working paper
department
of economics

A UNIFIED APPROACH TO ROBUST, REGRESSION-BASED SPECIFICATION TESTS

Jeffrey M. Wooldridge

No. 480

Revised
November 1988

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

A UNIFIED APPROACH TO ROBUST, REGRESSION-BASED SPECIFICATION TESTS

Jeffrey M. Wooldridge

No. 480

Revised
November 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
Revised: November 1988

*I am grateful to Adrian Pagan, Peter Phillips, Jim Poterba, Danny Quah, Tom Stoker, Hal White, and an anonymous referee for helpful comments and suggestions on a previous version.

ABSTRACT: This paper develops a general approach to robust, regression-based specification tests for (possibly) dynamic econometric models. The key feature of the proposed tests is that, in addition to 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. For the leading cases of conditional mean and/or conditional variance tests, the proposed statistics are robust to departures from distributional assumptions that are not being tested. Moreover, the statistics can be computed using any $\sqrt{T}$-consistent estimator, resulting in significant simplifications in some otherwise difficult contexts. Among the examples covered are conditional mean tests for models estimated by weighted nonlinear least squares under misspecification of the conditional variance, tests of jointly parameterized conditional means and variances estimated by quasi-maximum likelihood under nonnormality, and some new, computationally simple specification tests for the tobit model.

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. The most popular specification tests are those that can be computed via ordinary least squares regressions. Examples are the Lagrange Multiplier (LM) test for nested hypotheses, versions of Hausman's [13] specification tests, White's [24] information matrix (IM) test, and regression-based versions of various nonnested hypotheses tests. In fact, Newey [17], Tauchen [21], and White [26] have shown that all of these tests are asymptotically equivalent to a particular conditional moment (CM) test. In a maximum likelihood setting with independent observations, Newey [17] and Tauchen [21] have devised outer product-type auxiliary regressions for computing CM tests. White [26] has extended these methods to a general dynamic setting.

The simplicity of most popular regression-based procedures currently employed, including the Newey-Tauchen-White (NTW) procedure, 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 or unconditional heteroskedasticity. Except in special cases the NTW outer product statistic is also invalid. Other examples include the various tests for heteroskedasticity: currently used regression forms require constancy of the conditional fourth moment of the regression errors under the null hypothesis. In addition, the Lagrange Multiplier and other CM tests for jointly parameterized conditional means and variances are inappropriate under various departures from normality.

The above situations are all characterized by the same feature: validity of the tests requires imposition of more than just the hypotheses of interest under H_0. In addition, traditional econometric 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 [22,23,24,26], Domowitz and White [7], Hansen [11], and Newey [17], there now exist general methods of computing robust statistics. In the context of linear regression models, Pagan and Hall [19] discuss how to compute conditional mean tests that are robust to heteroskedasticity. Their discussion centers around the use of the White [22] heteroskedasticity-consistent standard errors. For one degree of freedom tests the Pagan and Hall suggestion leads to easily computable tests, and it is certainly an alternative to the current approach for regression models. But computation of the statistics for tests with more than one degree of freedom requires explicit inversion of the White covariance matrix estimator; this matrix must then be used in a quadratic form to obtain the heteroskedasticity-robust Wald statistic. The tests proposed here are very much in the spirit of the LM approach: computation requires estimation of the model only under the null, so that any particular model can be subjected to a battery of robust specification tests without ever reestimating the model. More importantly, the tests can be computed using any standard regression package.

Although there are some fairly general formulas available for robust LM statistics (e.g. Engle [9], White [24,25]), formulas for general nonlinear restrictions involve an analytical expression for the derivative of the implicit constraint function and a generalized inverse. In specific instances computationally simple robust LM statistics are available. A notable example is the paper by Davidson and MacKinnon [6], which develops a regression-based heteroskedasticity-robust LM test in a nonlinear regression model with independent errors and unconditional heteroskedasticity.

It is a safe bet that the substantial analytical and computational work required to obtain robust statistics is a primary reason that they are used infrequently in applied work. Evidence of this statement is the growing use of the White [22] heteroskedasticity-robust t-statistics, which are now computed by many econometrics packages. Only occasionally does one see 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 is 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 via least squares regressions which I believe is easily accessible to applied econometricians. The general method suggested here can be viewed as an extension of the Davidson and MacKinnon [6] approach. In fact, in the context of nonlinear regression models, their procedure is shown to be valid for quite general dynamic models with conditional as well as unconditional heteroskedasticity. In the the same context the approach here can be viewed as the Lagrange Multiplier version of the robust Wald strategy suggested by Pagan and Hall [19]. This paper also extends Wooldridge's [30] robust, regression based conditional mean and conditional variance tests in the context of quasi-maximum likelihood estimation in multivariante linear exponential families. The current framework is more general because it applies anytime a generalized residual function (defined in section 2) is the basis for the test.

For the leading cases of conditional first and second moments, the regression-based tests proposed maintain only the hypotheses of interest under the null, and they are applicable to specification testing of dynamic multivariate models of first and second moments without imposing further assumptions on the conditional distribution (except regularity conditions). Moreover, in classical situations, these tests are asymptotically equivalent under the null and local alternatives to their traditional counterparts. Robustness is obtained without sacrificing asymptotic efficiency.

For some specification tests the current approach does impose auxiliary assumptions under the null hypothesis. This is the price one pays for the regression-based nature of the tests. Still, in most cases encountered so far, the current framework imposes fewer auxiliary assumptions under the null than popular nonrobust tests. This does not mean that robust tests are not available in such circumstances, but only that regression-based forms of these tests are not known. The goal here is to provide a unified approach to robust, regression-based specification tests, and not to robust tests in general. Nevertheless, the coverage is fairly broad. A general treatment of robust tests is contained in White [26].

A second aspect of the proposed statistics is that they may be computed using any $\sqrt{T}$-consistent estimator. The asymptotic distribution of the test statistic under the null and local alternatives is invariant with respect to the asymptotic distribution of the estimators used in computation; this can be viewed as another kind of robustness. Consequently, the methodology leads to some interesting new tests in cases where the computational burden based on previous approaches can be prohibitive. This is true whether or not robustness to violation of auxiliary assumptions is an issue; in fact, the procedure can be profitably applied to situations which assume correct specification of the entire conditional distribution provided that the test statistic can be put into the form considered in section 2. In such cases the proposed tests have properties similar to Neyman's [18] $C(\alpha)$ tests, but they are applicable even whether or not the score of the log-likelihood is the basis for test statistic. When restricted to LM tests, the new statistics offer generalized residual alternatives to outer product-type $C(\alpha)$ statistics, provided of course that the score of the log-likelihood can be put into the appropriate generalized residual form.

Section 2 of the paper 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 $\{(y_t, z_t): t=1,2,\ldots\}$ be a sequence of observable random vectors with y_t 1xJ, z_t 1xK. 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 $x_t \equiv (z_t, y_{t-1}, z_{t-1}, \ldots, y_1, z_1)$ denote the predetermined variables. Note that current z_t can be excluded from x_t or, if there are no "exogenous" variables, one may take $x_t \equiv (y_{t-1}, y_{t-2}, \ldots, y_1)$. For cross section applications set $x_t \equiv z_t$ and assume that the observations are independently distributed.

The conditional distribution of y_t given 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 $\{(y_{t-1}, z_{t-1}), \ldots, (y_1, z_1)\}$ or $\{y_{t-1}, y_{t-2}, \ldots, y_1\}$, 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 irrelevant.

For motivational purposes and to illustrate the notation, it is useful to introduce a couple of examples. The first example concerns specification testing of a conditional mean. 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

$$\{m_t(x_t, \alpha): \alpha \in A,\ t=1,2,\ldots\}, \qquad (2.1)$$

where $A \subset \mathbb{R}^p$, and the null hypothesis is

$$H_0\text{: } E(y_t|x_t) = m_t(x_t,\alpha_o), \text{ some } \alpha_o \in A,\ t=1,2,\ldots. \tag{2.2}$$

If $\hat{\alpha}_T$ is a $\sqrt{T}$-consistent estimator of α_o under H_0 then the residuals are defined as $u_t(y_t,x_t,\hat{\alpha}_T) \equiv y_t - m_t(x_t,\hat{\alpha}_T)$. A test of H_0 can be based on the sample covariance

$$T^{-1}\sum_{t=1}^{T} \lambda_t(x_t,\hat{\alpha}_T,\hat{\pi}_T)' u_t(y_t,x_t,\hat{\alpha}_T) \tag{2.3}$$

$$\equiv T^{-1}\sum_{t=1}^{T} \hat{\lambda}_t'\hat{u}_t \tag{2.4}$$

where $\lambda_t(x_t,\alpha,\pi)$ is a 1xQ vector function of "misspecification indicators" that can depend on $\hat{\alpha}_T$ and a nuisance parameter estimator $\hat{\pi}_T$. The standard LM approach leads to a test based on the (uncentered) r-squared from the regression

$$\hat{u}_t \text{ on } \nabla_\alpha\hat{m}_t,\ \hat{\lambda}_t \qquad t=1,\ldots,T. \tag{2.5}$$

If $\hat{\alpha}_T$ is asymptotically equivalent to the NLS estimator then under H_0 <u>and</u> conditional homoskedasticity, TR_u^2 is asymptotically χ_Q^2. Thus, the LM approach effectively takes the null hypothesis to be

$$H_0'\text{: } H_0 \text{ holds and } V(y_t|x_t) = \sigma_o^2 \text{ for some } \sigma_o^2 > 0,\ t=1,2,\ldots \tag{2.6}$$

but it is inconsistent for the alternative

$$H_1'\text{: } H_0 \text{ holds but } H_0' \text{ does not.}$$

It also essentially requires that $\hat{\alpha}_T$ be the NLS estimator.

