working paper
department
of economics

AN ENCOMPASSING APPROACH TO CONDITIONAL MEAN TESTS WITH APPLICATIONS TO TESTING NONNESTED HYPOTHESES

Jeffrey M. Wooldridge

No. 511 February 1989

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

AN ENCOMPASSING APPROACH TO CONDITIONAL
MEAN TESTS WITH APPLICATIONS
TO TESTING NONNESTED HYPOTHESES

Jeffrey M. Wooldridge

No. 511 February 1989

AN ENCOMPASSING APPROACH TO CONDITIONAL MEAN TESTS WITH APPLICATIONS TO TESTING NONNESTED HYPOTHESES

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

February 1988
Latest Revision: January 1989

Paper presented at the 1988 Winter Meetings of the Econometric Society, New York. I wish to thank Quang Vuong and two anonymous referees for helpful comments and suggestions.

ABSTRACT: A general class of tests designed to detect conditional mean misspecification for cross section or time series applications is proposed. The tests are derived from a particular application of the encompassing principle. The resulting conditional mean encompassing (CME) tests contain as special cases a version of the Lagrange Multiplier test for nested models, a new test in the presence of nonnested alternatives, and a version of the Durbin-Wu-Hausman test that compares two weighted nonlinear least squares estimators. The tests are valid without any assumption on the conditional variance of the dependent variable and can be computed using any $\sqrt{T}$-consistent estimators. Moreover, CME tests for nonlinear, dynamic models are computable from linear least squares regressions.

1. Introduction

This paper develops a general class of tests intended to detect misspecification of a conditional expectation for cross section or time series models. The approach is based on the encompassing principle (Hendry and Richard (1982), Mizon (1984), and Mizon and Richard (1986)) in the sense that it exploits certain implications of estimating an alternative model when the model taken to be the null is true. However, for nonlinear, dynamic models the present application of the encompassing principle results in "conditional mean encompassing" (CME) tests that are more operational than the "complete parametric encompassing" (CPE) tests proposed by Mizon and Richard (1986). In particular, there is no need to solve for the "pseudo-true value" of the estimator from the alternative model, nor does one need to compute the null limiting distribution of the estimator in the alternative model. Both of these tasks can be difficult in the context of nonlinear regression with dependent observations.

The main results of the paper can be briefly summarized. For nested hypotheses the CME test is asymptotically equivalent to the Lagrange Multiplier (LM) test. For nonnested models, the CME test is based on the correlation between the residuals under the null and the gradient of the alternative regression function. The results of Wooldridge (1988) are applied throughout to produce tests that can be computed from linear regressions but do not maintain homoskedasticity or other second moment assumptions under H_0. Further, the test statistics can be computed using any $\sqrt{T}$-consistent estimators. These features make the tests applicable in situations more general than the usual LM test and standard tests of nonnested hypotheses.

When the approach is extended to weighted nonlinear least squares (WNLS) estimation, a statistic that is asymptotically equivalent to the Durbin (1954) - Wu (1973) - Hausman (1978) (DWH) statistic that compares two WNLS estimators can be shown to be a special case. Again, because this test is a special case of the general approach, the form of the statistic proposed here is regression-based but does not require either estimator to be relatively efficient under H_0. Additional robust tests in the presence of nonnested alternatives are available from WNLS estimation.

2. Setup and Motivation

Let $\{(y_t, z_t): t=1,2,\ldots\}$ be a sequence of random vectors where y_t is a scalar and z_t is a 1xK vector of conditioning variables. In a time series context, let $x_t \equiv (z_t, y_{t-1}, z_{t-1}, \ldots, y_1, z_1)$ or $x_t \equiv (y_{t-1}, z_{t-1}, \ldots, y_1, z_1)$ or $x_t \equiv (y_{t-1}, y_{t-2}, \ldots, y_1)$ be the set of predetermined variables. The choice of x_t depends on whether there are, in addition to past values of y_t, other conditioning variables $\{z_t\}$, and on whether or not the researcher wishes to condition on contemporaneous z_t. Including the entire observed past history of $\{(y_t, z_t)\}$ or $\{y_t\}$ in x_t restricts the analysis to cases where one is interested in getting the dynamics of the conditional mean correctly specified. In a cross section context, set $x_t \equiv z_t$ and assume that the observations are independently distributed.

