working paper
department
of economics

REGRESSION BASED LAGRANGE MULTIPLIER STATISTIC THAT IS ROBUST IN THE PRESENCE OF HETEROSKEDASTICITY

by

Jeffrey M. Wooldridge

No. 478 December 1987

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

A REGRESSION BASED LAGRANGE MULTIPLIER STATISTIC THAT IS ROBUST IN THE PRESENCE OF HETEROSKEDASTICITY

by

Jeffrey M. Wooldridge

No. 478 December 1987

A REGRESSION-BASED LAGRANGE MULTIPLIER STATISTIC THAT IS ROBUST IN THE PRESENCE OF HETEROSKEDASTICITY

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

May 1987
Revised: December 1987

ABSTRACT: This paper derives a form of the Lagrange Multiplier statistic for nonlinear regression models that does not assume conditional homoskedasticity under the null hypothesis. In addition to the initial nonlinear least squares estimation, computation of the statistic requires only two linear least squares regressions. The problems of computing tests of exclusion restrictions and tests for serial correlation in dynamic linear models with unknown heteroskedasticity are presented as simple applications.

1. Introduction

This paper modifies the Lagrange Multiplier (LM) (or efficient score) testing procedure to allow correct inference in the presence of heteroskedasticity of unknown form. The test statistic is computable from linear least squares regressions, and is applicable to both cross section and time series models that have been estimated by nonlinear least squares (NLLS). The cost of the heteroskedasticity-robust procedure is one linear least squares regression. Simple regression-based procedures for computing tests of exclusion restrictions and tests for serial correlation in dynamic linear models with conditional heteroskedasticity of unknown form are presented as applications.

2. The Robust Lagrange Multiplier Test

Let $\{(Y_t, Z_t): t=1,2,\ldots\}$ be a sequence of observable random variables with Y_t a scalar, Z_t a $1 \times L$ vector. For each $t \geq 1$, let $X_t \equiv (Z_t, Y_{t-1}, Z_{t-1}, \ldots, Y_1, Z_1)$ denote the $1 \times L+(t-1)(L+1)$ vector of predetermined variables. The purpose of the analysis is to test hypotheses about the conditional expectation of Y_t given the predetermined variables X_t, $E(Y_t|X_t)$. For cross section applications the observations would be assumed independent across t, so that $E(Y_t|X_t) = E(Y_t|Z_t)$; in this case, replace X_t by Z_t throughout. For time series data it is assumed that interest lies in testing hypotheses about the expectation conditional on current exogenous variables and all past information. Consequently, a correctly specified conditional expectation necessarily excludes the existence of serial correlation.

The starting point is a correctly specified parameterized version of the conditional mean:

$$E(Y_t|X_t) = m_t(X_t,\alpha_o,\beta_o) \qquad \alpha_o \in A,\ \beta_o \in B,\ t=1,2,... \tag{2.1}$$

where $A \subset \mathbb{R}^P$, $B \subset \mathbb{R}^Q$, and m_t is a known real-valued function defined on $\mathbb{R}^{L+(t-1)(L+1)} \times A \times B$. Let $\delta \equiv (\alpha',\beta')'$ and $\Delta \equiv A \times B$. Among other regularity conditions, it is assumed that A is compact and that $m_t(x_t,\cdot)$ is twice continuously differentiable on the interior of Δ for each $x_t \in \mathbb{R}^{L+(t-1)(L+1)}$.

