This paper develops a general approach to robust, regression-based specification tests for (possibly) dynamic econometric models. A useful 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, while maintaining asymptotic efficiency under ideal conditions. 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 robust specification tests for a dynamic linear model estimated by two-stage least squares.

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 [11] specification tests, White's [23] information matrix (IM) test, and regression-based versions of various nonnested hypotheses tests. In fact, Newey [15], Tauchen [20], and White [25] 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 [15] and Tauchen [20] have devised outer product-type auxiliary regressions for computing CM tests. White [25] has extended these methods to a general dynamic setting.

Adrian Pagan, Peter Phillips, Jim Poterba, Danny Quah, Tom Stoker, Hal White, and an anonymous referee provided helpful comments and suggestions on previous versions of this paper.

© 1990 Cambridge University Press 0261-4666/90 $5.00 + .00
The simplicity of most popular regression-based procedures currently employed is not without cost. Particularly when testing hypotheses about the conditional mean and/or the conditional variance, the validity of popular regression-based procedures 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. The Newey-Tauchen-White (NTW) approach pertains to maximum likelihood specification testing and generally takes the null to be correct specification of the entire distribution. Except in special cases, the NTW statistic in the context of nonlinear least squares is also invalid under conditional or unconditional heteroskedasticity. Other examples include the various tests for heteroskedasticity: currently used regression forms require constancy of the conditional fourth moment of the errors under the null hypothesis. Also, 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 nit-picking about regularity conditions.

Due primarily to the work of White [21,22,23,25], Domowitz and White [6], Hansen [10], and Newey [15], there now exist general methods of computing robust statistics. In the context of linear regression models, Pagan and Hall [18] discuss how to compute conditional mean tests that are robust to heteroskedasticity. Their discussion centers around the use of the White [21] heteroskedasticity-consistent standard errors. For 1 degree of freedom tests the Pagan and Hall suggestion leads to easily computable statistics since most regression packages now compute the White standard errors. Computation of the Wald-type statistics for tests with more than 1 degree of freedom is somewhat more cumbersome and cannot be carried out with all regression programs. 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 reestimating the model. Also, 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 [8], White [23,24]), formulas for general nonlinear restrictions involve an analytical expression for the derivative of the implicit constraint function and a generalized inverse. In specific instances computa-
tionally simple robust LM statistics are available. A notable example is the paper by Davidson and MacKinnon [5], 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 [21] 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 hypotheses 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 linear least-squares regressions. The general method suggested here can be viewed as an extension of the Davidson and MacKinnon [5] 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 same context the approach here can be viewed as the Lagrange multiplier version of the robust Wald strategy suggested by Pagan and Hall [18]. This paper also extends Wooldridge's [28] robust, regression-based conditional mean and conditional variance tests in the context of quasi-maximum likelihood estimation in multivariate 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 nonrobust counterparts; robustness is obtained without sacrificing asymptotic efficiency.

For some specification tests the current approach does impose auxiliary assumptions under the null hypothesis. 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 of this paper is to develop a fairly broad class
of robust tests that can be computed via linear regressions. The unified approach substantially reduces the analytical and in many cases the computational burden that is associated with a case-by-case analysis.

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 shown to be invariant with respect to the asymptotic distribution of the estimators used in computation; this can be viewed as another type 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 violations 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 [17] \( C(\alpha) \) tests, but they are applicable whether or not the score of the log-likelihood is the basis for the test statistic. When restricted to LM tests the new statistics offer generalized residual alternatives to outer product-type \( C(\alpha) \) statistics.

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 \( y_t \times J, z_t \times K; y_t \) 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 \( x_t = (z_t, y_{t-1}, z_{t-1}, \ldots, y_1, z_1) \) denote the predetermined variables. Current \( z_t \) can be excluded from \( x_t \) or, if there are no "exogenous" variables, one may take \( x_t = (y_{t-1}, y_{t-2}, \ldots, y_1) \). For cross section applications set \( x_t = z_t \) 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. 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 some examples. The first example concerns specification testing of a conditional mean. Suppose that 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

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

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

$$H_0 : E(y_t|x_t) = m_t(x_t, \alpha_0), \quad \text{some } \alpha_0 \in A, \quad t = 1, 2, \ldots$$

(2.2)

If $\hat{\alpha}_T$ is a $\sqrt{T}$-consistent estimator of $\alpha_0$ under $H_0$ then the residuals are defined as

$$u_t(y_t, x_t, \hat{\alpha}_T) = y_t - m_t(x_t, \hat{\alpha}_T).$$

A test of $H_0$ can be based on the sample covariance

$$T^{-1} \sum_{t=1}^{T} \lambda_t(x_t, \hat{\alpha}_T, \hat{\pi}_T)^\prime u_t(y_t, x_t, \hat{\alpha}_T)$$

(2.3)

$$= T^{-1} \sum_{t=1}^{T} \tilde{\lambda}_t \hat{u}_t,$$

(2.4)

where $\lambda_t(x_t, \alpha, \tau)$ is a $1 \times Q$ 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

$$\hat{u}_t \text{ on } \nabla_\alpha \hat{m}_t, \tilde{\lambda}_t, \quad t = 1, \ldots, T.$$ (2.5)

If $\hat{\alpha}_T$ is asymptotically equivalent to the nonlinear least squares (NLS) estimator then under $H_0$ and conditional homoskedasticity, $TR^2$ is asymptotically $\chi^2_Q$. Thus, the LM approach effectively takes the null hypothesis to be

$$H^*_0 : H_0 \text{ holds and } V(y_t|x_t) = \sigma^2_0 \quad \text{for some } \sigma^2_0 > 0, \quad t = 1, 2, \ldots$$

(2.6)

but it is inconsistent for the alternative

$$H^*_1 : H_0 \text{ holds but } H^*_0 \text{ does not.}$$

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

The NTW regression for the same problem is

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

In general, $H^*_0$ is also required for $TR^2$ from this regression to be asymptotically $\chi^2_Q$, 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 [18], a robust test is available from the regression (2.5). The White [21] 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 this is somewhat more complicated as 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 only when $\hat{\alpha}_T$ is asymptotically equivalent to the NLS estimator.

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

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

$$H_0: E(y_t|x_t) = m_t(x_t, \alpha_0) \quad \text{and} \quad V(y_t|x_t) = \sigma^2_0, \quad \alpha_0 \in A, \quad \sigma^2_0 > 0,$$

where $A$ is the set of potential parameter values. 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 $\theta = (\alpha', \sigma^2)$. A general class of tests is based on

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

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

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

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

$TR_c^2$ is asymptotically $\chi^2_Q$ under

$$H^*_0: H_0 \text{ holds and, in addition, } E[(u_t^o)^4|x_t] = \kappa^2_0 > 0, \quad t = 1, 2, \ldots$$

where $u_t^o = y_t - m_t(x_t, \alpha_o)$. Regression (2.8) yields the “studentized” version of the Breusch-Pagan [2] test as derived by Koenker [14]. The studentized form of the test is robust to certain departures from normality and is now widely used in the literature (see, e.g., Engle [7], Pagan and Hall [18], and Pagan, Trivedi, and Hall [19]). 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 de-
degrees of freedom that otherwise depend on \( x_t \), then \( H_0 \) holds but \( H'_0 \) does not. Hsieh [12] and Pagan and Hall [18] note that, just as with conditional mean tests, the White [21] covariance matrix can be used to compute heteroskedasticity tests that are robust to heterokurtosis. Pagan, Trivedi, and Hall [19] report the White [21] \( t \)-statistic in a model for the variance of inflation when \( \lambda_t \) is a scalar. When \( \lambda_t \) is a vector, the LM-type robust form of the test, which is discussed in Example 3.2, is easy to compute via linear regressions and 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 so that \( H_0 \) includes only the hypotheses of interest. To be attractive these tests should be easy to compute in 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 \( L \times 1 \) random function defined on a parameter set \( \Theta \subset \mathbb{R}^K \). The null hypothesis of interest is expressed as

\[
H_0 : E[\phi_t(y_t, x_t, \theta_0)|x_t] = 0, \quad \text{for some } \theta_0 \in \Theta, \quad t = 1, 2, \ldots \quad (2.9)
\]

By definition, \( \theta_0 \) is the "true" parameter vector under \( H_0 \). Because the null hypothesis specifies that the conditional expectation of \( \phi_t(y_t, x_t, \theta_0) \) given the predetermined variables \( x_t \) is zero, it is natural to call \( \phi_t \) a "generalized residual vector." For the conditional mean tests in a nonlinear regression model, \( L = 1, \theta = \alpha, \) and \( \phi_t(y_t, x_t, \theta) = u_t(y_t, x_t, \alpha) = y_t - m_t(x_t, \alpha) \). The tests for heteroskedasticity take \( L = 1, \theta = (\alpha', \sigma^2)' \), and \( \phi_t(y_t, x_t, \theta) = u_t^2(\alpha) - \sigma^2 \).