The Newey-Tauchen-White regression for the same problem is

$$1 \text{ on } \hat{u}_t\nabla_\alpha\hat{m}_t,\ \hat{u}_t\hat{\lambda}_t \qquad t=1,\ldots,T. \tag{2.7}$$

In general, H_0' is also required for TR_u^2 from this regression to be asmptotically χ_Q^2, although there are some cases, such as testing for serial correlation in a static regression model with static conditional heteroskedasticity (i.e. $V(y_t|x_t) = V(y_t|z_t)$), where the NTW regression is robust. The validity of the NTW procedure also generally relies on $\hat{\alpha}_T$ being the NLS estimator.

As pointed out by Pagan and Hall [19], a robust test is available from the regression (2.5). The White [22] heteroskedasticity-robust covariance matrix estimator can be used to compute a robust Wald statistic for the hypothesis that $\hat{\lambda}_t$ can be excluded from the regression (2.5). When $\hat{\lambda}_t$ is a scalar this is simple because a robust test statistic is simply the robust t-statistic on $\hat{\lambda}_t$. When $\hat{\lambda}_t$ is a vector computation of the robust Wald statistic is somewhat more complicated since it involves inversion of the White covariance matrix estimator as well as explicit construction of the appropriate quadratic form. In addition, the Wald procedure is valid essentially only when $\hat{\alpha}_T$ is the NLS estimator.

The regression-based heteroskedasticity-robust form of the test, which in addition is valid for any $\sqrt{T}$-consistent estimator, is a special case of Example 3.1 discussed in section 3.

As a second example, consider testing for heteroskedasticity. The null hypothesis is taken to be

$$H_0\colon E(y_t|x_t) = m_t(x_t,\alpha_o) \text{ and } V(y_t|x_t) = \sigma_o^2,\ \alpha_o \in A,\ \sigma_o^2 > 0,\ t=1,2,\ldots.$$

Again let $u_t(y_t,x_t,\alpha)$ be the residual function, and let $\lambda(x_t,\theta,\pi)$ be a $1\times Q$ vector of heteroskedasticity indicators, where $\theta \equiv (\alpha',\sigma^2)'$. A general class of tests is based on

$$T^{-1}\sum_{t=1}^{T} \lambda_t(x_t,\hat{\theta}_T,\hat{\pi}_T)'[u_t^2(y_t,x_t,\hat{\alpha}_T) - \hat{\sigma}_T^2)]$$

$$\equiv T^{-1} \sum_{t=1}^{T} \hat{\lambda}_t'(\hat{u}_t^2 - \hat{\sigma}_T^2)$$

where $\hat{\sigma}_T^2 \equiv T^{-1} \sum_{t=1}^{T} \hat{u}_t^2$. A standard LM-type statistic is obtained from the centered r-squared from the regression

$$\hat{u}_t^2 \text{ on } 1,\ \hat{\lambda}_t,\quad t=1,\ldots,T. \tag{2.8}$$

TR_c^2 is asymptotically χ_Q^2 under

H_0': H_0 holds and, in addition, $E[(u_t^o)^4|x_t] = \kappa_o^2 > 0$, $t=1,2,\ldots$

where $u_t^o \equiv y_t - m_t(x_t,\alpha_o)$. Regression (2.8) yields the "studentized" version of the Breusch-Pagan [2] test as derived by Koenker [16]. The studentized form of the test is robust to certain departures from normality, and it is now widely used in the literature (see, e.g,. Engle [8], Pagan and Hall [19], and Pagan, Trivedi, and Hall [20]). Unfortunately, this form of the test is not completely robust in the sense defined in this paper. The constancy of $E[(u_t^o)^4|x_t]$ is an auxiliary assumption imposed under H_0 that is required for (2.8) to lead to a valid test. Normality of u_t^o conditional on x_t rules out heterokurtosis under H_0, but it is easy to construct examples to illustrate that the auxiliary assumption of homokurtosis is binding. If the regression errors u_t^o have a conditional t-distribution with constant variance but degrees of freedom that otherwise depend on x_t then H_0 holds but H_0' does not. Hsieh [14] and Pagan and Hall [19] have noted that just as with conditional mean tests, the White [22] covariance matrix can be used to compute heteroskedasticity tests that are robust to heterokurtosis. However, except when $\hat{\lambda}_t$ is a scalar, computation of the statistic requires several matrix operations. Pagan, Trivedi, and Hall [20] report the White [22] t-statistic in a model for the variance of inflation when $\hat{\lambda}_t$ is a scalar. An alternative robust form of the test is provided in Example 3.2 in section 3. It is almost as easy to compute as the nonrobust form even when $\hat{\lambda}_t$ is a vector, but it allows for heterokurtosis under the null.

There are other examples where the goal is to test hypotheses about certain aspects of a conditional 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 are usually 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. To be attractive these tests must be easy to compute under reasonably broad circumstances. The remainder of this section develops a general approach to constructing robust, regression-based tests.

Many specification tests, including those for conditional means and variances, have asymptotically equivalent versions that can be derived as follows. Let $\phi_t(y_t,x_t,\theta)$ be an Lx1 random function defined on a parameter set $\Theta \subset R^P$. The null hypothesis of interest is expressed as

$$H_0: E[\phi_t(y_t,x_t,\theta_o)|x_t] = 0, \text{ for some } \theta_o \in \Theta, \quad t=1,2,\ldots.. \tag{2.9}$$

By definition, θ_o is the "true" parameter vector under H_0. Because the null hypothesis specifies that the conditional expectation of $\phi_t(y_t,x_t,\theta_o)$ given the predetermined variables x_t is zero, it is natural to call ϕ_t a "generalized residual vector." For the conditional mean tests in a nonlinear regression model, L = 1, $\theta \equiv \alpha$, and $\phi_t(y_t,x_t,\theta) \equiv u_t(y_t,x_t,\alpha) \equiv y_t$ -

$m_t(x_t,\alpha)$. The tests for heteroskedasticity take $L = 1$, $\theta \equiv (\alpha',\sigma^2)'$, and $\phi_t(y_t,x_t,\theta) \equiv u_t^2(\alpha) - \sigma^2$.

The validity of (2.9) can be tested by choosing functions of the predetermined variables x_t and checking whether the sample covariances between these functions and $\phi_t(y_t,x_t,\theta_o)$ are significantly different from zero. In order to cover a broad range of circumstances that are of interest to economists, it is useful to allow the misspecification indicators to depend on θ and some nuisance parameters. Let $\pi \in \Pi$ denote a Nx1 vector of nuisance parameters. Let $\Lambda_t(x_t,\theta,\pi)$ be an LxQ matrix of misspecification indicators and let $C_t(x_t,\theta,\pi)$ be an LxL, symmetric and positive semi-definite weighting 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 $T^{1/2}(\hat{\pi}_T - \pi_T^o) = O_p(1)$ under H_0, where $\{\pi_T^o$: $T=1,2,\ldots\}$ is a nonstochastic sequence in Π. It is because $\hat{\pi}_T$ need not have an interpretable probability limit under H_0 that π is called a nuisance parameter.

A computable test statistic is the Qx1 vector

$$T^{-1}\sum_{t=1}^{T} \hat{\Lambda}_t'\hat{C}_t\hat{\phi}_t \qquad (2.10)$$

where "^" denotes that each function is evaluated at $\hat{\theta}_T$ or $(\hat{\theta}_T',\hat{\pi}_T')'$ (dependence of the summands in (2.10) on the sample size T is suppressed for convenience). For the conditional mean tests and the heteroskedasticity tests, $\Lambda_t(x_t,\theta,\pi)$ is the 1xQ vector denoted $\lambda_t(x_t,\theta,\pi)$.

From the point of view of simply obtaining tests with known asymptotic size under H_0, the p.s.d. matrix C_t could be absorbed into Λ_t. But the structure in (2.10) is exploited below to generate regression-based tests with the additional property that they are asymptotically equivalent to better known tests in classical circumstances. In the examples discussed thus far $C_t(x_t,\theta,\pi) \equiv 1$. Section 3 covers some cases where it is profitable to allow C_t to be random.

To use (2.10) as the basis for a test of (2.9), the limiting distribution of

$$\hat{\xi}_T \equiv T^{-1/2} \sum_{t=1}^{T} \hat{\Lambda}_t' \hat{C}_t \hat{\phi}_t \tag{2.11}$$

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

$$\xi_T^o \equiv T^{-1/2} \sum_{t=1}^{T} \Lambda_t^{o\prime} C_t^o \phi_t^o \tag{2.12}$$

(values with "o" superscripts are evaluated at θ_o or $(\theta_o', \pi_T^{o\prime})'$) 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 generally derivable from a central limit theorem (provided that $\{\Lambda_t^{o\prime} C_t^o \phi_t^o\}$ is also weakly dependent in an appropriate sense). 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 a standard mean value expansion. In principle, deriving a quadratic form in $\hat{\xi}_T$ which has an asymptotic chi-square distribution is straightforward. But nothing guarantees that the resulting test statistic is easy to compute.

In certain instances test statistics based on $\hat{\xi}_T$ can be computed from simple OLS regressions. The Newey-Tauchen-White approach can be applied when $\hat{\theta}_T$ is the maximum likelihood estimator and the conditional density of y_t given x_t is correctly specified under H_0. In addition to $\hat{\phi}_t$, $\hat{\Lambda}_t$, and $\hat{C}_t$ the score $\hat{s}_t$ of the conditional log-likelihood is needed for computation. The NTW regression is simply

$$1 \text{ on } \hat{s}_t, \hat{\phi}_t'\hat{C}_t\hat{\Lambda}_t, \quad t=1,\ldots,T \tag{2.13}$$

and one uses TR_u^2 as asymptotically χ_Q^2. If interest lies in the case where the entire conditional density is correctly specified under H_0, and $\hat{\theta}_T$ is the maximum likelihood estimator of θ_o, then the Newey-Tauchen-White approach is computationally easier than the present approach. It should be noted, however, that the NTW regression is valid only when $\hat{\theta}_T$ is the MLE, whereas the procedure described below is valid when $\hat{\theta}_T$ is any $\sqrt{T}$-consistent estimator of θ_o. Wooldridge [31] discusses a $C(\alpha)$ version of the NTW statistic that allows $\hat{\theta}_T$ to be any $\sqrt{T}$-consistent estimator. Another possible drawback to the NTW regression is that there is growing evidence that it can yield tests with poor finite sample properties even in the best possible circumstances (Davidson and MacKinnon [6], Bollerslev and Wooldridge [1], Kennan and Neumann [15]). This is at least in part because the NTW regression ignores the generalized residual structure in (2.2) in always using the outer product of the gradient in computing an estimate of the information matrix.