Suppose that one is considering the following parametric model for $E(y_t|x_t)$:

$$\{m_t(x_t,\alpha): \alpha \in A, \ t=1,2,\ldots\}, \ A \subset \mathbb{R}^P. \qquad (2.1)$$

The null hypothesis that (2.1) is correctly specified for $E(y_t|x_t)$ is stated formally as

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

A general approach to testing the validity of H_0 is to compare the performance of model (2.1) in light of alternative parametric specifications for $E(y_t|x_t)$. Let

$$\{\mu_t(x_t,\beta): \beta \in B, \ t=1,2,\ldots\}, \ B \subset \mathbb{R}^Q \tag{2.3}$$

be another such parametric family. In what follows, m_t may or may not be nested within μ_t. The idea is to look for departures from H_0 in the "direction" of model (2.3), which for convenience is labelled the "alternative model." It is sometimes useful to refer explicitly to the specific alternative hypothesis H_1:

$$H_1: E(y_t|x_t) = \mu_t(x_t,\beta_o), \text{ for some } \beta_o \in B, \ t=1,2,\ldots.$$

Under H_0 and standard regularity conditions the nonlinear least squares estimator $\hat{\alpha}_T$ is weakly consistent for α_o; under further regularity conditions, $\sqrt{T}(\hat{\alpha}_T - \alpha_o)$ is asymptotically normal or at least $O_p(1)$. Explicit regularity conditions are not provided here; one set of sufficient conditions is contained in Wooldridge (1988).

Whether or not H_1 is also true, the NLS estimator $\hat{\beta}_T$ for model (2.3) can be computed by solving

$$\min_{\beta\in B} \sum_{t=1}^{T} (y_t - \mu_t(x_t,\beta))^2. \tag{2.4}$$

White and Domowitz (1984) have shown that, under H_0, $\hat{\beta}_T$ generally converges in probability to a nonstochastic sequence $\{\beta_T^o: T=1,2,\ldots\} \subset B$ which has the following optimality property: β_T^o solves the nonstochastic minimization problem

$$\min_{\beta \in B} \sum_{t=1}^{T} E[(y_t - \mu_t(x_t, \beta))^2]. \qquad (2.5)$$

Further, $\sqrt{T}(\hat{\beta}_T - \beta_T^o)$ typically has a limiting normal distribution. When the models are nonnested the asymptotic covariance matrix of $\sqrt{T}(\hat{\beta}_T - \beta_T^o)$ can be fairly complicated, especially in time series contexts. (This is because the implied errors $\{y_t - \mu_t(x_t, \beta_T^o): t=1,..,T\}$ do not constitute a martingale difference sequence under H_0; thus, they are usually serially correlated). Tests that require calculation of the asymptotic covariance matrix of $\sqrt{T}(\hat{\beta}_T - \beta_T^o)$ and/or characterization of the pseudo-true value function $\beta_T^o \equiv b_T(\alpha_o)$ are unattractive from a computational viewpoint. The complete parametric encompassing tests of Mizon and Richard (1986) have this feature for nonlinear dynamic models. The next section develops simple, regression-based tests which require only that $\sqrt{T}(\hat{\beta}_T - \beta_T^o) = O_p(1)$ and $\sqrt{T}(\hat{\alpha}_T - \alpha_o) = O_p(1)$ under H_0.

3. A New Test Based on the Encompassing Principle

The basis for the tests derived here is the statistical optimality of the sequence $\{\beta_T^o: T=1,2,...\}$. In particular, the idea is to exploit the testable implications of β_T^o solving the minimization problem (2.5); it is here that the encompassing principle is invoked.

Define the residual function for model (2.1) as $e_t(\alpha) \equiv y_t - m_t(x_t, \alpha)$. Under H_0 the true errors $e_t^o \equiv e_t(\alpha_o)$ are defined and $E(e_t^o | x_t) = 0$. Therefore, under H_0,

$$E[(y_t - \mu_t(x_t,\beta))^2] = E[(m_t(x_t,\alpha_o) - \mu_t(x_t,\beta))^2] + E[(e_t^o)^2] \tag{3.1}$$

$$\equiv E[(m_t(\alpha_o) - \mu_t(\beta))^2] + E[(e_t^o)^2].$$

Assume that $\mu_t(x_t,\cdot)$ is differentiable on int(B), $\{\beta_T^o\colon T=1,2,\ldots\} \subset$ int(B) uniformly in T, and derivatives and expectations can be interchanged. Then β_T^o must solve the first order condition

$$T^{-1}\sum_{t=1}^{T} E[\nabla_\beta\mu_t(\beta_T^o)'(m_t(\alpha_o) - \mu_t(\beta_T^o))] = 0, \tag{3.2}$$

