Delta\}$ is a sequence of $L\times L$ matrices satisfying the measurability requirements, $C_t(x_t,\delta)$ is symmetric and positive semi-definite for all x_t and δ, and $C_t(x_t,\cdot)$ is differentiable on $\text{int}(\Delta)$ for all x_t, $t=1,2,\dots$;

(d) $\{\Lambda_t(x_t,\delta): \delta \in \Delta\}$ is a sequence of LxQ matrices satisfying the measurability requirements, and $\Lambda_t(x_t,\cdot)$ is differentiable on int(Δ) for all x_t, t=1,2,...;

(iv) (a) $T^{1/2}(\hat{\theta}_T - \theta_o) = O_p(1)$;

(b) $T^{1/2}(\hat{\pi}_T - \pi_T^o) = O_p(1)$;

(v) (a) $\{\nabla_\theta\varphi_t(\theta)'C_t(\delta)\nabla_\theta\varphi_t(\theta)\}$ and $\{\nabla_\theta\varphi_t(\theta)'C_t(\delta)\Lambda_t(\delta)$ satistfy the UWLLN and UC conditions;

(b) $\{T^{-1}\sum_{t=1}^{T} E[\nabla_\theta\varphi_t^{o\prime}C_t^o\nabla_\theta\varphi_t^o]\}$ is uniformly positive definite;

(vi) (a) $\{\nabla_\theta\varphi_t(\theta)'C_t(\delta)\nabla_\theta\eta_t(\theta)\}$, $\{[I_P \otimes \phi_t(\theta)'C_t(\delta)]\nabla_\theta^2\varphi_t(\theta)\}$, and $\{\nabla_\theta\varphi_t(\theta)'[I_L \otimes \phi_t(\theta)']\nabla_\delta C_t(\delta)\}$ satisfy the UWLLN and UC conditions;

(b) $T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\varphi_t^{o\prime}C_t^o\phi_t^o = O_p(1)$;

(vii) (a) $\{\Lambda_t(\delta)'C_t(\delta)\nabla_\theta\eta_t(\theta)\}$, $\{\Lambda_t(\delta)'C_t(\delta)\nabla_\theta\varphi_t(\theta)\}$,

$\{[I_Q \otimes \phi_t(\theta)'C_t(\delta)]\nabla_\delta\Lambda_t(\delta)'\}$,

$\{[I_P \otimes \phi_t(\theta)'C_t(\delta)]\nabla_\theta^2\varphi_t(\theta)\}$,

$\{\Lambda_t(\delta)'[I_L \otimes \phi_t(\theta)']\nabla_\delta C_t(\delta)\}$, and

$\{\nabla_\theta\varphi_t(\delta)'[I_L \otimes \phi_t(\theta)']\nabla_\delta C_t(\delta)\}$

satisfy the UWLLN and UC requirements;

(viii) (a) $\Xi_T^o \equiv T^{-1}\sum_{t=1}^{T} E[(\Lambda_t^o-\nabla_\theta\varphi_t^oB_T^o)'C_t^o\phi_t^o\phi_t^{o\prime}C_t^o(\Lambda_t^o-\nabla_\theta\varphi_t^oB_T^o)]$ is uniformly p.d.;

(b) $\Xi_T^{o-1/2}T^{-1/2}\sum_{t=1}^{T}(\Lambda_t^o - \nabla_\theta\varphi_t^oB_T^o)'C_t^o\phi_t^o \xrightarrow{d} N(0,I_Q)$;

(c) $\{\Lambda_t(\delta)'C_t(\delta)\phi_t(\theta)\phi_t(\theta)'C_t(\delta)\Lambda_t(\delta)\}$,

$\{\Lambda_t(\delta)'C_t(\delta)\phi_t(\theta)\phi_t(\theta)'C_t(\delta)\nabla_\theta\varphi_t(\theta)\}$, and

$\{\nabla_\theta\varphi_t(\theta)'C_t(\delta)\phi_t(\theta)\phi_t(\theta)'C_t(\delta)\nabla_\theta\varphi_t(\theta)\}$

satisfy the UWLLN and UC conditions. ■

Proof of Theorem 2.1: First, note that assumptions (i)-(vi) ensure existence of B_T^o and imply that $\hat{B}_T - B_T^o = o_p(1)$ by Lemma A.1. Therefore,

$$\ddot{\xi}_T = T^{-1/2}\sum_{t=1}^{T}[\hat{\Lambda}_t - \nabla_\theta\hat{\varphi}_t B_T^o]'\hat{C}_t\hat{\phi}_t \qquad (a.1)$$
$$- (\hat{B}_T - B_T^o)'T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\hat{\varphi}_t'\hat{C}_t\hat{\phi}_t.$$

Consider the term post-multiplying $(\hat{B}_T - B_T^o)'$. A standard mean value expansion about δ_T^o, assumption (vi.a), and Lemma A.1 yield

$$T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\hat{\varphi}_t'\hat{C}_t\hat{\phi}_t = T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\varphi_t^{o\prime}C_t^o\phi_t^o \qquad (a.2)$$
$$+ T^{-1}\sum_{t=1}^{T}\{\nabla_\theta\varphi_t^{o\prime}C_t^o\nabla_\theta\phi_t^o + [I_P \otimes \phi_t^{o\prime}C_t^o]\nabla_\theta^2\varphi_t^o\}\, T^{1/2}(\hat{\theta}_T - \theta_o)$$
$$+ T^{-1}\sum_{t=1}^{T}\{[I_L \otimes \phi_t^{o\prime}]\nabla_\delta C_t^o\}\, T^{1/2}(\hat{\delta}_T - \delta_o) + o_p(1).$$