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 \( \phi_t(y_t, x_t, \theta_0) \) are significantly different from zero. In order to cover a wide range of circumstances that are of interest to economists, it is useful to allow the misspecification indicators to depend on \( \theta \) and some nuisance parameters. Let \( \pi \in \Pi \) denote an \( N \times 1 \) vector of nuisance parameters. Let \( \Lambda_t(x_t, \theta, \pi) \) be an \( L \times Q \) matrix of misspecification indicators and let \( C_t(x_t, \theta, \pi) \) be an \( L \times L \) symmetric and positive semidefinite (p.s.d.) weighting matrix. Assume the availability of an estimator \( \hat{\theta}_T \) such that \( T^{1/2}(\hat{\theta}_T - \theta_0) = o_p(1) \) under \( H_0 \). Also, assume that the nuisance parameter estimator \( \hat{\pi}_T \) is such that \( T^{1/2}(\hat{\pi}_T - \pi_0^2) = o_p(1) \) under \( H_0 \) for some nonstochastic sequence \{\pi_0^2 : T = 1, 2, \ldots \} \) in \( \Pi \). It is because \( \hat{\pi}_T \) need not have an interpretable probability limit under \( H_0 \) that \( \pi \) is called a nuisance parameter.
A computable test statistic is the \( Q \times 1 \) vector
\[
T^{-1} \sum_{t=1}^{T} \hat{\Lambda}_t \hat{C}_t \hat{\phi}_t, \tag{2.10}
\]
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 \( 1 \times Q \) vector earlier 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 \( \Lambda_t \). 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 \( C_t(x_t, \theta, \pi) = 1 \). Section 3 covers some cases where it is profitable to allow \( C_t \) to be nonconstant.

To use (2.10) as the basis for a test of (2.9), the limiting distribution of
\[
\hat{\xi}_T = T^{-1/2} \sum_{t=1}^{T} \hat{\Lambda}_t \hat{C}_t \hat{\phi}_t, \tag{2.11}
\]
under \( H_0 \) is needed. In general, finding the asymptotic distribution of \( \hat{\xi}_T \) under \( H_0 \) entails finding the limiting distribution of
\[
\xi_T^o = T^{-1/2} \sum_{t=1}^{T} \Lambda_t^o \Lambda_t^{o'} C_t^o \phi_t^o \tag{2.12}
\]
[values with "o" superscripts are evaluated at \( \theta_o \) or \((\theta_o', \pi_o')'\)] and the limiting distribution of \( T^{1/2}(\hat{\theta}_T - \theta_o) \) [the limiting distribution of \( T^{1/2}(\hat{\pi}_T - \pi_o) \) does not affect that of \( \hat{\xi}_T \) under \( H_0 \)]. Because \( \xi_T^o \) 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 \( \{\Lambda_t^o, C_t^o \phi_t^o\} \) 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 \( \xi_T^o \) and \( T^{1/2}(\hat{\theta}_T - \theta_o) \) and differentiability assumptions on \( \Lambda_t, C_t, \) and \( \phi_t \), 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. This standard approach generally requires estimation of the asymptotic variance of \( \hat{\theta}_T \), and 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 ordinary least squares (OLS) regressions. The NTW 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

\[
1 \text{ on } \hat{s}_t, \quad \hat{\phi}_t^\prime \hat{C}_t \hat{\Lambda}_t, \quad t = 1, \ldots, T
\]  

(2.13)

and one uses \( TR^2_0 \) as asymptotically \( \chi^2_0 \). 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 \( \theta_0 \), then the NTW 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 maximum likelihood estimate (MLE), whereas the procedure described below is valid when \( \hat{\theta}_T \) is any \( \sqrt{T} \)-consistent estimator of \( \theta_0 \). Wooldridge [29] derives 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 [5], Bollerslev and Wooldridge [1], Kennan and Neumann [13]). This is at least in part because the NTW regression ignores the generalized residual structure in (2.2) by 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 \( \theta_0 \in \text{int}(\theta) \) and that \( \phi_t \) is differentiable on \( \text{int}(\Theta) \). Define \( \hat{\phi}_t(x_t, \theta_0) = E[\nabla_\theta \phi_t(y_t, x_t, \theta_0)]|x_t \), where an implicit assumption is that this conditional expectation can be computed under \( H_0 \). 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 \( \hat{\phi}_t \equiv \hat{\phi}_t(x_t, \hat{\theta}_T) \). That is, consider the modified statistic

\[
\hat{\xi}_T = T^{-1/2} \sum_{t=1}^T \left( \hat{C}_t^{1/2} \hat{\Lambda}_t - \hat{\phi}_t \hat{B}_T \right)' \hat{C}_t^{1/2} \hat{\phi}_t
\]  

(2.14)

where

\[
\hat{B}_T = \left( \sum_{t=1}^T \hat{\phi}_t' \hat{C}_t \hat{\phi}_t \right)^{-1} \sum_{t=1}^T \hat{\phi}_t' \hat{C}_t \hat{\Lambda}_t
\]  

(2.15)

is the \( P \times Q \) matrix of regression coefficients from the matrix regression

\[
\hat{C}_t^{1/2} \hat{\Lambda}_t \text{ on } \hat{C}_t^{1/2} \hat{\phi}_t, \quad t = 1, \ldots, T.
\]  

(2.16)

\( \hat{\xi}_T \) can be written more concisely as

\[
\hat{\xi}_T = T^{-1/2} \sum_{t=1}^T \hat{\Lambda}_t' \hat{\phi}_t
\]  

(2.17)
where $\hat{\lambda}_t \equiv \hat{C}_t^{1/2}(\hat{\lambda}_t - \hat{\phi}_t \hat{B}_T)$, $t = 1, \ldots, T$ are the $L \times Q$ matrix residuals from the regression (2.16), and $\hat{\phi}_t \equiv \hat{C}_t^{1/2} \hat{\phi}_t$. Note that by construction $\hat{\lambda}_t$ is weighted by $\hat{C}_t^{1/2}$.

It is important to realize that $\hat{\xi}_T$ and $\hat{\xi}_T$ are not always asymptotically equivalent in the sense that $\hat{\xi}_T - \hat{\xi}_T \equiv 0$ under $H_0$. The indicators $\hat{\lambda}_t$ and $(\hat{\lambda}_t - \hat{\phi}_t \hat{B}_T)$ generally test for misspecification in different directions. Nevertheless, the robust form of the test almost always has a straightforward interpretation, and in many cases it is asymptotically equivalent to a statistic based on $\hat{\xi}_T$ when the latter is valid. I return to this issue below.

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

**THEOREM 2.1.** Assume that the following conditions hold under $H_0$:

(i) Regularity conditions A.1 in the appendix;
(ii) For some $\theta_0 \in \text{int}(\Theta)$,
   
   (a) $E[\phi_t(y_t, x_t, \theta_0)|x_t] = 0$, $t = 1, 2, \ldots$;
   
   (b) $\Phi_t(x_t, \theta_0) = E[\nabla \phi(y_t, x_t, \theta_0)|x_t]$, $t = 1, 2, \ldots$;
   
   (c) $T^{1/2}(\theta_T - \theta_0) = O_p(1), \quad T^{1/2}(\pi_T - \pi_0) = O_p(1)$.