For simplicity, assume that the hypothesis of interest can be expressed as

$$H_0\colon\ \beta_o = 0. \tag{2.2}$$

The LM statistic is based upon the sample covariance of the NLLS residuals obtained under the null hypothesis and the gradient of m_t with respect to β evaluated under the null hypothesis. More precisely, let $\hat{\alpha}_T$ be the NLS estimator of α_o under H_0, so that $\hat{\alpha}_T$ is a solution to

$$\min_{\alpha\in A} \sum_{t=1}^{T} (Y_t - m_t(X_t,\alpha,0))^2. \tag{2.3}$$

Define the residual function as $U_t(\alpha) \equiv Y_t - m_t(X_t,\alpha,0)$. Let $\hat{U}_t \equiv Y_t - m_t(X_t,\hat{\alpha}_T,0)$, $\nabla_\alpha\hat{m}_t \equiv \nabla_\alpha m_t(X_t,\hat{\alpha}_T,0)$ and $\nabla_\beta\hat{m}_t \equiv \nabla_\beta m_t(X_t,\hat{\alpha}_T,0)$. The true residuals under H_0 are $U_t^o \equiv U_t(\alpha_o)$. Note that under H_0, $\{U_t^o\colon t=1,2,...\}$ is a martingale difference sequence with respect to the σ-fields $\{\sigma(Y_t,X_t)\colon t=1,2,...\}$.

The LM statistic is a quadratic form in the Qx1 vector

$$\sum_{t=1}^{T} \nabla_\beta\hat{m}_t'\hat{U}_t. \tag{2.4}$$

If, in addition to H_0, it is assumed that $V(Y_t|X_t) = \sigma_o^2$ for some $\sigma_o^2 > 0$, then asymptotically equivalent versions of the LM test (under appropriate regularity conditions) are computed as T times the uncentered R^2 from one of the following two regressions:

$$\hat{U}_t \quad \text{on} \quad \nabla_\alpha \hat{m}_t,\ \nabla_\beta \hat{m}_t \qquad t=1,\ldots,T \tag{2.5}$$

or

$$1 \quad \text{on} \quad \nabla_\alpha \hat{m}_t \hat{U}_t,\ \nabla_\beta \hat{m}_t \hat{U}_t \qquad t=1,\ldots,T. \tag{2.6}$$

Under H_0 and conditional homoskedasticity, the resulting statistics have limiting χ^2_Q distributions. The statistics differ in the estimators used for the asymptotic covariance matrix appearing in the LM statistic.

As emphasized by White (1980a,b,1982,1984) and White and Domowitz [1984], the statistic obtained by (2.5) generally does not have a limiting χ^2_Q distributions under H_0 in the presence of heteroskedasticity. The same is typically true of the TR^2 statistic obtained from (2.6). Therefore, if the variance of Y_t conditional on X_t is not constant, these regression procedures can lead to inference with the wrong asymptotic size concerning the null hypothesis of interest, H_0. Moreover, the standard procedures can have poor power properties if heteroskedasticity is present under local alternatives.

In closely related contexts, White (1980a,b,1982,1984), Domowitz and White (1982), and White and Domowitz (1984) have derived statistics for testing H_0 that are robust in the presence of heteroskedasticity of unknown form. Computation of these statistics requires special programming. Moreover, the estimated covariance matrices used in computing the statistics are not always positive semi-definite. This paper proposes a regression-based method for testing H_0 without imposing homoskedasticity.

To motivate the procedure, consider the regressions in (2.5) and (2.6). Note that in each regression the first term appearing on the right hand side is, by construction, orthogonal to the dependent variable in sample, i.e.

$$\sum_{t=1}^{T} \nabla_\alpha \hat{m}_t' \hat{U}_t \equiv 0. \tag{2.7}$$

Equation (2.7) is simply the first order condition for $\hat{\alpha}_T$. For derivation of the robust form of the test, interest centers on equation (2.5). Given (2.7), the uncentered R^2 from the regression in (2.5) is equal to the uncentered R^2 resulting if $\nabla_\beta \hat{m}_t$ is first projected onto $\nabla_\alpha \hat{m}_t$ and the resulting residuals are used as the only regressors in (2.5). More formally, let $\hat{\Gamma}_T$ be the PxQ matrix of regression parameters from a multivariate regression of $\nabla_\beta \hat{m}_t$ on $\nabla_\alpha \hat{m}_t$:

$$\hat{\Gamma}_T \equiv \left(\sum_{t=1}^{T} \nabla_\alpha \hat{m}_t' \nabla_\alpha \hat{m}_t\right)^{-1} \sum_{t=1}^{T} \nabla_\alpha \hat{m}_t' \nabla_\beta \hat{m}_t. \tag{2.8}$$