A relatively simple statistic that typically imposes fewer assumptions than the NTW approach is available if $\hat{\xi}_T$ is appropriately modified. Assume that $\theta_o \in \text{int}(\Theta)$ and that ϕ_t is differentiable on $\text{int}(\Theta)$. Define $\Phi_t(x_t,\theta_o) \equiv$

$E[\nabla_\theta \phi_t(y_t, x_t, \theta_o) | x_t]$. Then, instead of basing a test statistic on the covariance of the weighted misspecification indicator $\hat{C}_t^{1/2}\hat{\Lambda}_t$ and the weighted generalized residuals $\hat{C}_t^{1/2}\hat{\phi}_t$, the idea is to first purge from $\hat{C}_t^{1/2}\hat{\Lambda}_t$ its linear projection onto $\hat{C}_t^{1/2}\hat{\Phi}_t$, where $\hat{\Phi}_t \equiv \Phi_t(x_t, \hat{\theta}_T)$. That is, consider the modified statistic

$$\ddot{\xi}_T \equiv T^{-1/2} \sum_{t=1}^{T} \hat{C}_t^{1/2}[\hat{\Lambda}_t - \hat{\Phi}_t \hat{B}_T]' \hat{C}_t^{1/2}\hat{\phi}_t \tag{2.14}$$

where

$$\hat{B}_T \equiv \left[\sum_{t=1}^{T} \hat{\Phi}_t' \hat{C}_t \hat{\Phi}_t\right]^{-1} \sum_{t=1}^{T} \hat{\Phi}_t' \hat{C}_t \hat{\Lambda}_t \tag{2.15}$$

is the PxQ matrix of regression coefficients from the matrix regression

$$\hat{C}_t^{1/2}\hat{\Lambda}_t \quad \text{on} \quad \hat{C}_t^{1/2}\hat{\Phi}_t \qquad t=1,\ldots,T. \tag{2.16}$$

$\ddot{\xi}_T$ can be written more concisely as

$$\ddot{\xi}_T \equiv T^{-1/2} \sum_{t=1}^{T} \ddot{\Lambda}_t' \tilde{\phi}_t \tag{2.17}$$

where $\ddot{\Lambda}_t \equiv \hat{C}_t^{1/2}[\hat{\Lambda}_t - \hat{\Phi}_t \hat{B}_T]$, $t=1,\ldots,T$ are the LxQ matrix residuals from the regression (2.16) and $\tilde{\phi}_t \equiv \hat{C}_t^{1/2}\hat{\phi}_t$. Note that by construction $\ddot{\Lambda}_t$ is weighted by $\hat{C}^{1/2}$.

It is important to realize that $\hat{\xi}_T$ and $\ddot{\xi}_T$ are <u>not</u> always asymptotically equivalent in the sense that $\hat{\xi}_T - \ddot{\xi}_T \xrightarrow{p} 0$ under H_0. The indicators $\hat{\Lambda}_t$ and $[\hat{\Lambda}_t - \hat{\Phi}_t \hat{B}_T]$ generally yield tests with different power functions. Nevertheless, the robust form of the test almost always has a straightforward interpretation. 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) $E[\phi_t(y_t,x_t,\theta_o)|x_t] = 0, \; t=1,2,\ldots;$

(b) $\Phi_t(x_t,\theta_o) \equiv E[\nabla_\theta \phi(y_t,x_t,\theta_o)|x_t], \; t=1,2,\ldots;$

(c) $T^{1/2}(\hat{\theta}_T - \theta_o) = O_p(1), \; T^{1/2}(\hat{\pi}_T - \pi_T^o) = O_p(1).$

Then

$$\ddot{\xi}_T = T^{-1/2} \sum_{t=1}^{T} [\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \phi_t^o + o_p(1) \tag{2.18}$$

where

$$B_T^o \equiv \left(\sum_{t=1}^{T} E[\Phi_t^{o\prime} C_t^o \Phi_t^o] \right)^{-1} \sum_{t=1}^{T} E[\Phi_t^{o\prime} C_t^o \Lambda_t^o].$$

In addition,

$$TR_u^2 \overset{d}{\to} \chi_Q^2,$$

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

$$1 \quad \text{on} \quad [\hat{C}_t^{1/2}\hat{\phi}_t]' \hat{C}_t^{1/2}(\hat{\Lambda}_t - \hat{\Phi}_t \hat{B}_T) \qquad t=1,\ldots,T \tag{2.19}$$

and $\hat{B}_T$ is given by (2.15).

Equation (2.18) has a very useful interpretation. Viewing $\ddot{\xi}_T$ as a function of θ, π, and B evaluated at the estimators $\hat{\theta}_T$, $\hat{\pi}_T$, and $\hat{B}_T$, equation (2.18) demonstrates that the asymptotic distribution of this vector is unchanged when the estimators are replaced by their probability limits. Note that the original statistic $\hat{\xi}_T$ does not generally have this property.

Theorem (2.1) can be applied as follows:

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

(2) Run the matrix regression

$$\tilde{\Lambda}_t \text{ on } \tilde{\Phi}_t \qquad t=1,\ldots,T \tag{2.20}$$

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

(3) Run the regression

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

and use $TR_u^2 = T - RSS$ as asymptotically χ_Q^2 under H_0, assuming that $\ddot{\Lambda}_t$ does not contain redundant indicators.

It must be emphasized that condition (ii.b), which requires that $E[\nabla_\theta \phi_t(y_t, x_t, \theta_o)|x_t]$ be computable under the null hypothesis, can impose additional restrictions on ϕ_t that must be satisfied in order for (1)-(3) to be a valid procedure under H_0. If additional assumptions are used in forming $\Phi_t(x_t, \theta_o)$ then the "implicit null hypothesis" includes more than just (2.9). But as shown in Wooldridge [30], Φ_t is always computable under the relevant null hypothesis for conditional mean (hence conditional probability) or conditional variance testing in a linear exponential family. These are leading - but certainly not the only - cases where one would like to be robust against other distributional misspecifications. Example 3.3 in section 3 shows that no auxiliary assumptions are needed to compute regression-based specification tests of jointly parameterized mean and variance functions that are robust to nonnormality.

In many other situations $\Phi_t(x_t,\theta_o)$ is easily computed if some additional and in many cases standard assumptions are imposed under H_0. For example, in the nonlinear regression example suppose that $\phi_t(y_t,x_t,\theta) = [y_t - m_t(x_t,\alpha)]^3$ where θ contains α and any conditional variance parameters. $\Phi_t(\theta_o)$ is easily seen to be $\Phi_t(\theta_o) = -3\nabla_\alpha m_t(\alpha_o)V(y_t|x_t)$. Most tests for skewness in the literature impose homoskedasticity or some other conditional variance assumption under the null, and $\Phi_t(\theta_o)$ is readily computed once a model for $V(y_t|x_t)$ has been specified. Tests for skewness are typically carried out after the first two moments are thought to be correctly specified. If this is the case, Theorem 2.1 imposes no auxiliary assumptions under the null. However, it should be noted that the choice of Λ_t is limited in this example by the form of Φ_t. If $\hat{\Lambda}_t$ is linearly related to $\hat{\Phi}_t$ then the modified indicator $\tilde{\Lambda}_t$ is simply zero. In a linear model with conditional homoskedasticity and regressors w_t, $\hat{\Phi}_t$ is proportional to w_t. Thus, $\hat{\Lambda}_t$ cannot contain linear combinations of w_t. In particular, if w_t contains unity then the choice $\Lambda_t \equiv 1$ is unavailable; this rules out a standard test for unconditional skewness based on $\sum_{t=1}^{T} \hat{u}_t^3$. A Newey-Tauchen-White test would allow more flexibility in the choice of Λ_t for this example.

As another example, consider testing for nonconstancy of the conditional first absolute moment of the regression errors. Under H_0, $E[|y_t - m_t(x_t,\alpha_o)||x_t] = \kappa_o > 0$. The generalized residual is $\phi_t(y_t,x_t,\theta) = |y_t - m_t(x_t,\alpha)| - \kappa$ where $\theta = (\alpha',\kappa)'$. Although $\phi_t(\theta)$ is not strictly differentiable in α, it is differentiable almost surely under the usual assumptions imposed in these contexts. The quasi-gradient with respect to α is $\nabla_\alpha\phi_t(\theta) = (1[y_t-m_t(\alpha_o) > 0] - 1[y_t-m_t(\alpha_o) \le 0])\nabla_\alpha m_t(\alpha_o)$. Under conditional symmetry of the distribution of y_t given x_t, $E(1[y_t - m_t(\alpha_o) > 0]|x_t) = E(1[y_t - m_t(\alpha_o) \le 0]|x_t)$, so that $E[\nabla_\alpha \phi_t(\theta_o)|x_t] = 0$. Also, $E[\nabla_\kappa \phi_t(\theta_o)|x_t] = 1$, and $\Phi_t(x_t, \theta_o)$ is simply $(0,1)$. If m_t is the conditional mean function and $\hat{\alpha}_T$ is an M-estimator other than the NLS estimator (e.g. the least absolute deviations estimator) then conditional symmetry is needed anyway for $\hat{\alpha}_T$ to be consistent for α_o.