Then

$$\hat{\xi}_T = T^{-1/2} \sum_{t=1}^{T} [\hat{\phi}_t - \Phi_t B_T^{\phi}]' C_t^{\phi} \hat{\phi}_t + o_p(1) \quad (2.18)$$

where

$$B_T^{\phi} \equiv \left( \sum_{t=1}^{T} E[\Phi_t' C_t^{\phi} \Phi_t] \right)^{-1} \sum_{t=1}^{T} E[\Phi_t' C_t^{\phi} \hat{\phi}_t].$$

In addition,

$$TR_u^{2} \stackrel{d}{\rightarrow} x_\theta^{2},$$

where $R_u^{2}$ is the uncentered $r$-squared from the regression 1 on $[\hat{C}_t^{1/2} \hat{\phi}_t]' \hat{C}_t^{1/2}(\hat{\lambda}_t - \hat{\phi}_t \hat{B}_T)$, $t = 1, \ldots, T$ (2.19) and $\hat{B}_T$ is given by (2.15).

Equation (2.18) has a very useful interpretation. Viewing $\hat{\xi}_T$ as a function of $\theta$, $\pi$, and $B$ evaluated at the estimators $\hat{\theta}_T$, $\hat{\pi}_T$, and $\hat{B}_T$, equation (2.18) states 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:

**Procedure 2.1**

(i) Given \( \Lambda_t, C_t, \phi_t, \hat{\theta}_t, \) and \( \hat{z}_t, \) compute \( \hat{\Lambda}_t, \hat{C}_t, \hat{\phi}_t, \) and \( \hat{\theta}_t. \) Define \( \tilde{\Lambda}_t = \hat{C}_t^{1/2} \hat{\Lambda}_t, \) \( \tilde{\phi}_t = \hat{C}_t^{1/2} \hat{\phi}_t, \) and \( \tilde{\theta}_t = \hat{C}_t^{1/2} \hat{\theta}_t. \)

(ii) Run the matrix regression

\[
\tilde{\Lambda}_t \text{ on } \tilde{\phi}_t, \quad t = 1, \ldots, T
\]

and save the residuals, say \( \tilde{\lambda}_t; \)

(iii) Run the regression

\[
I \text{ on } \tilde{\phi}_t^2 \tilde{\lambda}_t, \quad t = 1, \ldots, T
\]

and use \( TR^2_u = T - SSR \) (sum of squared residuals) as asymptotically \( \chi^2_Q \) under \( H_0, \) assuming that \( \tilde{\lambda}_t \) does not contain redundant indicators.

It must be emphasized that condition (ii.b), which is stated as a definition, requires that \( E[\phi_t(y_t, x_t, \theta_0)|x_t] \) be computable under the null hypothesis. This can impose additional restrictions on \( \phi_t \) that must be satisfied in order for Procedure 2.1 to be valid under \( H_0. \) If additional assumptions are used in forming \( \Phi_t(x_t, \theta_0) \) then the "implicit null hypothesis" includes more than just (2.9). But as shown in Wooldridge [28], \( \Phi_t \) is always computable under the relevant null hypothesis for conditional mean (hence conditional probability) and 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 \( \Phi_t(x_t, \theta_0) \) 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 \( \theta \) contains \( \alpha \) and any conditional variance parameters. \( \Phi_t(\theta_0) \) is easily seen to be \( \Phi_t(\theta_0) = -3 \nabla \alpha m_t(\alpha_0)V(y_t|x_t). \) Most tests for skewness in the literature impose homoskedasticity or some other conditional variance assumption under the null, and \( \Phi_t(\theta_0) \) 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, the choice of \( \Lambda_t \) is limited in this example by the form of \( \Phi_t. \) If \( \hat{\Lambda}_t \) is linearly related to \( \hat{\phi}_t \) then the modified indicator \( \hat{\Lambda}_t \) is simply zero. In a linear model with conditional homoskedasticity and regressors \( w_t, \hat{\phi}_t \) is proportional to \( w_t, \) and so \( \hat{\Lambda}_t \) cannot contain linear combinations of \( w_t. \) In
particular, if \( w \) contains unity then the choice \( \Lambda_t = 1 \) is unavailable, ruling out a standard test for unconditional skewness based on \( \sum_{i=1}^{T} \hat{u}_i^2 \). An NTW test would allow more flexibility in the choice of \( \Lambda_t \) but would generally be less robust.

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 \( \theta = (\alpha', \kappa)' \). Strictly speaking, the regularity conditions for Theorem 2.1 are not satisfied for this choice of \( \phi_t(\theta) \). Nevertheless, \( \phi_t(\theta) \) is differentiable almost surely in \( \alpha \) under the usual assumptions imposed in these contexts. At the cost of complicating the analysis, the conditions of Theorem 2.1 could be relaxed to handle this case. The quasigradient with respect to \( \alpha \) is \( \nabla_{\alpha} \phi_t(\theta) = \{1[y_t - m_t(\alpha_o) > 0] - 1[y_t - m_t(\alpha_o) \leq 0]\} \nabla_{\alpha} m_t(\alpha_o) \). Under conditional symmetry of the distribution of \( y_t \) given \( x_t \), \( E[1[y_t - m_t(\alpha_o) > 0]|x_t] = E[1[y_t - m_t(\alpha_o) \leq 0]|x_t] \), so that \( E[\nabla_{\alpha} \phi_t(\theta_o)|x_t] = 0 \).

Also, \( E[\nabla_{\alpha} \phi_t(\theta_o)|x_t] = 1 \), and \( \hat{\Phi}(x_t, \theta_o) \) 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 \( \alpha_o \).

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 \( \theta_o \) and \( \pi^{2} \) is a fairly weak requirement, and allows relatively simple specification tests when \( \theta_o \) (as well as \( \pi^{2} \)) has been estimated by an inefficient procedure (under classical assumptions). This is useful for certain tests in the presence of nonnested hypotheses as well as for conditional variance tests when the null imposes a nonconstant variance function.

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

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

\[
(iii) \quad T^{-1/2} \sum_{t=1}^{T} \hat{\phi}_t \hat{C}_t \hat{\phi}_t = o_p(1),
\]

then

\[
\bar{\xi}_T - \bar{\xi}_T = o_p(1). \tag{2.21}
\]

When (iii) holds, \( \bar{\xi}_T \) and \( \bar{\xi}_T \) are asymptotically equivalent under \( H_0 \). Condition (iii) is usefully interpreted as the sample covariance between \( \{\hat{C}_t^{1/2} \hat{\phi}_t : t = 1, \ldots, T\} \) and \( \{\hat{C}_t^{1/2} \hat{\phi}_t : t = 1, \ldots, T\} \) being asymptotically zero. It is trivially satisfied if
\[ \sum_{i=1}^{\tau} \Phi_i(\theta)C_i(\theta, \hat{\theta}_T)\phi_i(\theta) = 0 \]  

(2.22)

is the first-order condition that defines \( \hat{\theta}_T \). 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 [28]) 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 \( \hat{\theta}_T \) 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.

Lemma 2.2 does not directly describe the local behavior of \( \hat{\xi}_T \) when (iii) fails to hold under local alternatives, but viewed from a slightly different perspective it provides useful insight. Note that Theorem 2.1 implies that the quadratic form in \( \hat{\xi}_T \) 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 against which \( \hat{\xi}_T \) has power when (iii) does not hold. Fortunately, it is frequently the case that \( \hat{\xi}_T \) is asymptotically equivalent to some 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 \( \hat{\xi}_T \), it is useful to be somewhat more explicit about the nature of the local alternatives. Let \( \theta_T^* \) and \( \pi_T^* \) be nonstochastic sequences such that \( \sqrt{T}(\theta_T^* - \theta_0) = O(1) \) and \( \sqrt{T}(\pi_T^* - \pi_0^*) = O(1) \). \( \{\theta_T^*: T = 1, 2, \ldots\} \) indexes the sequence of local alternatives, but, as with \( \theta_0 \) under \( H_0 \), \( \theta_T^* \) need not uniquely index the nonnull probability measure. \( \pi_T^* \) is the plim of the estimator \( \hat{\pi}_T \) under the sequence of local alternatives \( \{H_{T_1}: T = 1, 2, \ldots\} \). Assume that the conditions of Theorem 2.1 are supplemented with conditions of the form

\[ E_{\theta_T^*} \left[ T^{-1} \sum_{t=1}^{T} G_t(y_t, x_t, \theta_T^*, \pi_T^*) \right] - E_{\theta_0} \left[ T^{-1} \sum_{t=1}^{T} G_t(y_t, x_t, \theta_0, \pi_0^*) \right] \rightarrow 0 \]