Then TR^2 from the regression

$$\hat{U}_t \quad \text{on} \quad (\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \hat{\Gamma}_T) \qquad t=1,\ldots,T \tag{2.9}$$

is identical to that obtained in (2.5). This shows that the LM test is equivalently based on the Qx1 vector

$$\sum_{t=1}^{T} (\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \hat{\Gamma}_T)' \hat{U}_t \tag{2.10}$$

which is identical, again by (2.7), to (2.4). The advantage of working with (2.10) instead of (2.4) is that, under H_0 and regularity conditions, it can be shown that (see the appendix)

$$\hat{\xi}_T \equiv T^{-1/2}\sum_{t=1}^{T}(\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \hat{\Gamma}_T)'\hat{U}_t \tag{2.11}$$

$$= T^{-1/2}\sum_{t=1}^{T}(\nabla_\beta m_t^o - \nabla_\alpha m_t^o \Gamma_T^o)'U_t^o \quad + \quad o_p(1)$$

where $\nabla_\alpha m_t^o \equiv \nabla_\alpha m_t(X_t,\alpha_o,0)$, $\nabla_\beta m_t^o \equiv \nabla_\beta m_t(X_t,\alpha_o,0)$, and

$$\Gamma_T^o \equiv \left[T^{-1}\sum_{t=1}^{T}E[\nabla_\alpha m_t^{o\prime}\nabla_\alpha m_t^o]\right]^{-1} T^{-1}\sum_{t=1}^{T}E[\nabla_\alpha m_t^{o\prime}\nabla_\beta m_t^o].$$

The term appearing on the right hand side of (2.11) is a function of the data and unknown parameters. A useful interpretation of (2.11) is that the limiting distribution of this random vector is unaffected when the unknown parameters are replaced by $\sqrt{T}$-consistent estimators. This feature yields a simple derivation of the regression-based robust test.

Under H_0, the right hand side of (2.11) has a limiting normal distribution with mean zero and covariance matrix

$$\Lambda_T^o \equiv T^{-1}\sum_{t=1}^{T}E[U_t^{o2}(\nabla_\beta m_t^o - \nabla_\alpha m_t^o\Gamma_T^o)'(\nabla_\beta m_t^o - \nabla_\alpha m_t^o\Gamma_T^o)]. \tag{2.12}$$

Note that Λ_T^o is the correct expression whether or not the conditional second moment of Y_t is constant. As pointed out by White (1980a,b,1982,1984), Λ_T^o is consistently estimated by

$$\hat{\Lambda}_T \equiv T^{-1}\sum_{t=1}^{T}\hat{U}_t^2(\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t\hat{\Gamma}_T)'(\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t\hat{\Gamma}_T). \tag{2.13}$$

Thus, under H_0,

$$\hat{\xi}_T'\hat{\Lambda}_T^{-1}\hat{\xi}_T \overset{d}{\to} \chi_Q^2 \tag{2.14}$$

in the presence of fairly arbitrary forms of heteroskedasticity. From a computational viewpoint, it is useful to note that the statistic in (2.14) is T times the uncentered R^2 from the regression

$$1 \quad \text{on} \quad (\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \hat{\Gamma}_T)\hat{U}_t \qquad t=1,\ldots,T. \qquad (2.15)$$

The procedure for testing H_0 can be summarized:

(i) Estimate $\hat{\alpha}_T$ by NLLS imposing $\beta_o = 0$. Compute the constrained residuals $\hat{U}_t$, and the gradients $\nabla_\alpha \hat{m}_t \equiv \nabla_\alpha m_t(\hat{\alpha}_T, 0)$ and $\nabla_\beta \hat{m}_t \equiv \nabla_\beta m_t(\hat{\alpha}_T, 0)$;

(ii) Perform a multivariate regression of $\nabla_\beta \hat{m}_t$ on $\nabla_\alpha \hat{m}_t$ and keep the residuals, $\nabla_\beta \ddot{m}_t \equiv \nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \hat{\Gamma}_T$, where $\hat{\Gamma}_T$ is given by (2.8);

(iii) Run the regression

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

and use TR^2 as asymptotically χ^2_Q under H_0.