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 $\sqrt{T}$-consistent estimators of θ_o and π_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 (under classical assumptions). An application to the tobit model is given in section 3. The tobit example has the feature of imposing correct specification of the conditional distribution under H_0 but, unlike the usual LM or NTW regressions, Theorem 2.1 can be used when an estimator other than the MLE is available.

A yet unresolved issue is the relationship between $\tilde{\xi}_T$ and $\hat{\xi}_T$. There is a simple characterization of their asymptotic equivalence under H_0. The proof of the following lemma follows immediately from the construction of $\tilde{\xi}_T$.

<u>Lemma</u> <u>2.2</u>: Let the conditions of Theorem 2.1 hold. If, in addition,

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

then

$$\tilde{\xi}_T - \hat{\xi}_T = o_p(1). \tag{2.21}$$

When (iii) holds, $\ddot{\xi}_T$ and $\hat{\xi}_T$ are asymptotically equivalent under H_0. Condition (iii) is usefully interpreted as the sample covariance between $\{\hat{C}_t^{1/2}\hat{\Phi}_t: t=1,..,T\}$ and $\{\hat{C}_t^{1/2}\hat{\phi}_t: t=1,...,T\}$ being asymptotically zero. It is trivially satisfied if

$$\sum_{t=1}^{T} \Phi_t(\theta)' C_t(\theta,\hat{\pi}_T)\phi_t(\theta) \equiv 0 \qquad (2.22)$$

is the defining first-order condition for $\hat{\theta}_T$. This is frequently the case, in particular when $\hat{\theta}_T$ is a quasi-maximum likelihood estimator (QMLE) of the parameters of a conditional mean (see Wooldridge [30]) or of the parameters of a jointly parameterized conditional mean and conditional variance (see Example 3.3 below). In these examples (2.21) also holds (trivially) for local alternatives, so that the difference between the test based on $\ddot{\xi}_T$ and, say, the NTW test based on $\hat{\xi}_T$, is simply that different estimators have been used for the moment matrices 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; robustness is achieved without losing asymptotic efficiency. In addition to having known asymptotic size under H_0, the robust test has a limiting noncentral chi-square distribution even when the auxiliary assumptions are violated under local alternatives (e.g. heteroskedasticity is present in a dynamic regression model).

Lemma 2.2 does not directly provide a description of the local behavior of $\tilde{\xi}_T$ when (iii) fails to hold under local alternatives, but viewed from a slightly different angle it provides useful insight. Note that Theorem 2.1 implies that the quadratic form in $\tilde{\xi}_T$ has an asymptotic chi-square distribution under H_0 regardless of whether or not (iii) holds; the issue is how to characterize the directions of misspecification that $\ddot{\xi}_T$ has power against when (iii) does not hold. Fortunately, it is frequently the case that $\ddot{\xi}_T$ is asymptotically equivalent to <u>some</u> well-known statistic under local alternatives, when classical assumptions hold. This facilitates interpreting a rejection when (iii) fails to hold.

To characterize the local behavior of $\ddot{\xi}_T$, it is useful to be somewhat more explicit about the nature of the local alternatives. Let θ_T^* and π_T^* be nonstochastic sequences such that $\sqrt{T}(\theta_T^* - \theta_o) = O(1)$ and $\sqrt{T}(\pi_T^* - \pi_T^o) = O(1)$. $\{\theta_T^*: T=1,2,\ldots\}$ indexes the sequence of local alternatives, but, as with θ_o under H_0, θ_T^* need not uniquely index the nonnull probability measure. π_T^* is the plim of the estimator $\hat{\pi}_T$ under the sequence of local alternatives $\{H_{T1}: T=1,2,\ldots\}$. Assume that the conditions of Theorem 2.1 are supplemented with conditions of the form

$$E_{\theta_T^*}\left[T^{-1}\sum_{t=1}^{T} G_t(y_t,x_t,\theta_T^*,\pi_T^*)\right] - E_{\theta_o}\left[T^{-1}\sum_{t=1}^{T} G_t(y_t,x_t,\theta_o,\pi_T^o)\right] \to 0$$

as $T \to \infty$ for various functions G_t. This corresponds to standard assumptions in the analysis of the local behavior of test statistics. The arguments of Theorem 2.1 can be used to show that under the sequence of local alternatives $\{H_{T1}\}$,

$$\ddot{\xi}_T = T^{-1/2}\sum_{t=1}^{T}[\Lambda_t^* - \Phi_t^* B_T^*]' C_t^* \phi_t^* + o_p(1) \qquad (2.23)$$

where

$$B_T^* \equiv \left\{\sum_{t=1}^{T} E[\Phi_t^{*\prime} C_t^* \Phi_t^*]\right\}^{-1} \sum_{t=1}^{T} E[\Phi_t^{*\prime} C_t^* \Lambda_t^*]$$

and values with a "*" superscript are evaluated at θ_T^* or (θ_T^*, π_T^*). Equation

(2.23) is the extension of (2.18) to local alternatives and implies that the local limiting distribution of $\ddot{\xi}_T$ is the same when $\hat{\theta}_T$ and $\hat{\pi}_T$ are replaced by their plims θ_T^* and π_T^*, provided that $\sqrt{T}(\hat{\theta}_T - \theta_T^*) = O_p(1)$ and $\sqrt{T}(\hat{\pi}_T - \pi_T^*) = O_p(1)$ under $\{H_{T1}\}$. This implies that if $(\hat{\theta}_{T1},\hat{\pi}_{T1})$ and $(\hat{\theta}_{T2},\hat{\pi}_{T2})$ are both $\sqrt{T}$-consistent estimators of (θ_T^*,π_T^*) under $\{H_{T1}\}$ then

$$\ddot{\xi}_{T1} - \ddot{\xi}_{T2} = o_p(1) \tag{2.24}$$

under $\{H_{T1}\}$, where $\ddot{\xi}_{T1}$ is evaluated at $(\hat{\theta}_{T1},\hat{\pi}_{T1})$ and $\ddot{\xi}_{T2}$ is evaluated at $(\hat{\theta}_{T2},\hat{\pi}_{T2})$. Suppose, in addition, that $\hat{\theta}_{T2}$ and $\hat{\pi}_{T2}$ are <u>chosen</u> to satisfy (iii), i.e.

$$T^{-1/2}\sum_{t=1}^{T}\Phi_t(\hat{\theta}_{T2})'C_t(\hat{\theta}_{T2},\hat{\pi}_{T2})\phi_t(\hat{\theta}_{T2}) = o_p(1). \tag{2.25}$$

Then, by the analog of Lemma 2.2 for local alternatives, $\ddot{\xi}_{T2} - \hat{\xi}_{T2} = o_p(1)$ where $\hat{\xi}_{T2}$ is evaluated at $(\hat{\theta}_{T2},\hat{\pi}_{T2})$. Along with (2.24) this implies that

$$\ddot{\xi}_{T1} - \hat{\xi}_{T2} = o_p(1) \tag{2.26}$$

under H_0 and local alternatives. Conclusion (2.26) is simple yet very powerful. It means that for <u>any</u> $\sqrt{T}$-consistent estimator $\hat{\theta}_{T1}$, $\ddot{\xi}_{T1}$ is asymptotically equivalent to $\hat{\xi}_{T2}$ because $\hat{\xi}_{T2}$ has been evaluated at an estimator that satisfies the asymptotic first order condition (2.22). Whenever such an estimator is available the interpretation of $\ddot{\xi}_T$ is straightforward: $\ddot{\xi}_T$ is asymptotically equivalent to the vector that originally motivated the test statistic, $\hat{\xi}_T$, when $\hat{\xi}_T$ is evaluated at the estimator that solves the first order condition. It does not matter which estimator is used in computing $\ddot{\xi}_T$, provided that it is $\sqrt{T}$-consistent. Thus, the interpretation of $\ddot{\xi}_T$ does <u>not</u> depend on the estimator used in computing it. In many situations there is available an estimator that satisfies (2.22), and typically it solves a well-known problem. Interpreting $\ddot{\xi}_T$ even whether or not (iii) holds typically reduces to interpreting an LM-type statistic in a particular weighted nonlinear regression model or in a model estimated by MLE under normality. An example of a case where an estimator satisfying (2.22) does not have a simple interpretation involves testing for skewness as discussed above. In a linear model with homoskedasticity, the estimator that solves the first order condition (2.22) sets the correlation between the regressors and the third moment of the errors equal to zero. This method of moments estimator is not particularly easy to interpret.

The reasoning of the previous paragraph is applied to the tobit model in the next section. There it is seen that a test for the conditional mean using, e.g., Heckman's [12] two-step estimator, is asymptotically equivalent to a standard Davidson-MacKinnon [5] test for comparing two readily interpretable weighted nonlinear regression models.

The results of Theorem 2.1 and Lemma 2.2 are asymptotic. Very little is known about the finite sample performance of the statistics of Theorem 2.1, especially for nonlinear dynamic models. It should be emphasized, however, that even though the regression in step (3) uses unity as the dependent variable, these statistics do not necessarily have the same finite sample biases sometimes exhibited by outer product-type regressions. Unlike standard outer product regressions, the robust form does exploit the generalized residual form of the test statistic. In fact, the simulations of Davidson and MacKinnon [6] for a static regression model and of Bollerslev and Wooldridge [1] for an AR-GARCH model suggest that the orthogonalization of $\hat{C}_t^{1/2}\hat{\Lambda}_t$ with respect to $\hat{C}_t^{1/2}\hat{\Phi}_t$ in step (2) improves the finite sample performance relative to the NTW outer product regression, even under classical assumptions. That this might be the case was previously suggested to me by Peter Phillips.