as \( T \rightarrow \infty \) 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 \( \{H_{T_1}\} \),
\[ \hat{\xi}_T = T^{-1/2} \sum_{i=1}^{T} (\Lambda_i^* - \Phi_i^* B_T^*)' C_i^* \phi_i^* + o_p(1) \] (2.23)

where

\[ B_T^* = \left[ \sum_{i=1}^{T} E(\Phi_i^* C_i^* \phi_i^*) \right]^{-1} \sum_{i=1}^{T} E(\Phi_i^* C_i^* \Lambda_i^*) \]

and values with a "*" superscript are evaluated at \( \theta_T^* \) or \( (\theta_T^*, \pi_T^*) \). Equation (2.23) is the extension of (2.18) to local alternatives and implies that the local limiting distribution of \( \hat{\xi}_T \) is the same when \( \hat{\theta}_T \) and \( \hat{\pi}_T \) are replaced by their plims \( \theta_T^* \) and \( \pi_T^* \), provided that \( \sqrt{T}(\hat{\theta}_T - \theta_T^*) = O_p(1) \) and \( \sqrt{T}(\hat{\pi}_T - \pi_T^*) = O_p(1) \) under \( \{H_T\} \). Therefore, if \( (\hat{\theta}_T, \hat{\pi}_T) \) and \( (\hat{\theta}_T^2, \hat{\pi}_T^2) \) are both \( \sqrt{T} \)-consistent estimators of \( (\theta_T^*, \pi_T^*) \) under \( \{H_T\} \) then

\[ \hat{\xi}_{T1} - \hat{\xi}_{T2} = o_p(1) \] (2.24)

under \( \{H_T\} \), where \( \hat{\xi}_{T1} \) is evaluated at \( \hat{\theta}_{T1}, \hat{\pi}_{T1} \) and \( \hat{\xi}_{T2} \) is evaluated at \( \hat{\theta}_{T2}, \hat{\pi}_{T2} \). If \( \hat{\theta}_{T2} \) and \( \hat{\pi}_{T2} \) are chosen to satisfy (iii), that is,

\[ T^{-1/2} \sum_{i=1}^{T} \Phi_i(\hat{\theta}_{T2})' C_i(\hat{\theta}_{T2}, \hat{\pi}_{T2}) \phi_i(\hat{\theta}_{T2}) = o_p(1), \] (2.25)

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

\[ \hat{\xi}_{T1} - \hat{\xi}_{T2} = o_p(1) \] (2.26)

under \( H_0 \) and local alternatives.

Conclusion (2.26) is simple yet powerful. It implies that for any \( \sqrt{T} \)-consistent estimator \( \hat{\theta}_{T1} \), \( \hat{\xi}_{T1} \) is asymptotically equivalent to \( \hat{\xi}_{T2} \) because \( \hat{\xi}_{T2} \) has been evaluated at an estimator \( \hat{\theta}_{T2} \) that satisfies the asymptotic first-order condition (2.22). Whenever such an estimator is available the interpretation of \( \hat{\xi}_T \) is straightforward: \( \hat{\xi}_T \) 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 (2.22). It does not matter which estimator is used in computing \( \hat{\xi}_T \), provided that it is \( \sqrt{T} \)-consistent. Thus, the interpretation of \( \hat{\xi}_T \) does not depend on the estimator used in computing it, but only on the first-order condition (2.22). In many situations there is an estimator that satisfies (2.22) and solves a well-known problem; interpreting \( \hat{\xi}_T \) 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. A case where an estimator satisfying (2.22)
does not have particularly desirable properties arises in 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 cannot be expected to have any optimality properties.

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 (iii) of Procedure 2.1 uses unity as the regressand, 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 [5] for a static regression model and of Bollerslev and Wooldridge [1] for an AR-GARCH (autoregressive-generalized autoregressive conditional heteroskedasticity) 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 (ii) of Procedure 2.1 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 $\{m_t(x_t,\alpha) : \alpha \in A\}, A \subset \mathbb{R}^p$, be a parametric family for the conditional expectation of $y_t$ given $x_t$. The null hypothesis is

$$H_0: E(y_t|x_t) = m_t(x_t,\alpha_0), \quad \text{some } \alpha_0 \in A, \quad i = 1,2,\ldots$$

(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 $T^{1/2}(\hat{\gamma}_T - \gamma_0^*) = O_p(1)$, where $\{\gamma_0^*\} \subset \Gamma$. It is not assumed that $h_t(x_t,\gamma_0)$ is proportional to $V(y_t|x_t)$ for some $\gamma_0 \in \Gamma$. 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 $\alpha_0$. No matter which estimator of $\alpha_0$ is used, the tests are motivated by the WNLS first-order condition

$$\sum_{i=1}^T \nabla_{\alpha} m_t(\alpha) \left( \frac{y_t - m_t(\alpha)}{h_t(\hat{\gamma}_T)} \right) = 0.$$ 

(3.2)

A general class of diagnostics is obtained by replacing $\nabla_{\alpha} m_t(\alpha)$ in (3.2) with $1 \times Q$ vector of misspecification indicators evaluated at the estimators:

$$\sum_{i=1}^T \lambda_t(\hat{\alpha}_T, \hat{\gamma}_T) \left( \frac{y_t - m_t(\hat{\alpha}_T)}{h_t(\hat{\gamma}_T)} \right),$$

(3.3)
where \( \hat{\gamma}_T \) 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 \( \Phi_t(x_t, \theta_0) \) requires no auxiliary assumptions under \( H_0: \Phi_t(x_t, \theta_0) = -\nabla_x m_t(x_t, \alpha_0). \)

The usual LM-type statistic, which is \( TR^2 \) from the regression

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

(i) Let \( \hat{\alpha}_T \) be a \( \sqrt{T} \)-consistent estimator of \( \alpha_0 \). Compute the residuals \( \hat{u}_t, \) the gradient \( \nabla_x m_t(\hat{\alpha}_T) \), and the indicator \( \lambda_t(\hat{\alpha}_T, \hat{\gamma}_T) \). Define \( \hat{u}_t = \hat{h}_t^{-1/2}\hat{u}_t, \nabla_x \hat{m}_t = \hat{h}_t^{-1/2}\nabla_x \hat{m}_t, \) and \( \hat{\lambda}_t = \hat{h}_t^{-1/2}\hat{\lambda}_t. \)

(ii) Regress \( \hat{\lambda}_t \) on \( \nabla_x \hat{m}_t \) and save the \( 1 \times Q \) residuals, say \( \hat{\lambda}_t. \)

(iii) Regress 1 on \( \hat{\lambda}_t \hat{\lambda}_t \) and use \( TR^2 = T - \text{SSR from this regression as asymptotically } \chi^2. \)

This procedure with \( \hat{h}_t \equiv 1 \) was first proposed by Davidson and MacKinnon [5] in the context of a nonlinear regression model within independent errors and unconditional heteroskedasticity. It was also suggested by Wooldridge [27] 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 the Davidson-MacKinnon [4] test in the presence of nonnested alternatives, none of which requires 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 detail in Wooldridge [28].

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 any \( \sqrt{T} \)-consistent estimator is asymptotically equivalent to the LM statistic based on (3.3) when (3.3) is evaluated at the WNLS estimator, \( h_t(\gamma_0) \) is proportional to \( V(y_t|x_t) \), and \( \hat{\gamma}_T \) is a \( \sqrt{T} \)-consistent estimator of \( \gamma_0 \). For efficiency reasons it is prudent to put some thought into the choice of \( h_t \).