R^2 is of course the uncentered r-squared. Equivalently, use T − SSR where SSR is the sum of squared residuals from the regression in (iii). For time series applications, T might be replaced by the actual number of observations used in the regression in (iii). ■

The above procedure yields computationally simple, robust inference for nonlinear regression. Although previously recommended statistics are not restrictively difficult to program, they sometimes involve estimated covariance matrices that are not positive semi-definite. Moreover, the regression approach is available to researchers using standard econometrics packages, and is likely to lead to fewer programming errors.

In the linear case, the above procedure produces a statistic that is numerically identical to what would be obtained by applying

Theorem 4.32 in White (1984) to the case of heteroskedasticity. Therefore, in the linear case, the above procedure can be viewed as a computionally simple method for calculating the White robust LM test.

Before giving a specific example, it is interesting to compare the LM procedure based on (2.5) to the robust procedure just outlined. Suppose that, instead of estimating Λ_T^o by $\hat{\Lambda}_T$, the conditional homoskedasticity assumption is maintained and Λ_T^o is estimated by

$$\tilde{\Lambda}_T \equiv \hat{\sigma}_T^2 T^{-1} \sum_{t=1}^{T} (\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \hat{\Gamma}_T)'(\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \hat{\Gamma}_T). \qquad (2.16)$$

where $\hat{\sigma}_T^2 \equiv T^{-1} \sum_{t=1}^{T} \hat{U}_t^2$. Using $\tilde{\Lambda}_T$ in place of $\hat{\Lambda}_T$ yields the statistic $\hat{\xi}_T' \tilde{\Lambda}_T^{-1} \hat{\xi}_T$. This statistic is numerically equivalent to TR^2 from the regression in (2.5). Therefore, under H_0 and homoskedasticity, the robust statistic is asymptotically equivalent to the usual LM statistic. Because the statistics only differ in their estimates of the moment matrix appearing in the quadratic for of the LM statistic, they are asymptotically equivalent under local alternatives and homoskedasticity. The robust form remains asymptotically noncentral χ^2 for local alternatives under heteroskedasticity. When heteroskedasticity is present, the robust form will frequently have better power properties than the usual LM statistic.

Example 2.1: (Exclusion restrictions in a linear model): Consider the linear model

$$E(Y_t | X_t) = X_{t1}\alpha_o + X_{t2}\beta_o \qquad t=1,2,...$$

where X_{t1} is a 1xP vector and X_{t2} is a 1xQ vector, both of which may contain lagged values of Y_t and current and lagged values of elements of Z_t. Applying the methodology developed above leads to a very simple heteroskedasticity-robust test of H_0: $\beta_o = 0$:

(i) Run the regression

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

and save the residuals, $\hat{U}_t$.

(ii) Run the multivariate regression

$$X_{t2} \quad \text{on} \quad X_{t1} \qquad t=1,\ldots,T$$

and save the residuals, $\ddot{X}_{t2}$.

(iii) Run the regression

$$1 \quad \text{on} \quad \ddot{X}_{t2}\hat{U}_t \qquad t=1,\ldots,T$$

and use TR^2 as asymptotically χ^2_Q under H_0. ■

The regression in step (ii) is the cost to the researcher for being robust to heteroskedasticity. One application of Example 2.1 is testing for Granger causality in time series models without imposing constancy of the conditional second moment of the dependent variable. Because noncausality in mean implies nothing about conditional second moments, standard approaches can lead to inference with the wrong asymptotic size under the null hypothesis of interest. The above methodology alleviates this problem.