3. Examples of Robust, Regression-Based Tests

Example 3.1: Let y_t be a scalar and let $\{m_t(x_t,\alpha): \alpha \in A\}$, $A \subset R^P$, be a parametric family for the conditional expectation of y_t given x_t. The null hypothesis is

$$H_0: E(y_t|x_t) = m_t(x_t,\alpha_o), \text{ some } \alpha_o \in A, \ t=1,2,\ldots. \tag{3.1}$$

Let $\{h_t(x_t,\gamma): \gamma \in \Gamma\}$ be a sequence of weighting functions such that $h_t(x_t,\gamma) > 0$, and suppose that $\hat{\gamma}_T$ is an estimator such that $T^{1/2}(\hat{\gamma}_T - \gamma_T^o) = O_p(1)$, where $\{\gamma_T^o\} \subset \Gamma$. It is not assumed that $\{h_t(x_t,\gamma): \gamma \in \Gamma\}$ contains a version of $V(y_t|x_t)$ or that $h_t(x_t,\gamma_o)$ is proportional to $V(y_t|x_t)$ for some $\gamma_o \in \Gamma$. The researcher chooses a set of weights $\{h_t(x_t,\hat{\gamma}_T)\}$ and performs weighted NLS (WNLS), or uses some other $\sqrt{T}$-consistent estimator for α_o. However, no matter which estimator for α_o is used, the tests are motivated by the WNLS first order condition

$$\sum_{t=1}^{T} \nabla_\alpha m_t(\alpha)'[y_t - m_t(\alpha)]/h_t(\hat{\gamma}_T) = 0. \tag{3.2}$$

A general class of diagnostics is obtained by replacing $\nabla_\alpha m_t(\alpha)$ with a 1xQ vector of misspecification indicators evaluated at the estimators:

$$\sum_{t=1}^{T} \lambda_t(\hat{\alpha}_T,\hat{\pi}_T)'[y_t - m_t(\hat{\alpha}_T)]/h_t(\hat{\gamma}_T) \tag{3.3}$$

where $\hat{\pi}_T$ can contain $\hat{\gamma}_T$ and other nuisance parameters. In the notation of Theorem 2.1, $\theta \equiv \alpha$, $\phi_t(\theta) \equiv y_t - m_t(\alpha)$, $\Lambda_t(\theta,\pi) \equiv \lambda_t(\alpha,\pi)$, and $C_t(\theta,\pi) \equiv 1/h_t(\gamma)$. It is easy to see that computation of $\Phi_t(x_t,\theta_o)$ requires no auxiliary assumptions under H_0, and in fact $\Phi_t(x_t,\theta_o) \equiv -\nabla_\alpha m_t(x_t,\alpha_o)$.

The usual LM-type statistic, which is TR_u^2 from the regression

$$\hat{u}_t/\sqrt{\hat{h}_t} \quad \text{on} \quad \nabla_\alpha \hat{m}_t/\sqrt{\hat{h}_t}, \quad \hat{\lambda}_t/\sqrt{\hat{h}_t}, \quad t=1,\ldots,T,$$

requires that $\hat{\alpha}_T$ be asymptotically equivalent to the WNLS estimator and that $h_t(\hat{\gamma}_T)$ be a consistent estimator of $V(y_t|x_t)$ up to scale. The following procedure is valid under H_0 for any $\sqrt{T}$-consistent estimator $\hat{\alpha}_T$, without any assumptions about $V(y_t|x_t)$:

(i) Let $\hat{\alpha}_T$ be a $\sqrt{T}$-consistent estimator of α_o. Compute the residuals $\hat{u}_t$, the gradient $\nabla_\alpha m_t(\hat{\alpha}_T)$, and the indicator $\lambda_t(\hat{\alpha}_T,\hat{\pi}_T)$. Define $\tilde{u}_t \equiv \hat{h}_t^{-1/2}\hat{u}_t$, $\nabla_\alpha \tilde{m}_t \equiv \hat{h}_t^{-1/2}\nabla_\alpha \hat{m}_t$, and $\tilde{\lambda}_t \equiv \hat{h}_t^{-1/2}\hat{\lambda}_t$;

(ii) Regress $\tilde{\lambda}_t$ on $\nabla_\alpha \tilde{m}_t$ and save the 1xQ residuals, say $\ddot{\lambda}_t$;

(iii) Regress 1 on $\tilde{u}_t\ddot{\lambda}_t$ and use $TR_u^2 = T - RSS$ from this regression as asymptotically χ_Q^2 under H_0.

This procedure with $\hat{h}_t \equiv 1$ was first proposed by Davidson and MacKinnon [6] in the context of a nonlinear regression model with independent errors and unconditional heteroskedasticity. It was independently suggested by Wooldridge [29] for nonlinear, possibly dynamic regression models with conditional or unconditional heteroskedasticity under a martingale difference assumption on the regression errors. Theorem 2.1 further demonstrates that $\hat{\alpha}_T$ need not be the NLS estimator. The indicator $\hat{\lambda}_t$ can be chosen to yield LM tests, Hausman tests based on two WNLS regressions, and tests of nonnested hypotheses, such as the Davidson-MacKinnon [5] test, which do not require correct specification of the conditional variance of y_t given x_t. Conditional mean tests in the more general context of multivariate linear exponential families are considered in more detail in Wooldridge [30].

The estimator that satisfies condition (iii) of Lemma 2.2 is the WNLS estimator based on weights $1/h_t(\hat{\gamma}_T)$. From the remarks following Lemma 2.2, the robust test statistic employing <u>any</u> $\sqrt{T}$-consistent estimator is asymptotically equivalent to the LM statistic based on (3.3) when (3.3) is evaluated at the WNLS estimator, $h_t(\gamma_o)$ is proportional to $V(y_t|x_t)$, and $\hat{\gamma}_T$ is a $\sqrt{T}$-consistent estimator of γ_o. For efficiency reasons it is prudent to put some thought into the choice of h_t.

<u>Example</u> <u>3.2</u>: Suppose now, in the context of Example 3.1, the goal is to test whether for some $\gamma_o \in \Gamma$, $h_t(x_t,\gamma_o)$ is proportional to $V(y_t|x_t)$. Let $v_t(x_t,\gamma) \equiv \sigma^2 h_t(x_t,\gamma)$ where σ^2 is absorbed into γ. The null hypothesis is

$$H_0\colon E(y_t|x_t) = m_t(x_t,\alpha_o),\ V(y_t|x_t) = v_t(x_t,\gamma_o),\ \alpha_o \in A,\ \gamma_o \in \Gamma,\ t=1,2,\ldots. \tag{3.4}$$

Let $\hat{\alpha}_T$ be the WNLS or some other $\sqrt{T}$-consistent estimator of α_o, and let $\hat{\gamma}_T$ be any $\sqrt{T}$-consistent estimator of γ_o. Let $\lambda_t(x_t,\theta,\pi)$ be a 1xQ vector of indicators where $\theta \equiv (\alpha',\gamma')'$. Most tests for variances can be derived from a statistic of the form

$$T^{-1}\sum_{t=1}^{T}\hat{\lambda}_t'[\hat{u}_t^2 - \hat{v}_t]/\hat{v}_t^2 \tag{3.5}$$

Choosing $\lambda(x_t, \theta, \pi)$ to be the nonconstant, nonredundant elements of $\text{vech}[\nabla_\alpha m_t(\alpha)' \nabla_\alpha m_t(\alpha)]$ leads to the White [24] information matrix test in the context of quasi-maximum likelihood estimation in a linear exponential familiy (see Wooldridge [30]). When $v_t(\gamma) \equiv \sigma^2$ choosing $\lambda_t(x_t, \theta, \pi) \equiv w_t$, where w_t is a 1xQ subvector of x_t, leads to the Lagrange Multiplier test for a general form of heteroskedasticity (see Breusch and Pagan [2]). Setting $\lambda_t(x_t, \theta, \pi) \equiv (u_{t-1}^2(\alpha), \ldots, u_{t-Q}(\alpha))$ gives Engle's [8] test for ARCH(Q) under a null of conditional homoskedasticity.

The correspondences for Theorem 2.1 are $L \equiv 1$, $\theta \equiv (\alpha', \gamma')'$, $\phi_t(\theta) = u_t^2(\alpha) - v_t(\gamma)$, $C_t(\theta, \pi) \equiv 1/v_t^2(\gamma)$. Note that $\nabla_\theta \phi_t(\theta) = -2\nabla_\alpha m_t(\alpha) u_t(\alpha) - \nabla_\gamma v_t(\gamma)$. Under H_0, $E[u_t(\alpha_o)|x_t] = 0$ so that $\Phi_t(x_t, \theta_o) \equiv E[\nabla_\theta \phi_t(\theta_o)|x_t] = -\nabla_\gamma v_t(\gamma_o)$; no additional assumptions are needed under H_0 to compute $\Phi_t(x_t, \theta_o)$.

The choice $C_t(\theta, \pi) \equiv 1/v_t^2(\gamma)$ in (3.5) is motivated by the structure of the score of the normal log-likelihood with mean function $m_t(\alpha)$ and variance function $v_t(\gamma)$. In particular, the scaling $1/\hat{v}_t^2$ appears in the the variance tests of Godfrey [10] and Breusch and Pagan [3]. The standard LM statistic in this context is TR_u^2 from the regression

$$(\hat{u}_t^2 - \hat{v}_t)/\hat{v}_t, \quad \nabla_\gamma \hat{v}_t/\hat{v}_t, \quad \hat{\lambda}_t/\hat{v}_t, \quad t=1,\ldots,T. \tag{3.6}$$

In addition to (3.4) this test imposes

$$E[(u_t^o)^4|x_t] = \kappa_o^2[v_t(x_t, \gamma_o)]^2, \quad \text{some } \kappa_o^2 > 0 \tag{3.7}$$

under the null, so that it is nonrobust. Moreover, as pointed out by Breusch and Pagan [3], (3.6) is generally valid only if $\hat{\gamma}_T$ is the QMLE of γ_o under normality. Breusch and Pagan [3] offer a computationally simple $C(\alpha)$ test that allows $\hat{\gamma}_T$ to be any $\sqrt{T}$-consistent estimator of γ_o, but it still requires that (3.7) hold under the null. The NTW procedure applied to this case is valid essentially in the same cases as the usual LM statistic.