**Example 3.2.** 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 \( \sigma^2 \) is absorbed into \( \gamma \). The null hypothesis is

\[
H_0: E(y_t|x_t) = m_t(x_t, \alpha_o), \quad V(y_t|x_t) = v_t(x_t, \gamma_o), \quad (3.4)
\]

\( \alpha_o \in \mathcal{A}, \quad \gamma_o \in \Gamma, \quad t = 1, 2, \ldots. \)
Let $\hat{\alpha}_T$ be the WNLS or some other $\sqrt{T}$-consistent estimator of $\alpha_o$, and let $\hat{\gamma}_T$ be any $\sqrt{T}$-consistent estimator of $\gamma_o$. Let $\lambda_r(x_t, \theta, \pi)$ be a $1 \times Q$ vector of indicators where $\theta = (\alpha', \gamma')'$. Most tests for variances can be derived from a statistic of the form

$$T^{-1} \sum_{t=1}^{T} \hat{\lambda}_t' \left( \frac{\hat{u}_t^2 - \hat{v}_t}{\hat{b}_t^2} \right). \quad (3.5)$$

Choosing $\lambda(x_t, \theta, \pi)$ to be the nonconstant, nonredundant elements of $\text{vech}[\nabla_\alpha m_\tau(\alpha)'\nabla_\alpha m_\tau(\alpha)]$ leads to the White [23] information matrix test in the context of quasi-maximum likelihood estimation in a linear exponential family (see Wooldridge [28]). When $u_t(\gamma) = \sigma^2$ choosing $\lambda_r(x_t, \theta, \pi) = w_t$, where $w_t$ 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 [2]). Setting $\lambda_r(x_t, \theta, \pi) = [u_{t-1}^2(\alpha), \ldots, u_{t-Q}(\alpha)]$ gives Engle's [7] test for ARCH($Q$) under a null of conditional homoskedasticity.

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

The choice $C_r(\theta, \pi) = 1/u_t^2(\gamma)$ in (3.5) is motivated by the structure of the score of the normal log-likelihood with mean function $m_\tau(\alpha)$ and variance function $u_t(\gamma)$. In particular, the scaling $1/b_t^2$ appears in the variance tests of Godfrey [9] and Breusch and Pagan [3]. The standard LM statistic in this context is $TR^2_\tau$ from the regression

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

In addition to (3.4) this test imposes

$$E[(u_t^2)^4|x_t] = \kappa_5^2 [u_t(x_t, \gamma_o)]^2, \quad \text{some } \kappa_5^2 > 0 \quad (3.7)$$

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 $\gamma_o$ 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 $\gamma_o$ but 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 $\gamma_o$:

(i) Let $\hat{\alpha}_T$ be a $\sqrt{T}$-consistent estimator of $\alpha_o$, and let $\hat{\gamma}_T$ be a $\sqrt{T}$-consistent estimator of $\gamma_o$. Compute the residuals $\hat{u}_t$, the gradient $\nabla_\gamma \hat{u}_t(\hat{\gamma}_T)$, and the indicator $\lambda_r(\hat{\theta}_T, \hat{\pi}_T)$. Define $\hat{\phi}_t = (\hat{u}_t^2 - \hat{v}_t)/\hat{b}_t = \hat{u}_t^2/\hat{b}_t - 1$, $\nabla_\gamma \hat{\phi}_t = \nabla_\gamma \hat{u}_t/\hat{b}_t$, and $\hat{\lambda}_t = \hat{\lambda}_t/\hat{b}_t$. 


(ii) Regress $\tilde{\lambda}_t$ on $\tilde{\psi}_t, \tilde{\nu}_t$ and save the $1 \times Q$ residuals, say $\tilde{\lambda}_t$.

(iii) Regress 1 on $\tilde{\phi}_t, \tilde{\lambda}_t$ and use $TR_0^2 = T - SSR$ from this regression as asymptotically $\chi^2_0$ under $H_0$.

Interestingly, when $u_t(x_t, \gamma) = \sigma_t^2$, so that the null is conditional homoskedasticity, the regression in (ii) simply demeans the indicators. Given $\tilde{\nu}_t^2, \tilde{\theta}_t^2$, and a choice for $\tilde{\lambda}_t$, the $\chi^2_0$ statistic is obtained as $TR_0^2$ from the regression

$$1 \text{ on } (\tilde{\nu}_t^2 - \tilde{\theta}_t^2)(\tilde{\lambda}_t - \tilde{\lambda}_T), \quad t = 1, \ldots, T$$

(3.8)

where $\bar{\lambda}_T = T^{-1} \sum_{t=1}^T \tilde{\lambda}_t$. This procedure is asymptotically equivalent to the traditional regression form (2.8) under the additional assumption that $E[u_t^4(\sigma_o)|x_t]$ is constant. Note that (3.6) and (2.8) usually yield different test statistics that are not asymptotically equivalent under $H_0$. The demeaning of the indicators is only a slight modification, but it yields an asymptotically chi-square distributed statistic without the additional assumption of constant fourth moment for $u_t^2$. In the case of the White test in a linear time series model, the demeaning of the indicators yields a statistic that is asymptotically equivalent to Hsieh's [12] 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, $TR_0^2$ from the regression in (3.8) is asymptotically equivalent to $TR_0^2$ from the regression

$$1 \text{ on } (\tilde{\nu}_t^2 - \tilde{\theta}_t^2)(\tilde{\nu}_{t-1}^2 - \tilde{\theta}_t^2), \ldots, (\tilde{\nu}_t^2 - \tilde{\theta}_t^2)(\tilde{\nu}_{t-Q}^2 - \tilde{\theta}_t^2), \quad t = Q + 1, \ldots, T.$$  

(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 $u_t^2$. Nevertheless, it is asymptotically equivalent to the usual ARCH test under normality.

Example 3.3. 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

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

(3.10)

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

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

(3.11)

Take the null hypothesis to be

$$H_0: \delta_o = r(\theta_o), \quad \text{for some } \theta_o \in \Theta \subset \mathbb{R}^P$$

(3.12)

where $P < M$ and $r$ is continuously differentiable on int($\Theta$). Let $m_t(\theta) = \mu_t[r(\theta)]$ and $v_t(\theta) = \omega_t[r(\theta)]$ be the constrained mean and variance func-
tions. QMLE is carried out under the null hypothesis. Let \( \hat{\delta}_T \) be the estimator of \( \delta_0 \) under \( H_0 \), and let \( \tilde{\delta}_T = r(\hat{\delta}_T) \) be the constrained estimator of \( \delta_0 \). \( \nabla_0 \hat{m}_t \) and \( \nabla_0 \hat{\delta}_t \) are the 1 \times P gradients of \( m_t \) and \( \delta_t \) under \( H_0 \). Note that \( \hat{\omega}_t = \hat{\delta}_t \) 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 \( \tilde{\delta}_T \). The transpose of the score is

\[
\begin{align*}
\begin{pmatrix} s_t(\delta)^\prime \end{pmatrix} &= \nabla_0 \mu_t(\delta)^\prime u_t(\delta)/\omega_t(\delta) + \nabla_0 \omega_t(\delta)^\prime [u_t^2(\delta) - \omega_t(\delta)]/2\omega_t^2(\delta) \\
&= \begin{pmatrix} \nabla_0 \mu_t(\delta)^\prime \\ \nabla_0 \omega_t(\delta)^\prime \end{pmatrix} \begin{pmatrix} 1/\omega_t(\delta) & 0 \\ 0 & 1/[2\omega_t(\delta)^2] \end{pmatrix} \begin{pmatrix} u_t(\delta) \\ u_t^2(\delta) - \omega_t(\delta) \end{pmatrix}.
\end{align*}
\] (3.13)

Evaluating \( s_t \) at \( r(\theta) \) gives

\[
\begin{align*}
\begin{pmatrix} s_t[r(\theta)]^\prime \end{pmatrix} &= \Lambda_t(\theta)^\prime C_t(\theta) \phi_t(\theta) \\
&= \begin{pmatrix} \nabla_0 \mu_t[r(\theta)]^\prime \\ \nabla_0 \omega_t[r(\theta)]^\prime \end{pmatrix} \begin{pmatrix} 1/\omega_t(r(\theta)) & 0 \\ 0 & 1/[2\omega_t(r(\theta))^2] \end{pmatrix} \begin{pmatrix} u_t[r(\theta)] \\ u_t^2[r(\theta)] - u_t[r(\theta)] \end{pmatrix}.
\end{align*}
\] (3.14)