where $\nabla_\beta\mu_t(\beta) \equiv \nabla_\beta\mu_t(x_t,\beta)$ is the 1xQ gradient of $\mu_t(x_t,\beta)$. Equation (3.2) is a testable implication of performing NLS on model (2.3) when H_0 is true. To operationalize (3.2), remove the expectations operator and replace the unknown values α_o and β_T^o by consistent estimators under H_0. Initially, let $\hat{\alpha}_T$ and $\hat{\beta}_T$ denote the NLS estimators; however, the robust testing procedure subsequently derived is valid if $\hat{\alpha}_T$ and $\hat{\beta}_T$ are <u>any</u> estimators such that $\sqrt{T}(\hat{\alpha}_T - \alpha_o) = O_p(1)$ and $\sqrt{T}(\hat{\beta}_T - \beta_T^o) = O_p(1)$ under H_0. In certain cases it is computationally convenient to use estimators other than NLS for both the null and alternative models. An example is provided in section 4.

A computable statistic is the Qx1 vector

$$T^{-1}\sum_{t=1}^{T} \nabla_\beta\mu_t(\hat{\beta}_T)'[m_t(\hat{\alpha}_T) - \mu_t(\hat{\beta}_T)] \tag{3.3}$$

$$= -T^{-1}\sum_{t=1}^{T} \nabla_\beta\mu_t(\hat{\beta}_T)'[(y_t - m_t(\hat{\alpha}_T)) - (y_t - \mu_t(\hat{\beta}_T))] \tag{3.4}$$

$$= -T^{-1}\sum_{t=1}^{T} \nabla_\beta\mu_t(\hat{\beta}_T)'\hat{e}_t \tag{3.5}$$

where $\hat{e}_t \equiv y_t - m_t(x_t,\hat{\alpha}_T)$, t=1,...,T are the residuals for model (2.1). Equation (3.5) follows from (3.4) and the first order condition for $\hat{\beta}_T$.

Thus, the optimality criterion leads to a test based on the covariance of the gradient of the alternative regression function μ_t and the residuals fitted under H_0. The statistic (3.5) is seen to be of the conditional moment form analysed by Newey (1985), Tauchen (1985), White (1987), and others. In order to distinguish the tests based on (3.5) from the complete parametric encompassing tests of Mizon and Richard (1986), the former will be called "conditional mean encompassing" (CME) tests. A CME test is simply a Newey-Tauchen-White conditional moment test using $\nabla_\beta \hat{\mu}_t$ as the misspecification indicator.

In a nested hypotheses framework, where $m_t(x_t,\alpha) \equiv \mu_t(x_t,r(\alpha))$ for some differentiable function $r\colon A \to B$, the statistic

$$T^{-1/2} \sum_{t=1}^{T} \nabla_\beta \mu_t(\hat{\beta}_T)' \hat{e}_t \qquad (3.6)$$

is closely related to the statistic underlying the Lagrange Multiplier test. Equation (3.6) leads exactly to the LM test if $\hat{\beta}_T$ in $\nabla_\beta \mu_t(\hat{\beta}_T)$ is replaced by the constrained estimator $r(\hat{\alpha}_T)$. When the unconstrained estimator $\hat{\beta}_T$ is used the resulting statistic is asymptotically equivalent to the LM statistic under the null hypothesis and under local nested alternatives.

In general, even if m_t is not nested within μ_t, (3.5) is the covariance that arises in the construction of the LM statistic for testing exclusion of $\nabla_\beta \mu_t(\hat{\beta}_T)$ in the regression (2.1). More precisely, consider the LM test for $\delta_o = 0$ in the artificial regression model

$$y_t = m_t(x_t,\alpha_o) + \nabla_\beta \mu_t(\hat{\beta}_T)\delta_o \;+\; \text{error}_t. \qquad (3.7)$$

One candidate test statistic is the TR_u^2 form of the LM test, where R_u^2 is the uncentered r-squared from the regression

$$\hat{e}_t \quad \text{on} \quad \nabla_\alpha \hat{m}_t,\ \nabla_\beta \hat{\mu}_t \qquad t=1,\ldots,T. \tag{3.8}$$

Unfortunately, even when $\hat{\alpha}_T$ is the NLS estimator of α_o, the resulting statistic does not always have a limiting chi-square distribution under H_0. It <u>is</u> true that if $\hat{\alpha}_T$ is the NLS estimator then under

$$H_0': H_0 \text{ holds } \underline{\text{and}}\ V(y_t|x_t) = \sigma_o^2, \text{ some } \sigma_o^2 > 0,\ t=1,2,\ldots, \tag{3.9}$$