The robust procedure obtained from Theorem 2.1 is easy to compute, imposes only (3.4) under the null, and allows $\hat{\gamma}_T$ to be any $\sqrt{T}$-consistent estimator of γ_o.

(i) Let $\hat{\alpha}_T$ be a $\sqrt{T}$-consistent estimator of α_o, and let $\hat{\gamma}_T$ be a $\sqrt{T}$-consistent estimator of γ_o. Compute the residuals $\hat{u}_t$, the gradient $\nabla_\gamma v_t(\hat{\gamma}_T)$, and the indicator $\lambda_t(\hat{\theta}_T,\hat{\pi}_T)$. Define $\tilde{\phi}_t \equiv (\hat{u}_t^2 - \hat{v}_t)/\hat{v}_t = \hat{u}_t^2/\hat{v}_t - 1$, $\nabla_\gamma \tilde{v}_t \equiv \nabla_\gamma \hat{v}_t/\hat{v}_t$, and $\tilde{\lambda}_t \equiv \hat{\lambda}_t/\hat{v}_t$;

(ii) Regress $\tilde{\lambda}_t$ on $\nabla_\gamma \tilde{v}_t$ and save the 1xQ residuals, say $\ddot{\lambda}_t$;

(iii) Regress 1 on $\tilde{\phi}_t\ddot{\lambda}_t$ and use $TR_u^2 = T - RSS$ from this regression as asymptotically χ_Q^2 under H_0.

Interestingly, when $v_t(x_t,\gamma) \equiv \sigma^2$, so that the null is conditional homoskedasticity, the regression in (ii) simply demeans the indicators. Given $\hat{u}_t^2$, $\hat{\sigma}_T^2$, and a choice for $\hat{\lambda}_t$, the χ_Q^2 statistic is obtained as TR_u^2 from the regression

$$1 \quad \text{on} \quad (\hat{u}_t^2 - \hat{\sigma}_T^2)(\hat{\lambda}_t - \bar{\lambda}_T) \qquad t=1,\ldots,T \tag{3.8}$$

where $\bar{\lambda}_T \equiv T^{-1}\sum_{t=1}^{T}\hat{\lambda}_t$. This procedure is asymptotically equivalent to the traditional regression form (2.8) under the additional assumption that $E[u_t^4(\alpha_o)|x_t]$ is constant. Note that (3.6) and the regression (2.8) usually yield different test statistics that are not asymptotically equivalent under

H_0. The demeaning of the indicators may not seem like much of a modification, but it yields an asymptotically chi-square 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 [14] suggestion for a robust form of the White test, but the above statistic is significantly easier to compute.

In the case of the ARCH test, TR_u^2 from the regression in (3.8) is asymptotically equivalent to TR^2 from the regression

$$1 \quad \text{on} \quad (\hat{u}_t^2-\hat{\sigma}_T^2)(\hat{u}_{t-1}^2-\hat{\sigma}_T^2), \ldots, (\hat{u}_t^2-\hat{\sigma}_T^2)(\hat{u}_{t-Q}^2-\hat{\sigma}_T^2) \quad t=Q+1,\ldots T. \tag{3.9}$$

The regression based form in (3.9) 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. Nevertheless, it is asymptotically equivalent to the usual ARCH test under normality.

<u>Example 3.3</u>: Theorem 2.1 can also be applied to models that jointly parameterize the conditional mean and conditional variance. Again, let y_t be a scalar, and consider LM tests that are robust to nonnormality. The unconstrained conditional mean and variance functions are

$$\{\mu_t(x_t,\delta),\ \omega_t(x_t,\delta):\ \delta \in \Delta\} \tag{3.10}$$

where $\Delta \subset R^M$. It is assumed that

$$E(y_t|x_t) = \mu_t(x_t,\delta_o),\quad V(y_t|x_t) = \omega_t(x_t,\delta_o),\ \text{some } \delta_o \in \Delta. \tag{3.11}$$

Take the null hypothesis to be

$$H_0:\ \delta_o = r(\theta_o) \qquad \text{for some } \theta_o \in \Theta \subset R^P \tag{3.12}$$

where $P < M$ and r is continuously differentiable on $\text{int}(\Theta)$. Let $m_t(\theta) \equiv \mu_t(r(\theta))$ and $v_t(\theta) \equiv \omega_t(r(\theta))$ 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 $\hat{\delta}_T \equiv r(\hat{\theta}_T)$ be the constrained estimator of γ_o. $\nabla_\theta \hat{m}_t$ and $\nabla_\theta \hat{v}_t$ are the 1xP gradients of m_t and v_t under H_0. Note that $\hat{\omega}_t = \hat{v}_t$ and $\hat{\mu}_t = \hat{m}_t$ by definition. The LM test of (3.12) is based on the unrestricted score of the quasi-log likelihood evaluated at $\hat{\delta}_T$. The transpose of the score is

$$s_t(\delta)' = \nabla_\delta \mu_t(\delta)' u_t(\delta)/\omega_t(\delta) + \nabla_\delta \omega_t(\delta)' [u_t^2(\delta) - \omega_t(\delta)]/2\omega_t^2(\delta) \tag{3.13}$$

$$= \begin{bmatrix} \nabla_\delta \mu_t(\delta) \\ \nabla_\delta \omega_t(\delta) \end{bmatrix}' \begin{bmatrix} 1/\omega_t(\delta) & 0 \\ 0 & 1/[2\omega_t(\delta)^2] \end{bmatrix} \begin{bmatrix} u_t(\delta) \\ u_t^2(\delta) - \omega_t(\delta) \end{bmatrix}. \tag{3.14}$$

Evaluating s_t at $r(\theta)$ gives

$$s_t(r(\theta))' \equiv \Lambda_t(\theta)' C_t(\theta) \phi_t(\theta) \tag{3.15}$$

where $\Lambda_t(\theta)' \equiv [\nabla_\delta \mu_t(r(\theta))' \mid \nabla_\delta \omega_t(r(\theta))']$, $C_t(\theta)$ is the diagonal matrix in the middle of (3.14) evaluated at $r(\theta)$, and $\phi_t(\theta)' \equiv [u_t(r(\theta)), u_t^2(r(\theta)) - v_t(\theta)]$. The standardized score evaluated at $r(\hat{\theta}_T)$ is

$$T^{-1/2} \sum_{t=1}^{T} s_t(r(\hat{\theta}_T)) \equiv T^{-1/2} \sum_{t=1}^{T} \hat{s}_t. \tag{3.16}$$

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

$$1 \quad \text{on} \quad \hat{s}_t \qquad t=1,\ldots,T \tag{3.17}$$

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, nor does it have systematic power for detecting nonnormality. Theorem (2.1) suggests a robust form of the test. In this case,

$$\underset{2\times 1}{\hat{\phi}_t} \equiv \begin{bmatrix} \hat{u}_t \\ \hat{u}_t^2 - \hat{v}_t \end{bmatrix}, \quad \underset{2\times M}{\hat{\Lambda}_t} \equiv \begin{bmatrix} \nabla_\delta \hat{\mu}_t \\ \nabla_\delta \hat{\omega}_t \end{bmatrix}$$

$$\underset{2\times 2}{\hat{C}_t} \equiv \begin{bmatrix} 1/\hat{v}_t & 0 \\ 0 & 1/[2\hat{v}_t^2] \end{bmatrix}, \quad \underset{2\times P}{\hat{\Phi}_t} \equiv \begin{bmatrix} \nabla_\theta \hat{m}_t \\ \nabla_\theta \hat{v}_t \end{bmatrix},$$

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

$$\tilde{\phi}_t \equiv \begin{bmatrix} \hat{u}_t/\sqrt{\hat{v}_t} \\ [\hat{u}_t^2 - \hat{v}_t]/(\sqrt{2}\hat{v}_t) \end{bmatrix}, \quad \tilde{\Lambda}_t \equiv \begin{bmatrix} \nabla_\delta \hat{\mu}_t/\sqrt{\hat{v}_t} \\ \nabla_\delta \hat{\omega}_t/(\sqrt{2}\hat{v}_t) \end{bmatrix}$$

$$\tilde{\Phi}_t \equiv \begin{bmatrix} \nabla_\theta \hat{m}_t/\sqrt{\hat{v}_t} \\ \nabla_\theta \hat{v}_t/(\sqrt{2}\hat{v}_t) \end{bmatrix}.$$

The robust test statistic is obtained by first running the regression

$$\tilde{\Lambda}_t \text{ on } \tilde{\Phi}_t \qquad t=1,\ldots,T \tag{3.18}$$

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

$$1 \text{ on } \tilde{\phi}_t' \ddot{\Lambda}_t \qquad t=1,\ldots,T \tag{3.19}$$

and use $TR_u^2 = T - RSS$ as asymptotically χ_Q^2 under H_0. Note that the regression in (3.19) 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 a RSS; for those that do not, P regressors can be omitted from (3.19).

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

$$\sum_{t=1}^{T} \Phi_t(\hat{\theta}_T)' C_t(\hat{\theta}_T) \phi_t(\hat{\theta}_T) \equiv 0,$$

so that the robust indicator is asymptotically equivalent to the usual LM indicator. The matrix regression in (3.18) is the cost to the researcher in guarding against nonnormality.