When Q = 1, so that X_{t2} is a scalar, the statistic computed by steps (i)-(iii) is the square of the LM form of the White heteroskedasticity-robust t-statistic. Note that the regression in step (ii) is now a standard OLS regression. Consequently, an asymptotically equivalent version of a White (1980a) t-statistic can be computed by three OLS regressions. ■

Example 2.2: (LM test for AR(1) serial correlation): Consider a linear model with AR(1) serial correlation:

$$E(Y_t|X_t) = X_{t1}\alpha_o + \rho_o(Y_{t-1} - X_{t-1,1}\alpha_o) \qquad (2.17$$

where $|\rho_o| < 1$ and X_{t1} is a $1 \times K$ subvector of X_t. X_{t1} may contain lagged values of Y_t as well as current and lagged values of Z_t. The null hypothesis is

$$H_0 : \rho_o = 0.$$

In the above notation, $\beta = \rho$, $\nabla_\alpha \hat{m}_t = X_{t1}$, and $\nabla_\rho \hat{m}_t = \hat{U}_{t-1}$ where $\{\hat{U}_t\}$ are the OLS residuals from the regression of Y_t on X_{t1}. The following procedure is valid for testing H_0 in the presence of heteroskedasticity:

(i) Run the regression

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

and save the residuals $\hat{U}_t$.

(ii) Run the regression of $\hat{U}_{t-1}$ on W_t, and keep the residuals, say $\ddot{U}_{t-1}$.

(ii) Perform the regression

$$1 \quad \text{on} \quad \ddot{U}_{t-1}\hat{U}_t$$

and use $(T-1)R^2$ from this regression as asymptotically χ^2_1 under H_0. R^2 is the uncentered r-squared. ■

Several features of this example are worth emphasizing. First, the robust procedure follows from a straightforward modification of the LM principle, and the test is asymptotically equivalent to the usual LM statistic for AR(1) serial correlation under homoskedasticity. Also, there are essentially no restrictions on what X_{t1} can include, so that the procedure allows specification testing for dynamic regression models. The usual LM approach also has this feature, but the above procedure is valid in the presence of conditional heteroskedasticity and computationally is almost as simple as the usual LM test. Because certain economic time series exhibit conditional heteroskedasticity (see, for example Engle (1982) and Bollerslev (1986)), the heteroskedasticity-robust approach should be a useful innovation. If the conditional mean is the primary interest of the researcher, then it may be undesirable to explicitly model the second moment. The above procedure allows for asymptotically correct inference without worrying about the mechanism determining the second moments.

As pointed out by White (1985), if the procedure used in Example 2.2 rejects H_0, it is not necessarily true that (2.17) represents the true conditional expectation of Y_t given X_t. The test for AR(1) serial correlation may be detecting some other form of dynamic misspecification. In general, the statistic will reject for certain deviations from

$$H_0\colon E(Y_t | X_t) = X_{t1}\alpha_o, \qquad t=1,2,\ldots,$$

although the test is not consistent for every alternative to H_0 (see Bierens (1982,1984)).

This section concludes with a formal result. The regularity conditions imposed are not the weakest possible; in particular, it is assumed that moment matrices are stablized after normalization by T^{-1}. These assumptions rule out many nonstationary time series. They could be generalized along the lines of Wooldridge (1986) to allow for deterministic trends (this would not at all change the testing procedure), but the theorem cannot be expected to extend to certain time series models with unit roots.

A definition simplifies the statement of the theorem.

<u>Definition 2.1</u>: A sequence of random functions $\{q_t(Y_t, X_t, \theta): \theta \in \Theta, t=1,2,\ldots\}$, where $q_t(Y_t, X_t, \cdot)$ is continuous on Θ and Θ is a compact subset of $\mathbb{R}^G$, is said to satisfy the <u>Uniform Weak Law of Large Numbers</u> (UWLLN) and <u>Uniform Continuity</u> (UC) conditions provided that

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

and

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

continuous on Θ uniformly in T. ■

In the statement of the theorem, the argument β is often excluded when it is equal to 0. The dependence of functions on the predetermined variables X_t is suppressed for notational convenience.