TR_u^2 obtained from (3.8) has an asymptotically χ_Q^2 distribution (assuming that $\nabla_\beta \mu_t^o$ does not contain redundancies with respect to $\nabla_\alpha m_t^o$). However, the assumption of conditional homoskedasticity under H_0 is frequently implausible in economic applications, especially when y_t is a nonnegative variable. Further, by definition, a conditional mean hypothesis imposes no restrictions on the conditional variance. One goal of this paper is to develop tests based on (3.5) that do not make additional second moment assumptions under H_0. This is straightforward since the statistic (3.5) is of the general form that I have considered elsewhere (Wooldridge (1988)). The following procedure, which first purges from $\nabla_\beta \hat{\mu}_t$ its linear projection onto $\nabla_\alpha \hat{m}_t$, is valid under the regularity conditions of Theorem 2.1 in Wooldridge (1988):

PROCEDURE 3.1:

(i) Obtain $\hat{\alpha}_T$ and $\hat{\beta}_T$ by NLS, or some other procedure such that $\sqrt{T}(\hat{\alpha}_T - \alpha_o) = O_p(1)$ and $\sqrt{T}(\hat{\beta}_T - \beta_T^o) = O_p(1)$. Save the residuals $\hat{e}_t \equiv y_t - m_t(x_t,\hat{\alpha}_T)$ and the gradients $\nabla_\alpha \hat{m}_t \equiv \nabla_\alpha m_t(\hat{\alpha}_T)$ and $\hat{\lambda}_t \equiv \nabla_\beta \hat{\mu}_t \equiv \nabla_\beta \mu_t(\hat{\beta}_T)$;

(ii) Run the multivariate regression

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

and save the 1xQ vector residuals, say $\hat{\xi}_t$;

(iii) Run the regression

$$1 \quad \text{on} \quad \hat{e}_t\hat{\xi}_t \qquad t=1,..,T$$

and use $TR_u^2 = T - SSR$ as asymptotically χ_Q^2 under H_0, where SSR is the sum of squared residuals. Let $\nabla_\alpha m_t^o \equiv \nabla_\alpha m_t(\alpha_o)$, $\lambda_t^o \equiv \nabla_\beta \mu_t^o \equiv \nabla_\beta \mu_t(\beta_T^o)$ and define $\{\xi_t^o: t=1,...,T\}$ to be the residuals from the population regression

$$\lambda_t^o \quad \text{on} \quad \nabla_\alpha m_t^o,\ t=1,...,T, \tag{3.10}$$

and let $\Xi_T \equiv T^{-1}\sum_{t=1}^{T} V(e_t^o\xi_t^o)$. If $\{\Xi_T: T=1,2,...\}$ is not uniformly positive definite in T for T sufficiently large then some elements of $\hat{\lambda}_t$ are redundant with respect to $\nabla_\alpha \hat{m}_t$; the redundant elements in $\hat{\lambda}_t$ should be discarded and the degrees of freedom reduced accordingly. ■

Typically it is obvious upon inspection whether redundancies appear in $\nabla_\beta \hat{\mu}_t$. A simple instance is when both models are linear and contain overlapping regressors, a case considered more fully in the following section.

The robust procedure not only has a limiting chi-square distribution under H_0 in the presence of heteroskedasticity (conditional or unconditional) of unknown form, but it also remains asymptotically efficient in the event that $V(y_t|x_t)$ is constant. More precisely it is shown in Wooldridge (1988) that under alternatives local to H_0 that maintain conditional homoskedasticity, the robust form of the test is asymptotically equivalent to the more traditional regression test (3.8); robustness is obtained without sacrificing asymptotic efficiency under ideal conditions. It follows that any asymptotic power calculations under local alternatives and homoskedasticity for the nonrobust statistic also hold for the robust statistic. But the robust test has the further advantange of having an asymptotic noncentral chi-square distribution under alternatives local to H_0 when heteroskedasticity is present.