Example 3.4: Suppose that y_t is a random scalar censored below zero, and let x_{t1} be a 1xM vector of predetermined variables from x_t. A popular model for y_t is the tobit model. Among other things, the tobit model implies that

$$E(y_t|y_t>0,x_t) = x_{t1}\alpha_o + \sigma_o\rho(x_{t1}\alpha_o/\sigma_o)$$

$$\equiv m_t(\theta_o) \tag{3.20}$$

and

$$V(y_t|y_t>0,x_t) = \sigma_o^2\{1 - (x_{t1}\alpha_o/\sigma_o)\rho(x_{t1}\alpha_o/\sigma_o) - [\rho(x_{t1}\alpha_o/\sigma_o)]^2\}$$

$$\equiv v_t(\theta_o) \tag{3.21}$$

where $\rho(\cdot)$ is the Mill's ratio and $\theta_o \equiv (\alpha_o,\sigma_o)$. Here σ_o^2 is the conditional variance usually associated with the underlying "latent" variable, and $w_t\alpha_o$ is conditional mean of the latent variable. From a statistical standpoint, the tobit model is no more sensible than

$$\log y_t|y_t>0,x_t \sim N(x_{t1}\beta_o,\eta_o^2) \tag{3.22}$$

((3.22) also seems reasonable for many economic applications; see Cragg [4]). If (3.22) is valid, β_o and η_o^2 can be estimated by OLS of

$$\log y_t \quad \text{on} \quad x_{t1}$$

using only the positive values of y_t. Recall that (3.22) implies

$$E(y_t|y_t>0,x_t) = \exp(\eta_o^2/2 + x_{t1}\beta_o).$$

$$\equiv \mu_t(\beta_o,\eta_o^2) \tag{3.23}$$

and

$$V(y_t|y_t>0,x_t) = [\exp(\eta_o^2) - 1][\exp(\eta_o^2/2 + x_{t1}\beta_o)]^2$$

$$\equiv \omega_t(\beta_o,\eta_o^2) \tag{3.24}$$

Let $\hat{\alpha}_T$, $\hat{\sigma}_T^2$ be any $\sqrt{T}$-consistent estimators of α_o and σ_o^2 under H_0. These include the MLE's, Heckman's [12] two-step estimators, and various WNLS estimators. Define the Davidson-MacKinnon [5] indicator to be the following weighted difference of the predicted values from (3.23 and (3.20):

$$\hat{\lambda}_t \equiv (\hat{v}_t/\hat{\omega}_t)\{\exp[\hat{\eta}_T^2/2 + x_{t1}\hat{\beta}_T] - [x_{t1}\hat{\alpha}_T + \hat{\sigma}_T\rho(x_{t1}\hat{\alpha}_T/\hat{\sigma}_T)]\} \quad (3.25)$$

where $\hat{v}_t \equiv v_t(\hat{\alpha}_T,\hat{\sigma}_T)$ and $\hat{\omega}_t \equiv \omega_t(\hat{\beta}_T,\hat{\eta}_T^2)$. Then, if the tobit model is true, $\hat{\lambda}_t$ should be asymptotically uncorrelated with

$$\hat{u}_t \equiv y_t - x_{t1}\hat{\alpha}_T - \hat{\sigma}_T\rho(x_{t1}\hat{\alpha}_T/\hat{\sigma}_T). \quad (3.26)$$

A test which can be shown to be consistent against the alternative (3.23) is based on the weighted correlation between $\hat{u}_t$ and $\hat{\lambda}_t$. Unfortunately, the usual LM statistic is invalid even when the weighting $1/\sqrt{\hat{v}_t}$ is employed. The reason is that the estimators $(\hat{\alpha}_T,\hat{\sigma}_T)$ need not have been obtained from the weighted nonlinear least squares problem

$$\min_{\alpha,\sigma} \sum_{t=1}^{T} (y_t - x_{t1}\alpha - \sigma\rho(x_{t1}\alpha/\sigma))^2/\hat{v}_t. \quad (3.27)$$

Nevertheless, a statistic is available from Theorem 2.1. Let

$$\phi_t(\alpha,\sigma) \equiv y_t - x_{t1}\alpha - \sigma\rho(x_{t1}\alpha/\sigma) \equiv u_t(\theta)$$

and let $\nabla_\theta\hat{m}_t$ denote the 1 x (M+1) gradient of $x_{t1}\alpha + \sigma\rho(x_{t1}\alpha/\sigma)$ with respect to α and σ, evaluated at $(\hat{\alpha}_T,\hat{\sigma}_T)$. Then the following procedure is asymptotically valid:

(i) Define $\hat{\lambda}_t$ as in (3.25), $\hat{u}_t$ as in (3.26), and $\nabla_\theta \hat{m}_t$ as above. The weighted quantities are $\tilde{\lambda}_t \equiv \hat{\lambda}_t/\sqrt{\hat{v}_t}$, $\tilde{u}_t \equiv \hat{u}_t/\sqrt{\hat{v}_t}$, $\nabla_\theta \tilde{m}_t \equiv \nabla_\theta \hat{m}_t/\sqrt{\hat{v}_t}$.

(ii) Run the OLS regression

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

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

(iii) Run the regression

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

and use $TR_u^2 = T - RSS$ as asymptotically χ_1^2 under H_0.

This test takes the null hypothesis to be correct specification of the conditional density of y_t given x_t, i.e. the tobit model holds under H_0. In particular, it relies on linearity of the conditional expectation in x_{t1}, conditional homoskedasticity, and conditional normality in the underlying latent variable model; it is not intended to be robust to departures from any of these assumptions. Instead the test is devised to have power against departures from the tobit model that invalidate the conditional expectation (3.20). Equation (3.20) is of course only one of many consequences of the tobit model that could be tested. The test is most useful if interest lies in determining the effect of explanatory variables on positive values of the dependent variable.

The discussion following Lemma 2.2 implies that (i)-(iii) is asymptotically equivalent to the Davidson-MacKinnon test for weighted NLS estimation of (3.20) and (3.23), provided $\hat{\alpha}_T$ and $\hat{\sigma}_T$ are $\sqrt{T}$-consistent estimators. If $\hat{\alpha}_T$ and $\hat{\sigma}_T$ are the MLE's then these estimators are more efficient than the WNLS estimators that solve (3.27). But Heckman's [12] two step estimators yield a test that is asymptotically equivalent to the test employing the MLE's or the WNLS estimators. One is essentially doing specification testing of the nonlinear regression model (3.20) with variance (3.21). This makes interpretation of procedure (i)-(iii) straightforward.

It must be emphasized that if the MLE's are available then a Newey-Tauchen-White regression (2.13), which requires the estimated scores of the conditional log-likelihood, is available. The procedure obtained from Theorem 2.1 is more flexible albeit less efficient.

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 (3.22) such as in Cragg [4].

Before leaving this example, it is useful to note that simple tests for exclusion restrictions can be developed along similar lines. The unrestricted mean function for $E(y_t|y_t>0,x_t)$ is

$$x_{t1}\alpha_{o1} + x_{t2}\alpha_{o2} + \sigma\rho([x_{t1}\alpha_{o1} + x_{t2}\alpha_{o2}]/\sigma_o). \qquad (3.28)$$

The null hypothesis is that $\alpha_{o2} = 0$, which reduces to (3.20). The indicator $\hat{\lambda}_t$ is now the gradient of (3.28) with respect to α_2 evaluated at the restricted estimates:

$$\hat{\lambda}_t = x_{t2} + \hat{\sigma}_T\nabla_z\rho(x_{t1}\hat{\alpha}_{T1}/\hat{\sigma}_T)x_{t2}$$

where $\nabla_z\rho(\cdot)$ is the derivative of the Mill's ratio. When this $\hat{\lambda}_t$ is used in (i)-(iii) a test asymptotically equivalent to the LM statistic in the context of WNLS is obtained. Again, $\hat{\alpha}_{T1}$ and $\hat{\sigma}_T$ are any $\sqrt{T}$ consistent estimators of α_{o1} and σ_o.

4. Conclusions

This paper has developed a general class of regression-based specification tests for (possibly) dynamic multivariate models which, in the leading cases, imposes under H_0 only the hypotheses being tested (correctness of the conditional mean and/or correctness of the conditional variance). The framework can be applied to testing other aspects of a conditional distribution under a modest number of additional assumptions. 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 WNLS procedure or the even more efficient MLE. Another application is to 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 applies directly to linear simultaneous equations models (SEM's). Computation of $\Phi_t(x_t,\theta_o)$ is straightforward provided that the reduced form for the endogenous right hand side variables is available. The parameter vector θ contains the reduced form parameters of the relevant endogenous variables as well as the structural parameters in the equation(s) of interest. If all equations of a simultaneous system are being tested jointly then θ is simply all structural parameters. An immediate application is to testing for serial correlation in linear dynamic simultaneous equations models in the presence of heteroskedasticity (conditional or unconditional) of unknown form. Also, tests for multivariate ARCH in SEM's that are robust to heterokurtosis are also easily constructed. The scope of applications to nonlinear SEM's is limited by one's ability to compute $\Phi_t(x_t, \theta_o) \equiv E[\nabla_\theta \phi(y_t, x_t, \theta_o) | x_t]$. This is exactly the problem of computing the optimal instrumental variables for nonlinear SEM's.

Theorem 2.1 can be extended to certain unit root time series models. The initial purging of $\hat{C}_t^{1/2}\hat{\Phi}_t$ from $\hat{C}_t^{1/2}\hat{\Lambda}_t$ can produce indicators $\ddot{\Lambda}_t$ that are effectively stationary. This happens for the LM test in linear time series models when the regressors excluded under the null hypothesis are individually cointegrated (in a generalized sense) with the regressors included under the null. In this context the 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.

References

1. Bollerslev, T. and J.M. Wooldridge. Quasi-maximum likelihood estimation of dynamic models with time-varying covariances. MIT Department of Economics Working Paper No. 505, 1988.

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

3. Breusch, T.S. and A.R. Pagan. The Lagrange Multiplier statistic and its application to model specification in econometrics. *Review of Economic Studies* 47 (1980): 239-253.