Under \( H_0 \) and the assumption of conditional normality, \( TR_u^2 \) from the regression

\[
1 \text{ on } s_t, \quad t = 1, \ldots, T
\] (3.17)

is asymptotically \( \chi_Q^2 \), where \( Q = M - P \) is the number of restrictions under \( H_0 \). Unfortunately, this procedure is invalid under nonnormality, and it has no systematic power for detecting nonnormality. Theorem 2.1 suggests a robust form of the test. In this case,

\[
\begin{align*}
\dot{\hat{\phi}}_t &= \begin{pmatrix} \dot{\hat{\mu}}_t \\ \dot{\sqrt{\hat{\delta}}} \end{pmatrix}; \quad \dot{\hat{\Lambda}}_t = \begin{pmatrix} \nabla_0 \dot{\hat{\mu}}_t \\ \nabla_0 \dot{\sqrt{\hat{\delta}}} \end{pmatrix} \\
\dot{\hat{C}}_t &= \begin{pmatrix} 1/\dot{\sqrt{\hat{\delta}}} & 0 \\ 0 & 1/(2\dot{\sqrt{\hat{\delta}}}) \end{pmatrix} \\
\dot{\hat{\Phi}}_t &= \begin{pmatrix} \nabla_0 \dot{\hat{m}}_t/\dot{\sqrt{\hat{\delta}}} \\ \nabla_0 \dot{\hat{\delta}}/\sqrt{2\dot{\sqrt{\hat{\delta}}}} \end{pmatrix},
\end{align*}
\]

where \( \dot{\hat{\mu}}_t = y_t - m_t(\hat{\delta}_T) \). The transformed quantities are

\[
\begin{align*}
\tilde{\hat{\phi}}_t &= \begin{pmatrix} \dot{\hat{\mu}}_t/\sqrt{\dot{\sqrt{\hat{\delta}}}} \\ (\dot{\sqrt{\hat{\delta}}} - \dot{\hat{\delta}})/(\sqrt{2}\dot{\sqrt{\hat{\delta}}}) \end{pmatrix}; \quad \tilde{\hat{\Lambda}}_t = \begin{pmatrix} \nabla_0 \dot{\hat{\mu}}_t/\dot{\sqrt{\hat{\delta}}} \\ \nabla_0 \dot{\sqrt{\hat{\delta}}}/\sqrt{2\dot{\sqrt{\hat{\delta}}}} \end{pmatrix} \\
\tilde{\hat{\Phi}}_t &= \begin{pmatrix} \nabla_0 \dot{\hat{m}}_t/\sqrt{\dot{\sqrt{\hat{\delta}}}} \\ \nabla_0 \dot{\hat{\delta}}/\sqrt{2\dot{\sqrt{\hat{\delta}}}} \end{pmatrix}.
\end{align*}
\]

The robust test statistic is obtained by first running the regression

\[
\tilde{\Lambda}_t \text{ on } \tilde{\hat{\Phi}}_t, \quad t = 1, \ldots, T
\] (3.18)
and saving the matrix residuals \( \{ \tilde{A}_t: t = 1, \ldots, T \} \). Then run the regression
\[
1 \text{ on } \tilde{\phi}_t \tilde{A}_t, \quad t = 1, \ldots, T
\]
and use \( TR_u^2 = T - \text{SSR} \) as asymptotically \( \chi^2_0 \) under \( H_0 \). Note that the regression in (3.19) contains perfect multicollinearity since \( \tilde{A}_t \nabla \theta(\hat{\theta}_T) = 0, \)
where \( \nabla \theta(\hat{\theta}) \) is the \( M \times P \) gradient of \( \theta \). Many regression packages nevertheless compute an SSR; 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
\[
\sum_{t=1}^T \Phi_t(\hat{\theta}_T)'C_t(\hat{\theta}_T)\Phi_t(\hat{\theta}_T) = 0,
\]
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.** Consider a single equation from a linear simultaneous system of equations:
\[
y_{1t} = y_{12} \alpha_{o1} + w_{t1} \gamma_{o1} + u_{1t}
\]
where \( y_{12} \) is a \( 1 \times J_2 \) vector and \( w_{t1} \) is a \( 1 \times K_1 \) subvector of \( x_t \). Suppose that the first hypothesis of interest is
\[
H_0: E(u_{1t}|x_t) = 0,
\]
that is, the errors are unpredictable given the predetermined variables. Under \( H_0 \), suppose that \( w_t \) is a \( 1 \times K \) optimal set of instruments for \( y_{12} \) in the sense that for some \( K \times J_2 \) matrix \( \Delta_o \),
\[
E(y_{12}|x_t) = w_t \Delta_o.
\]
Then, in the notation of Procedure 2.1, \( \theta \equiv (\alpha_1, \gamma_1, \delta), \phi_t(y_t, x_t, \theta) = y_{1t} - y_{12} \alpha_1 - w_{t1} \gamma_1, \) and \( \nabla \phi_t(\theta) = (-y_{12}, w_{t1}, 0) \), where \( \delta \equiv \text{vec}(\Delta) \). Using (3.21),
\[
\Phi_t(x_t, \theta_o) = E[\nabla \theta \phi_t(y_t, x_t, \theta_o)|x_t] = -(w_t \Delta_o, w_{t1}, 0).
\]
If \( \lambda_t(x_t, \theta, \pi) \) is a \( 1 \times Q \) vector of misspecification indicators that depend on predetermined variables, then Procedure 2.1 can be applied immediately. First, \( \hat{\phi}_t \equiv \hat{u}_{1t} = y_{1t} - y_{12} \hat{\alpha}_{T1} - x_{t1} \hat{\gamma}_{T1} \) are the two stage least squares (2SLS) residuals using instruments \( w_t \) (the usual identification conditions are assumed to hold). The estimate of \( \Phi_t \) is
\[
\hat{\Phi}_t = -(\hat{y}_{12}, w_{t1}, 0),
\]
where \( \hat{y}_{12} \) are the fitted values from the first-stage regression of \( y_{12} \) on \( w_t \). From Theorem 2.1, the following procedure is valid under (3.20) and (3.21):

(i) Compute the 2SLS residuals \( \hat{u}_{1t} \), the first-stage fitted values \( \hat{y}_{12} \), and the misspecification indicator \( \hat{\lambda}_t \).
(ii) Run the regression
\[ \hat{\lambda}_t \text{ on } \hat{y}_{t2}, \, w_{t1} \]
and save the \( 1 \times Q \) residuals, \( \hat{\lambda}_t \).

(iii) Use \( TR^2_\theta = T - \text{SSR} \) from the regression
\[ 1 \text{ on } \hat{\mu}_{t1} \hat{\lambda}_t, \quad t = 1, \ldots, T \]
as asymptotically \( \chi^2_Q \) under \( H_0 \).

This procedure is valid under conditional or unconditional heteroskedasticity and for a variety of misspecification indicators. A heteroskedasticity-robust test for AR(\( Q \)) serial correlation is obtained by setting \( \hat{\lambda}_t = (\hat{\mu}_{t-1,1}, \ldots, \hat{\mu}_{t-Q,1}) \). Note that the restriction (3.21) simply requires that all relevant instruments have been used under \( H_0 \); it does not matter what the optimal instruments are under the alternative. Incidentally, if conditional homoskedasticity \( E(u^2_{t1} \mid x_t) = \sigma^2_{t1} \) is maintained under \( H_0 \) in addition to (3.20) and (3.21), then a valid test statistic is \( TR^2_\theta \) from
\[ \hat{\mu}_{t1} \text{ on } \hat{y}_{t2}, \, w_{t1}, \, \hat{\lambda}_t. \]

Now suppose that one wishes to test the hypothesis
\[ H_0: E(u^2_{t1} \mid x_t) = \sigma^2_{t1}, \] (3.22)
maintaining that (3.20) holds. Let \( \phi_t(\theta) = u^2_{t1}(\alpha_1, \gamma_1) - \sigma^2_t \) and note that
\[ E[\nabla_{\alpha_1} \phi_t(\theta_0) \mid x_t] = -2E(y_{t2} u_{t1} \mid x_t). \]

To make things simple, assume in addition that
\[ E(y_{t2} u_{t1} \mid x_t) = \psi_{01} \] (3.23)
for some \( J_2 \times 1 \) vector \( \psi_{01} \); this essentially imposes homoskedasticity on the system covariance matrix under \( H_0 \). The appropriate choice of \( \theta \) is then \( (\alpha, \gamma_1, \sigma^2_t, \psi_1) \), and under (3.20) and (3.23),
\[ \Phi_t(x_t, \theta_0) = E[\nabla_\theta \phi_t(\theta_0) \mid x_t] = -2(\psi_{01}, 0, 1, 0). \]

Because \( \Phi_t(x_t, \theta_0) \) is constant, the purging step (ii) of Procedure 2.1 simply demeans the misspecification indicator \( \lambda_t(x_t, \hat{\theta}_T, \hat{\pi}_T) \), as in the case of testing for heteroskedasticity in a model of a conditional expectation. If \( \hat{\sigma}^2_{t1} \) is the usual 2SLS estimator of \( \sigma^2_{t1} \) then the test statistic is \( T - \text{SSR} \) from the regression
\[ 1 \text{ on } (\hat{\mu}^2_{t1} - \hat{\sigma}^2_{t1})(\hat{\lambda}_t - \hat{\lambda}_T). \] (3.24)

A test for ARCH(\( Q \)) is obtained by choosing \( \hat{\lambda}_t = (\hat{\mu}_{t-1,1}, \ldots, \hat{\mu}_{t-Q,1}) \). If homokurtosis is maintained in addition to (3.20), (3.22), and (3.23), a valid statistic is \( TR^2_\theta \) from the regression
\[ \hat{\mu}^2_{t1} \text{ on } 1, \, \hat{\lambda}_t. \] (3.25)
The additional restriction (3.23) arises because the endogenous variables \( y_{t2} \) appear on the right-hand side of the original model; (3.23) embodies the information about \( y_{t2} \) that justifies a test based on (3.24) or (3.25).