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

(i) A is compact.

(ii) $\{m_t(x_t, \delta) : x_t \in \mathbb{R}^{L+(t-1)(L+1)}, \delta \in \Delta\}$ is a sequence of real-valued functions such that

(a) $m_t(\cdot, \delta)$ is Borel measurable for each $\delta \in \Delta$ and $m_t(x_t, \cdot, 0)$ is continuous on A for all x_t, $t=1,2,\ldots$;

(b) $m_t(x_t, \cdot)$ is twice continuously differentiable on the interior of Δ for all x_t, $t=1,2,\ldots$;

(iii) (a) $\{(m_t(\alpha)-m_t(\alpha_o))^2\}$ and $\{((m_t(\alpha)-m_t(\alpha_o))U_t^o\}$ satify the WULLN and UC conditions;

$$\text{(b)} \quad T^{-1} \sum_{t=1}^{T} [U_t^{o2} - E(U_t^{o2})] \overset{p}{\to} 0;$$

(c) α_o is the identifiably unique minimizer (see Bates and White (1985)) of

$$T^{-1} \sum_{t=1}^{T} E[(m_t(\alpha) - m_t(\alpha_o))^2];$$

(iv) α_o is in the interior of A;

(v) (a) $\{ \nabla_\alpha m_t(\alpha)' \nabla_\alpha m_t(\alpha) \}$, $\{ \nabla_\alpha m_t(\alpha)' \nabla_\beta m_t(\alpha) \}$, $\{ \nabla_{\alpha\alpha} m_t(\alpha) \}$, and $\{ \nabla_{\alpha\beta} m_t(\alpha) \}$ satisfy the WULLN and UC conditions;

(b) $T^{-1} \sum_{t=1}^{T} E[\nabla_\alpha m_t(\alpha_o)' \nabla_\alpha m_t(\alpha_o)]$ is O(1) and uniformly positive definite;

(c) $T^{-1} \sum_{t=1}^{T} E[\nabla_\alpha m_t(\alpha_o)' \nabla_\beta m_t(\alpha_o)]$ is O(1);

(vi) (a) $\{\nabla_{\alpha\alpha} m_t(\alpha)' U_t(\alpha)\}$, $\{\nabla_{\alpha\beta} m_t(\alpha)' U_t(\alpha)\}$ and $\{\nabla_\alpha m_t(\alpha)' U_t(\alpha)\}$ satisfy the WULLN and UC conditions;

(b) $T^{-1/2} \sum_{t=1}^{T} \nabla_\alpha m_t(\alpha_o)' U_t^o = O_p(1)$;

(vii) (a) $\{ U_t^2(\alpha)(\nabla_\beta m_t(\alpha) - \nabla_\alpha m_t(\alpha)\Gamma)'(\nabla_\beta m_t(\alpha) - \nabla_\alpha m_t(\alpha)\Gamma) \}$ satisfies the WULLN and UC conditions;

(b) $\Lambda_T^o \equiv T^{-1} \sum_{t=1}^{T} E[U_t^{o2}(\nabla_\beta m_t^o - \nabla_\alpha m_t^o \Gamma_T^o)'(\nabla_\beta m_t^o - \nabla_\alpha m_t^o \Gamma_T^o)]$ is O(1) and uniformly positive definite;

(c) $\Lambda_T^{o-1/2}\, T^{-1/2} \sum_{t=1}^{T} (\nabla_\beta m_t^o - \nabla_\alpha m_t^o \Gamma_T^o)' U_t^o \xrightarrow{d} N(0, I_Q)$.