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

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

6. Davidson, R. and J.G. MacKinnon. Heteroskedasticity-robust tests in regression directions. *Annales de l'INSEE* 59/60 (1985): 183-218.

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

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

9. Engle, R.F. Wald, Likelihood Ratio, and Lagrange Multiplier tests in econometrics. In Z. Griliches and M. Intriligator (ed.), *Handbook of Econometrics*, Vol. II, Amsterdam: North Holland (1984): 775-826.

10. Godfrey, L.G. Testing for multiplicative heteroskedasticity. *Journal of Econometrics* 8 (1978): 227-236.

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

12. 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.

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

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

15. Kennan, J. and G.R. Neumann. Why does the information matrix test reject too often? University of Iowa Department of Economics Working Paper No. 88-4, 1988.

16. Koenker, R. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics* 17 (1981): 107-112.

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

18. Neyman. J. Optimal asymptotic tests of composite statistical hypotheses. In U. Grenander (ed.), *Probability and Statistics*, Stockholm: Almquist and Wiksell (1959): 213-234.

19. Pagan, A.R. and A.D. Hall. Diagnostic tests as residual analysis. *Econometric Reviews* 2 (1983): 159-218.

20. Pagan, A.R., P.K. Trivedi, and A.D. Hall. Assessing the variability of inflation. *Review of Economic Studies* 50 (1983): 585-596.

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

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

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

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

25. White, H. *Asymptotic Theory for Econometricians*. New York: Academic Press, 1984.

26. White, H. Specification testing in dynamic models. In T. Bewley (ed.), *Advances in Econometrics - Fifth World Congress*, Vol. I, New York: Cambridge University Press (1987): 1-58.

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

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

29. Wooldridge, J.M. A regression-based Lagrange Multiplier statistic that is robust in the presence of heteroskedasticity. MIT Deparment of Economics Working Paper No. 478, 1987.

30. Wooldridge, J.M. Specification testing and quasi-maximum likelihood estimation. MIT Department of Economics Working Paper No. 479, 1987.

31. Wooldridge, J.M. On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. Manuscript, MIT Department of Economics, 1988.

Mathematical Appendix

For convenience, I include a lemma that 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 R^P, and the sequence of nonrandom functions $\{\bar{Q}_T(\theta): \theta \in \Theta, T=1,2,\ldots\}$, satisfy the following conditions:

(i) $\sup_{\theta\in\Theta} |Q_T(w_T,\theta) - \bar{Q}_T(\theta)| \xrightarrow{p} 0$;

(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 $\ddot{\theta}_T - \theta_T^o \xrightarrow{p} 0$ where $\{\theta_T^o\} \subset \Theta$. Then

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

Proof: see Wooldridge [28, 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 R^P, is said to satisfy the Uniform Weak Law of Large Numbers (UWLLN) and Uniform Continuity (UC) conditions provided that

$$\text{(i)} \quad \sup_{\theta\in\Theta} \left|T^{-1}\sum_{t=1}^{T} q_t(y_t,x_t,\theta) - E[q_t(y_t,x_t,\theta)]\right| \xrightarrow{p} 0$$

and

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

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 1xL function of the Px1 vector θ then, by convention, $\nabla_\theta a(\theta)$ is the LxP matrix $\nabla_\theta[a(\theta)']$. If $A(\theta)$ is a QxL matrix then the matrix $\nabla_\theta A(\theta)$ is the LQxP matrix defined as

$$\nabla_\theta A(\theta)' \equiv [\nabla_\theta A_1(\theta)' | \ldots | \nabla_\theta A_Q(\theta)']$$

where $A_j(\theta)$ is the jth row of $A(\theta)$ and $\nabla_\theta A_j(\theta)$ is the LxP gradient of $A_j(\theta)$ as defined as above. Also, for any Lx1 vector function φ, define the second derivative of φ to be the LPxP matrix

$$\nabla_\theta^2 \varphi(\theta) \equiv \nabla_\theta[\nabla_\theta \varphi(\theta)'].$$

Finally, define the parameter vector $\delta \equiv (\theta', \pi')'$.

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

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

(ii) $\theta_o \in \text{int}(\Theta)$, $\{\pi_T^o: T=1,2,\ldots\} \subset \text{int}(\Pi)$ uniformly in T;

(iii) (a) $\{\phi_t(y_t, x_t, \theta): \theta \in \Theta\}$ is a sequence of Lx1 functions such that $\phi_t(\cdot, \theta)$ is Borel measurable for each $\theta \in \Theta$ and $\phi_t(y_t, x_t, \cdot)$ is continuously differentiable on the interior of Θ for all y_t, x_t, $t=1,2,\ldots$;

(b) Define $\Phi_t(x_t, \theta_o) \equiv E_{\theta_o}[\nabla_\theta \phi_t(y_t, x_t, \theta_o) | x_t]$ for all $\theta_o \in \text{int}(\Theta)$. Assume that $\Phi_t(x_t, \cdot)$ is continuously differentiable on the interior of Θ for all x_t, $t=1,2,\ldots$;

(c) $\{C_t(x_t, \delta): \delta \in \Delta\}$ is a sequence of LxL 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, $t=1,2,\ldots$;

(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) $\{\Phi_t(\theta)' C_t(\delta)\Phi_t(\theta)\}$ and $\{\Phi_t(\theta)' C_t(\delta)\Lambda_t(\delta)$ satistfy the UWLLN and UC conditions;

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

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

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

(vii) (a) $\{\Lambda_t(\delta)' C_t(\delta)\nabla_\theta \phi_t(\theta)\}$, $\{[I_Q \otimes \phi_t(\theta)' C_t(\delta)]\nabla_\delta \Lambda_t(\delta)'\}$, $\{\Lambda_t(\delta)'[I_L \otimes \phi_t(\theta)']\nabla_\delta C_t(\delta)\}$, and $\{\Phi_t(\theta)'[I_L \otimes \phi_t(\theta)']\nabla_\delta C_t(\delta)\}$ satisfy the UWLLN and UC requirements;

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

(b) $\Xi_T^{o-1/2} T^{-1/2} \sum_{t=1}^{T} (\Lambda_t^o \ - \Phi_t^o B_T^o)' C_t^o \phi_t^o \ \stackrel{d}{\to} \ N(0,I_Q)$;

(c) $\{\Lambda_t(\delta)' C_t(\delta)\phi_t(\theta)\phi_t(\theta)' C_t(\delta)\Lambda_t(\delta)\}$, $\{\Lambda_t(\delta)' C_t(\delta)\phi_t(\theta)\phi_t(\theta)' C_t(\delta)\Phi_t(\theta)\}$, and $\{\Phi_t(\theta)' C_t(\delta)\phi_t(\theta)\phi_t(\theta)' C_t(\delta)\Phi_t(\theta)\}$ satisfy the UWLLN and UC conditions.

<u>Proof</u> <u>of</u> <u>Theorem</u> <u>2.1</u>: 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,

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

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

$$T^{-1/2}\sum_{t=1}^{T} \hat{\Phi}_t' \hat{C}_t \hat{\phi}_t = T^{-1/2}\sum_{t=1}^{T} \Phi_t^{o\prime} C_t^o \phi_t^o \qquad (a.2)$$

$$+ T^{-1}\sum_{t=1}^{T} \{\Phi_t^{o\prime} C_t^o \nabla_\theta \phi_t^o + [I_p \otimes \phi_t^{o\prime} C_t^o] \nabla_\theta \Phi_t^o\} \; T^{1/2}(\hat{\theta}_T - \theta_o)$$

$$+ T^{-1}\sum_{t=1}^{T} \{\Phi_t^{o\prime} [I_L \otimes \phi_t^{o\prime}] \nabla_\delta C_t^o\} \; T^{1/2}(\hat{\delta}_T - \delta_o) + o_p(1).$$

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,

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

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

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

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

$$\tilde{\xi}_T = T^{-1/2}\sum_{t=1}^{T} [\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \phi_t^o \qquad (a.5)$$

$$+ T^{-1}\sum_{t=1}^{T} \{[\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \nabla_\theta \phi_t^o - B_T^{o\prime} [I_p \otimes \phi_t^{o\prime} C_t^o] \nabla_\theta \Phi_t^o\} \; T^{1/2}(\hat{\theta}_T - \theta_o)$$

$$+ T^{-1}\sum_{t=1}^{T}\{[I_Q \otimes \phi_t^{o\prime} C_t^o]\nabla_\delta \Lambda_t^{o\prime} + [\Lambda_t^o - \Phi_t^o B_T^o]'[I_L \otimes \phi_t^{o\prime}]\nabla_\delta C_t^o\} \cdot T^{1/2}(\hat{\delta}_T - \delta_T^o)$$

$$+ o_p(1).$$

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 by definition of Φ_t^o and the law of iterated expectations,

$$E([\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \nabla_\theta \phi_t^o) = E\{E([\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \nabla_\theta \phi_t^o | x_t])\} \tag{a.6}$$

$$= E\{[\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \Phi_t^o\}.$$

Therefore,

$$T^{-1}\sum_{t=1}^{T} E\{[\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \nabla_\theta \phi_t^o\} = T^{-1}\sum_{t=1}^{T} E\{[\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \Phi_t^o\} = 0 \tag{a.7}$$

by definition of B_T^o. The regularity conditions imposed imply that each of the averages appearing in (a.7) satisfy the WLLN. Therefore

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

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:

$$\tilde{\xi}_T = T^{-1/2}\sum_{t=1}^{T} [\Lambda_t^o - \Phi_t^o B_T^o]' C_t^o \phi_t^o + o_p(1). \tag{a.9}$$

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

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

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

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

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

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

and $\tilde{\phi}_t$ and $\ddot{\Lambda}_t$ are as defined in the text. Because the dependent variable in regression (a.12) is unity, $TR_u^2 = T - RSS$, where RSS is the residual sum of squares from the regression (a.12).

3-22-89