The first term on the right hand side of (a.2) is $O_p(1)$ by (vi.b). By (vi.a) and (iv.a,b), the terms in lines two and three of (a.2) are also $O_p(1)$. Therefore,

$$T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\hat{\varphi}_t'\hat{C}_t\hat{\phi}_t = O_p(1). \qquad (a.3)$$

Along with $\hat{B}_T - B_T^o = o_p(1)$, this establishes that under H_0,

$$\ddot{\xi}_T = T^{-1/2}\sum_{t=1}^{T}[\hat{\Lambda}_t - \nabla_\theta\hat{\varphi}_t B_T^o]'\hat{C}_t\hat{\phi}_t + o_p(1). \qquad (a.4)$$

A mean value expansion, assumption (vii), and Lemma A.1 yield

$$\ddot{\xi}_T = T^{-1/2}\sum_{t=1}^{T}[\Lambda_t^o - \nabla_\theta\varphi_t^o B_T^o]'C_t^o\phi_t^o \qquad (a.5)$$
$$+ T^{-1}\sum_{t=1}^{T}\{[\Lambda_t^o-\nabla_\theta\varphi_t^o B_T^o]'C_t^o\nabla_\theta\phi_t^o - B_T^{o\prime}[I_P \otimes \phi_t^{o\prime}C_t^o]\nabla_\theta^2\varphi_t^o\}\, T^{1/2}(\hat{\theta}_T - \theta_o)$$
$$+ T^{-1}\sum_{t=1}^{T}\{[I_Q \otimes \phi_t^{o\prime}C_t^o]\nabla_\delta\Lambda_t^{o\prime} + [\Lambda_t^o-\nabla_\theta\varphi_t^o B_T^o]'[I_L \otimes \phi_t^{o\prime}]\nabla_\delta C_t^o\}$$
$$\cdot T^{1/2}(\hat{\delta}_T - \delta_T^o)$$

$+ \quad o_p(1).$

Consider the second line of (a.5). It must be shown that the average appearing there is $o_p(1)$ under H_0. First, note that

$$T^{-1}\sum_{t=1}^{T} [\Lambda_t^o - \nabla_\theta\varphi_t^o B_T^o]' C_t^o \nabla_\theta \phi_t^o = T^{-1}\sum_{t=1}^{T} [\Lambda_t^o - \nabla_\theta\varphi_t^o B_T^o]' C_t^o \nabla_\theta \eta_t^o \qquad (a.6)$$

$$- T^{-1}\sum_{t=1}^{T} [\Lambda_t^o - \nabla_\theta\varphi_t^o B_T^o]' C_t^o \nabla_\theta \varphi_t^o.$$

By (ii.b) of the text, $E[\nabla_\theta \eta_t^o | X_t] = 0$ under H_0. Note that Λ_t^o, $\nabla_\theta \varphi_t^o$, and C_t^o depend only on X_t. Also, B_T^o is defined such that

$$T^{-1}\sum_{t=1}^{T} E[(\Lambda_t^o - \nabla_\theta\varphi_t^o B_T^o)' C_t^o \nabla_\theta \varphi_t^o] = 0. \qquad (a.7)$$

The regularity conditions imposed imply that each of the averages on the right hand side of (a.6) satisfy the WLLN. Therefore

$$T^{-1}\sum_{t=1}^{T} [\Lambda_t^o - \nabla_\theta\varphi_t^o B_T^o]' C_t^o \nabla_\theta \phi_t^o = o_p(1). \qquad (a.8)$$

Because $E[\phi_t^o | X_t] = 0$ under H_0, it is even easier to show that the remaining sample averages in (a.5) are $o_p(1)$. Combined with $T^{1/2}(\hat{\delta}_T - \delta_T^o) = O_p(1)$ this establishes the first conclusion of the theorem:

$$\ddot{\xi}_T = T^{-1/2}\sum_{t=1}^{T} [\Lambda_t^o - \nabla_\theta\varphi_t^o B_T^o]' C_t^o \phi_t^o + o_p(1). \qquad (a.9)$$

Given (viii.a), the asymptotic covariance matrix of $\ddot{\xi}_T$ is uniformly positive definite. Moreover, $\Xi_T^{o-1/2}\ddot{\xi}_T \xrightarrow{d} N(0, I_Q)$ under H_0 by (viii.b). Condition (viii.c) ensures that

$$\ddot{\Xi}_T \equiv T^{-1}\sum_{t=1}^{T} [(\hat{\Lambda}_t - \nabla_\theta\hat{\varphi}_t \hat{B}_T)' \hat{C}_t \hat{\phi}_t \hat{\phi}_t' \hat{C}_t (\hat{\Lambda}_t - \nabla_\theta\hat{\varphi}_t \hat{B}_T)] \qquad (a.10)$$

is a consistent estimator of Ξ_T^o. It is easy to see that

$$\ddot{\xi}_T' \ddot{\Xi}_T^{-1} \ddot{\xi}_T = TR^2, \qquad (a.11)$$

where R^2 is the uncentered r-squared from the regression

$$1 \quad \text{on} \quad \tilde{\phi}_t' \ddot{\Lambda}_t \qquad t=1,\ldots,T, \qquad \text{(a.12)}$$

and $\tilde{\phi}_t$ and $\ddot{\Lambda}_t$ are as defined in the text. ■