Then

$$\hat{\xi}_T' \hat{\Lambda}_T^{-1} \hat{\xi}_T \xrightarrow{d} \chi_Q^2$$

where $\hat{\xi}_T$ is given by (2.11) and $\hat{\Lambda}_T$ is defined in (2.13). ■

Though the list of regularity conditions is long, the theorem is widely applicable. The WULLN and UC requirements hold under very general conditions on the process $\{(Y_t, Z_t)\}$. See Andrews (1987), Gallant and White (1986) and Newey (1987) for general treatments. The asymptotic normality assumed in (vii:c) can be established by application of the martingale central limit theorem discussed in Wooldridge (1986). Overall, the conditions of Theorem 2.1 are fairly weak, and can be expected to hold in many cases where the data are weakly dependent.

3. Conclusions and Further Applications

The approach to specification testing taken in this paper leads to simple robust inference for regression models that contain unknown forms of heteroskedasticity. The procedure should be useful for both cross section data and for time series data for which interest lies in correct dynamic specification of the regression function. Although it was assumed that the restrictions being tested are zero restrictions, a similar analysis shows that the procedure is valid when the $P+Q \times 1$ vector θ_o can be expressed as $\theta_o = r(\alpha_o)$ under H_0, where α_o is a $P \times 1$ vector and r is a differentiable function from $\mathbb{R}^P$ to $\mathbb{R}^{P+Q}$.

The approach developed here has several other applications. In Wooldridge (1987a), it is used to derive a class of regression-based conditional moment tests (Newey (1985), Tauchen (1984)) for conditional expectations estimated by QMLE. The resulting regression-based tests include the LM test discussed here, Hausman tests which do not assume relative efficiency of either estimator, and nonnested hypotheses tests that are robust to second moment misspecification under the null hypothesis. Slightly modifying the approach yields a simple regression White test for heteroskedasticity which does not assume constancy of the conditional fourth moment of the errors. This test is derived in Wooldridge (1987a) for the more general class of QMLE's assuming only that the first two conditional moments are correctly specified under the null hypothesis.

Appendix

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

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

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

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

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

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

Proof: see Wooldridge (1986, Lemma A.1, p.229).

Proof of Theorem 2.1: The major task of the proof is establishing the validity of equation (2.11).

By the the weak consistency analog of Bates and White (1985,Theorem 2.2), assumptions (i), (ii) and (iii) imply that $\hat{\alpha}_T \overset{p}{\to} \alpha_o$. Consistency of $\hat{\alpha}_T$ and (iv) imply that

$$P\left[\sum_{t=1}^{T} \nabla_\alpha \hat{m}_t' \hat{U}_t = 0 \right] \to 1 \quad \text{as } T \to \infty. \qquad (a.1)$$

Now (v:b,c) guarantee that Γ_T^o exists for sufficiently large T and is O(1).

Next, consider equation (2.11). Now

$$\hat{\xi}_T = T^{-1/2}\sum_{t=1}^{T}(\nabla_\beta \hat{m}_t - \nabla_\alpha \hat{m}_t \Gamma_T^o)'\hat{U}_t - T^{-1/2}(\hat{\Gamma}_T - \Gamma_T^o)'\sum_{t=1}^{T}\nabla_\alpha \hat{m}_t' \hat{U}_t . \quad \text{(a.2)}$$

But by (a.1), the second term on the right hand side equals zero with probability approaching one, so that the second term on the right hand side of (a.2) has no effect on the limiting distribution of $\hat{\xi}_T$. Given (v:a,b) and (vi:b), $T^{1/2}(\hat{\alpha}_T - \alpha_o) = O_p(1)$. Using this fact and the mean value expansion for random functions (Jennrich (1969, Lemma 3)), expand the first term on the right hand side of (a.2) about α_o and apply Lemma A.1 to get

$$\hat{\xi}_T = T^{-1/2}\sum_{t=1}^{T}(\nabla_\beta m_t^o - \nabla_\alpha m_t^o \Gamma_T^o)' U_t^o +$$

$$\left(T^{-1}\sum_{t=1}^{T}E[(\nabla_{\beta\alpha} m_t^o - \nabla_{\alpha\alpha} m_t^o \Gamma_T^o)' U_t^o] - T^{-1}\sum_{t=1}^{T}E[(\nabla_\beta m_t^o - \nabla_\alpha m_t^o \Gamma_T^o)'\nabla_\alpha m_t^o]\right)$$

$$\cdot\ T^{1/2}(\hat{\alpha}_T - \alpha_o) \quad + \quad o_p(1). \qquad \text{(a.3)}$$