4. CONCLUSION

This paper has developed a general class of regression-based specification tests for (possibly) dynamic multivariate models which, in many 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.

The possibility of generating simple test statistics when \( T^{1/2}(\hat{\theta} - \theta_0) \) has a complicated limiting distribution should be useful in several situations. One interesting application is choosing between log-linear and linear-linear specifications. In this case, both models can be estimated by OLS, and the log-linear estimates can be transformed into estimates of \( E(y_t|x_t) \) to be compared with the linear model. The Davidson–MacKinnon [4] test for non-nested hypotheses can be based on estimates of \( E(y_t|x_t) \) under each model. The distribution of the estimators, which is somewhat complicated, plays no role if Example 3.1 is used.

Example 3.4 shows that Theorem 2.1 applies directly to linear simultaneous equations models. The scope of applications to nonlinear simultaneous equations models is limited by one's ability to compute \( \Phi_t(x_t, \theta_0) = E[\nabla_\theta \phi(y_t, x_t, \theta_0)|x_t]. \) This is exactly the problem of computing the optimal instrumental variables for nonlinear SEMs. Estimating \( \Phi_t(x_t, \theta_0) \) by a nonparametric technique as in Newey [16] is one possible solution to this problem.

Theorem 2.1 can be extended to certain unit root time series models. The initial purging of \( \hat{\Phi}_t^{1/2} \) from \( \hat{\Phi}_t^{1/2} \hat{\lambda}_t \) can produce indicators \( \hat{\lambda}_t \) 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. & 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. & A.R. Pagan. A simple test for heteroskedasticity and random coefficient variation. *Econometrica* 47 (1979): 1287–1294.
3. Breusch, T.S. & A.R. Pagan. The Lagrange Multiplier statistic and its application to model specification in econometrics. *Review of Economic Studies* 47 (1980): 239–253.
4. Davidson, R. & J.G. MacKinnon. Several tests of model specification in the presence of alternative hypotheses. *Econometrica* 49 (1981): 781–793.
5. Davidson, R. & J.G. MacKinnon. Heteroskedasticity-robust tests in regression directions. *Annales de l'INSEE* 59/60 (1985): 183–218.
6. Domowitz, I. & H. White. Misspecified models with dependent observations. *Journal of Econometrics* 20 (1982): 35–58.
7. Engle, R.F. Autoregressive conditional heteroskedasticity with estimates of United Kingdom inflation. *Econometrica* 50 (1982): 987–1008.
8. Engle, R.F. Wald, Likelihood Ratio, and Lagrange Multiplier tests in econometrics. In Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Vol. II, Amsterdam: North Holland (1984): 775–826.
9. Godfery, L.G. Testing for multiplicative heteroskedasticity. *Journal of Econometrics* 8 (1978): 227–236.
10. Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* 50 (1982): 1029–1054.
11. Hausman, J.A. Specification tests in econometrics. *Econometrica* 46 (1978): 1251–1271.
12. Hsieh, D.A. A heteroskedasticity-consistent covariance matrix estimator for time series regressions. *Journal of Econometrics* 22 (1983): 281–290.
13. Kennan, J. & G.R. Neuman. Why does the information matrix test reject too often? University of Iowa Department of Economics Working Paper No. 88-4, 1988.
14. Koekker, R. A note on studentizing a test for heteroskedasticity. *Journal of Econometrics* 17 (1981): 107–112.
15. Newey, W.K. Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53 (1985): 1047–1070.
16. Newey, W.K. Efficient estimation of semiparametric models via moment restrictions. Manuscript. Princeton University.
17. Neyman. J. Optimal asymptotic tests of composite statistical hypotheses. In U. Grenander (ed.), *Probability and Statistics*, Stockholm: Almquist and Wiksell (1959): 213–234.
18. Pagan, A.R. & A.D. Hall. Diagnostic tests as residual analysis. *Econometric Reviews* 2 (1983): 159–218.
19. Pagan, A.R., P.K. Trivedi, & A.D. Hall. Assessing the variability of inflation. *Review of Economic Studies* 50 (1983): 585–596.
20. Tauchen, G. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30 (1985): 415–443.
21. White, H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (1980): 817–838.
22. White, H. Nonlinear regression on cross section data. *Econometrica* 48 (1980): 721–746.
23. White, H. Maximum likelihood estimation of misspecified models. *Econometrica* 50 (1982): 1–26.
24. White, H. *Asymptotic Theory for econometricians*. New York: Academic Press, 1984.
25. 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.
26. Wooldridge, J.M. Asymptotic properties of econometric estimators. UCSD Ph.D. dissertation, 1986.
27. Wooldridge, J.M. A regression-based Lagrange Multiplier statistic that is robust in the presence of heteroskedasticity. MIT Department of Economics Working Paper No. 478, 1987.
28. Wooldridge, J.M. Specification testing and quasi-maximum likelihood estimation. MIT Department of Economics Working Paper No. 479, 1987.
29. Wooldridge, J.M. A C(α) Version of the Newey-Tauchen-White Test. Manuscript, MIT Department of Economics, 1989.
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 \( \Theta \) and \( \Theta \) 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) \( \{ Q_T(\theta) : \theta \in \Theta, T = 1, 2, \ldots \} \) is continuous on \( \Theta \) uniformly in \( T \). Let \( \theta_T \) be a sequence of random vectors such that \( \theta_T \to \theta^* \overset{P}{\to} 0 \) where \( \{ \theta^*_T \} \subset \Theta \). Then \( Q_T(w_T, \theta_T) - \bar{Q}_T(\theta^*_T) \overset{P}{\to} 0. \)

**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 \( \Theta \) and \( \Theta \) 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

(i) \( \sup_{\theta \in \Theta} \left| T^{-1} \sum_{t=1}^T q_t(y_t, x_t, \theta) - E[q_t(y_t, x_t, \theta)] \right| \overset{P}{\to} 0; \)

(ii) \( \left\{ T^{-1} \sum_{t=1}^T E[q_t(y_t, x_t, \theta)] : \theta \in \Theta, T = 1, 2, \ldots \right\} \) is \( O(1) \) and continuous on \( \Theta \) 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 \( \theta \) then, by convention, \( \nabla_\theta a(\theta) \) is the \( L \times P \) matrix \( \nabla_\theta[a(\theta)^\prime] \). If \( A(\theta) \) is a \( Q \times L \) matrix then the matrix \( \nabla_\theta A(\theta) \) is the \( LQ \times P \) matrix defined as

\[ \nabla_\theta A(\theta)^\prime = [\nabla_\theta A_1(\theta)^\prime | \ldots | \nabla_\theta A_Q(\theta)^\prime] \]

where \( A_j(\theta) \) is the \( j \)th row of \( A(\theta) \) and \( \nabla_\theta A_j(\theta) \) is the \( L \times P \) gradient of \( A_j(\theta) \) as defined above. Also, for any \( L \times 1 \) vector function \( \varphi \), define the second derivative of \( \varphi \) to be the \( LP \times P \) matrix

\[ \nabla_\theta^2 \varphi(\theta) = \nabla_\theta[\nabla_\theta \varphi(\theta)^\prime]. \]

Finally, define the parameter vector \( \delta = (\theta^*, \pi^*)^\prime. \)

**Conditions A.1**

(i) \( \Theta \subset \mathbb{R}^p \) and \( \Pi \subset \mathbb{R}^N \) are compact and have nonempty interiors.

(ii) \( \theta_0 \in \text{int}(\Theta), \{ \pi^*_T : T = 1, 2, \ldots \} \subset \text{int}(\Pi) \) uniformly in \( T \).

(iii) (a) \( \{ \phi_t(y_t, x_t, \theta) : \theta \in \Theta \} \) is a sequence of \( L \times 1 \) 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 $\Theta$ for all $y_t, x_t, \ell = 1, 2, \ldots$

(b) Define $\Phi_t(x_t, \theta_0) = E_{\theta_0} [\nabla_{\theta} \phi_t(y_t, x_t, \theta_0)|x_t]$ for all $\theta_0 \in \text{int}(\Theta)$. Assume that $\Phi_t(x_t, \cdot)$ is continuously differentiable on the interior of $\Theta$ for all $x_t, \ell = 1, 2, \ldots$

(c) $\{C_t(x_t, \eta); \delta \in \Delta\}$ is a sequence of $L \times L$ matrices satisfying the measurability requirements, $C_t(x_t, \delta)$ is symmetric and positive semidefinite for all $x_t$ and $\delta$, and $C_t(x_t, \cdot)$ is differentiable on int$(\Delta)$ for all $x_t, \ell = 1, 2, \ldots$

(d) $\{A_t(x_t, \delta); \delta \in \Delta\}$ is a sequence of $L \times Q$ matrices satisfying the measurability requirements, and $A_t(x_t, \cdot)$ is differentiable on int$(\Delta)$ for all $x_t, \ell = 1, 2, \ldots$

(iv) (a) \(T^{1/2}(\hat{\theta}_T - \theta_0) = O_p(1)\).

(b) $T^{1/2}(\hat{\pi}_T - \pi_0) = O_p(1)$.

(v) (a) $[\Phi_t(\theta)'C_t(\delta)\Phi_t(\theta)]$ and $[\Phi_t(\theta)'C_t(\delta)A_t(\delta)]$ satisfy the UWLLN and UC conditions.

(b) $T^{-1/2} \sum_{t=1}^{T} \Phi_t^0 C_t^0 \Phi_t^0 = O_p(1)$.

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

(b) $T^{-1/2} \sum_{t=1}^{T} \Phi_t^0 C_t^0 \Phi_t^0 = O_p(1)$.

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

(viii) (a) $\{\Xi_t^0 = T^{-1} \sum_{t=1}^{T} E[(\Lambda^0_t - \Phi^0_t B^0_T)'C_t^0 \phi_t^0 \phi_t^0' C_t^0 (\Lambda^0_t - \Phi^0_t B^0_T)]\}$ is uniformly p.d.

(b) $\Xi_t^{0,1/2} T^{-1/2} \sum_{t=1}^{T} (\Lambda_t^0 - \Phi^0_t B^0_T)'C_t^0 \phi_t^0 \xrightarrow{d} N(0, I_Q)$.

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

PROOF OF THEOREM 2.1. First, note that assumptions (i)-(vi) ensure existence of $B_T^0$ and imply that $\hat{B}_T - B_T^0 = o_p(1)$ by Lemma A.1. Therefore,

\[ \hat{\xi}_T = T^{-1/2} \sum_{t=1}^{T} (\Lambda_t - \Phi_t B_T^0)' \hat{C}_t \hat{\phi}_t - (\hat{B}_T - B_T^0)' T^{-1/2} \sum_{t=1}^{T} \hat{C}_t \hat{\phi}_t. \]  \hspace{1cm} (A1)

Consider the term post-multiplying $(\hat{B}_T - B_T^0)'$. A standard mean value expansion about $\delta_T$, assumption (vi.a), and Lemma A.1 yield

\[ T^{-1/2} \sum_{t=1}^{T} \hat{C}_t \hat{\phi}_t = T^{-1/2} \sum_{t=1}^{T} \Phi_t^0 C_t^0 \phi_t^0 \]  \hspace{1cm} (A2)

\[ + T^{-1} \sum_{t=1}^{T} [\Phi_t^0 C_t^0 \nabla_{\phi_t}(\theta) + (I_P \otimes \phi_t' C_t^0) \nabla_{\phi_t}(\theta)] T^{1/2}(\delta_T - \delta_0) \]

\[ + T^{-1} \sum_{t=1}^{T} [\Phi_t^0 (I_L \otimes \phi_t' C) \nabla_{\phi_t}(\theta)] T^{1/2}(\delta_T - \delta_0) + o_p(1). \]
The first term on the right-hand side of (A2) is $O_p(1)$ by (vi.b). By (vi.a) and (iv.a.b), the terms in lines two and three of (A2) are also $O_p(1)$. Therefore,

$$T^{-1/2} \sum_{t=1}^{T} \hat{\Phi}_t \hat{C}_t \hat{\phi}_t = O_p(1). \quad (A3)$$

Along with $\hat{B}_T - B_T^0 = o_p(1)$, this establishes that under $H_0$,

$$\tilde{\xi}_T = T^{-1/2} \sum_{t=1}^{T} (\hat{\Lambda}_t - \hat{\Phi}_t B_T^0)' \hat{C}_t \hat{\phi}_t + o_p(1). \quad (A4)$$

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

$$\tilde{\xi}_T = T^{-1/2} \sum_{t=1}^{T} (\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \Phi_t^0 \quad (A5)$$

$$+ T^{-1} \sum_{t=1}^{T} [(\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \nabla_0 \Phi_t^0 - B_T^0 (I_P \otimes \Phi_t^0 C_t^0) \nabla_0 \Phi_t^0] T^{1/2} (\hat{\delta}_T - \delta_T^0) + o_p(1). \quad (A6)$$

Consider the second line of (A5). It must be shown that the average appearing there is $o_p(1)$ under $H_0$. First, note that by definition of $\Phi_t^0$ and the law of iterated expectations,

$$E[(\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \nabla_0 \Phi_t^0] = E[E[(\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \nabla_0 \Phi_t^0 | x_i]] = E[(\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \Phi_t^0]. \quad (A7)$$

Therefore,

$$T^{-1} \sum_{t=1}^{T} E[(\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \nabla_0 \Phi_t^0] = T^{-1} \sum_{t=1}^{T} E[(\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \Phi_t^0] = 0 \quad (A8)$$

by definition of $B_T^0$. The regularity conditions imposed imply that each of the averages appearing in (A7) satisfy the WLLN. Therefore,

$$T^{-1} \sum_{t=1}^{T} (\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \nabla_0 \Phi_t^0 = o_p(1). \quad (A9)$$

Because $E(\Phi_t^0 | x_i) = 0$ under $H_0$, it is even easier to show that the remaining sample averages in (A5) are $o_p(1)$. Combined with $T^{1/2}(\hat{\delta}_T - \delta_T^0) = O_p(1)$ this establishes the first conclusion of the theorem:

$$\tilde{\xi}_T = T^{-1/2} \sum_{t=1}^{T} (\Lambda_t^0 - \Phi_t^0 B_T^0)' C_t^0 \Phi_t^0 + o_p(1). \quad (A10)$$

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

$$\tilde{\sigma}_T = T^{-1} \sum_{t=1}^{T} [(\hat{\Lambda}_t - \hat{\Phi}_t B_T)' \hat{C}_t \hat{\Phi}_t \hat{\phi}_t \hat{C}_t (\hat{\Lambda}_t - \hat{\Phi}_t B_T)] \quad (A11)$$
is a consistent estimator of $\mathcal{E}_\theta$. It is easy to see that

$$\tilde{\xi}_T \tilde{\xi}_T^{-1} \tilde{\xi}_T = T R_u^2, \quad (A11)$$

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

$$1 \text{ on } \tilde{\phi}_t \tilde{A}_t, \quad t = 1, \ldots, T, \quad (A12)$$

and $\tilde{\phi}_t$ and $\tilde{A}_t$ are as defined in the text. Because the dependent variable in regression (A12) is unity, $T R_u^2 = T - SSR$, where SSR is the sum of squared residuals from the regression (A12).