Under H_o, $E[\nabla_{\beta\alpha} m_t^{o\prime} U_t^o] = E[\nabla_{\alpha\alpha} m_t^{o\prime} U_t^o] = 0$, and, by definition of Γ_T^o,

$$\sum_{t=1}^{T}E[(\nabla_\beta m_t^o - \nabla_\alpha m_t^o \Gamma_T^o)'\nabla m_t^o] \equiv 0.$$

Because $T^{1/2}(\hat{\alpha}_T - \alpha_o) = O_p(1)$, the second term on the right hand side of (a.3) is $o_p(1)$. This shows that (2.11) holds.

By (vii:b,c), $\Lambda_T^{o-1/2}\hat{\xi}_T \xrightarrow{d} N(0, I_Q)$. Condition (vii:a) ensures that $\hat{\Lambda}_T$ is a consistent estimator of Λ_T^o, and (vii:b) ensures that it is positive definite with probablity approaching one. Therefore, $\hat{\xi}_T'\hat{\Lambda}_T^{-1}\hat{\xi}_T \xrightarrow{d} \chi_Q^2$, and this completes the proof. ■

5562

References

Andrews, D.W.K., 1987, "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, forthcoming.

Bates, C. and H. White, 1985, "A Unified Theory of Consistent Estimation for Parametric Models," *Econometric Theory* 1, 151-175.

Bierens, H.J., 1982, "Consistent Model Specification Tests," *Journal of Econometrics* 20, 105-134.

Bierens, H.J., 1984, "Model Specification Testing of Time Series Regressions," *Journal of Econometrics* 26, 323-353.

Bollerslev, T., 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31, 307-327.

Domowitz I. and H. White, 1982, "Misspecified Models with Dependent Observations," *Journal of Econometrics* 20, 35-58.

Engle, R., 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of United Kingdom Inflation," *Econometrica* 50, 987-1008.

Gallant, A.R. and H. White, 1986, "A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models," UCSD Department of Economics Discussion Paper.

Jennrich, R., 1969, "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics* 40, 633-643.

Newey, W., 1985, "Maximum Likelihood Specification Testing and Conditional Moment Tests," *Econometrica* 53, 1047-1070.

Newey, W., 1987, "Expository Notes on Uniform Laws of Large Numbers," mimeo, Princeton University.

Tauchen, G., 1984, "Diagnostic Testing and Evaluation of Maximum Likelihood Models," Duke University Working Papers in Economics, 84-01.

White, H., 1980a, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817-838.

White, H., 1980b, "Nonlinear Regression on Cross Section Data," *Econometrica* 48, 721-746.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 1-26.

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

White, H., 1985, "Specification Testing in Dynamic Models," Invited Paper at the Fifth World Congress of the Econometric Society, Cambridge, Massachusetts, August 1985.

White, H. and I. Domowitz, 1984, "Nonlinear Regression with Dependent Observations," *Econometrica* 52, 143-162.

Wooldridge, J.M., 1986, "Asymptotic Properties of Econometric Estimators," UCSD Ph.D. dissertation.

Wooldridge, J.M., 1987a, "Specification Testing and Quasi-Maximum Likelihood Estimation," mimeo, MIT.

Wooldridge, J.M., 1987b, "Regression-based Specification Tests for Dynamic Simultaneous Equations Models," mimeo, MIT.

Date Due

AG 2 '89		
NOV 21 1990		
MAY 1 1 1991		
FEB. 2 1 1996		
MAY 2 1 1997		
AUG 0 9 2001		

Lib-26-67