Derivation of the limiting distribution of the CME statistic under alternatives local to H_0 is fairly standard and is only sketched. For the present purposes, a sequence of local alternatives to H_0 is characterized by a sequence of minimizers $\{\alpha_T^*: T=1,2,...\}$ of

$$\min_{\alpha \in A} T^{-1} \sum_{t=1}^{T} E[(y_t - m_t(x_t,\alpha))^2]$$

satisfying $\sqrt{T}(\hat{\alpha}_T - \alpha_T^*) = O_p(1)$, $\sqrt{T}(\alpha_T^* - \alpha_o) = O(1)$, and

$$T^{-1/2} \sum_{t=1}^{T} E[\nabla_\beta \mu_t(\beta_T^o)' e_t(\alpha_T^*)] = O(1).$$

Letting ξ_t^o again denote the residuals under H_0 from the population regression of λ_t^o on $\nabla_\alpha m_t(\alpha_o)$, under standard regularity conditions it is straightforward to show that

$$T^{-1/2} \sum_{t=1}^{T} \hat{\xi}_t' \hat{e}_t = T^{-1/2} \sum_{t=1}^{T} \xi_t^{o\prime} e_t^* + o_p(1)$$

under the sequence of local alternatives. Thus, letting $\pi_T^* \equiv T^{-1/2} \sum_{t=1}^{T} E(\xi_t^{o\prime} e_t^*) = O(1)$ and $\Xi_T^o \equiv T^{-1} \sum_{t=1}^{T} V(e_t^o \xi_t^o)$ (Ξ_T^o is computed under H_0), it follows that the CME test has a limiting noncentral chi-square distribution with sequence of noncentrality parameters $\{\pi_T^{*\prime} \Xi_T^{o\,-1} \pi_T^*\}$. For particular alternatives π_T^* can be further simplified, but this is not attempted here. Incidentally, unlike the robust test, the local distribution of the nonrobust test under heteroskedasticity is typically unknown.

Another useful property of the robust procedure is that it is valid when <u>any</u> $\sqrt{T}$-consistent estimator of α_o is used in step (i). This is in contrast to traditional testing procedures, where the limiting distributions of statistics typically depend on the limiting distribution of $\sqrt{T}(\hat{\alpha}_T - \alpha_o)$ (an exception is Neyman's $C(\alpha)$ test). This added flexibility of the robust procedure allows simple, regression-based tests in situations where standard approaches can be computationally difficult.

It should be emphasized that the CME test was derived under the assumption that H_0 is true. Another approach to comparing nonnested models is to allow both models to be misspecified under the null. Rossi (1985) offers a Bayesian approach to model selection when neither model is assumed to be true. Vuong (1989) considers a generalized likelihood ratio approach which assumes that neither model is correctly specified under H_0 but that, in a well-defined statistical sense, they explain the data equally well.

Before turning to some examples, note that $\hat{\beta}_T$ can be any estimator such that $\sqrt{T}(\hat{\beta}_T - \beta_T^o) = O_p(1)$ for <u>some</u> sequence $\{\beta_T^o\} \subset \text{int}(B)$ uniformly in T. If $\{\beta_T^o\}$ does not have the optimality properties based on (2.5) then the test statistics is not derivable from (3.3) - (3.5). Nevertheless, the test based on Procedure 3.1 could be a useful diagnostic.

4. Examples of Nonnested Tests

Because the heteroskedasticity-robust Lagrange Multiplier statistic has been considered elsewhere (Davidson and MacKinnon (1985), Wooldridge (1987a)), this section focuses on the application of CME tests to model specification testing in the presence of nonnested alternatives.

Example 4.1: The most well-known application of nonnested hypotheses testing is to two competing linear models with different regressors. In particular,

$$m_t(x_t,\alpha) = x_{t1}\alpha \tag{4.1}$$

$$\mu_t(x_t,\beta) = x_{t2}\beta \tag{4.2}$$

where x_{t1} and x_{t2} are 1xP and 1xQ subvectors of x_t, with lag lengths independent of t. Assume that there are a sufficient number of past observations to start the indexing in (4.1) and (4.2) at t = 1. Let w_{t2} be the 1xM vector of regressors in x_{t2} but not x_{t1}. Then the form of the test which assumes homoskedasticity in addition to H_0 (see (3.8)) is simply the LM test for $\delta_o = 0$ in the model

$$E(y_t|x_t) = x_{t1}\alpha_o + w_{t2}\delta_o \qquad t=1,2,\ldots. \tag{4.3}$$

Under H_0', the LM test is asymptotically equivalent to the standard Wald test for exclusion of w_{t2}. In models with nonrandom regressors, the F-statistic as a test in the presence of nonnested hypotheses has been studied extensively by, among others, Ericsson (1983) and, more recently, as a special case of the CPE test by Mizon and Richard (1986). The CME test is the same whether or not x_t contains lagged dependent variables or other random regressors. To ensure that the test has correct asymptotic size in the presence of heteroskedasticity, Procedure 3.1 can be applied with $\nabla_\alpha \hat{m}_t \equiv x_{t1}$ and $\hat{\lambda}_t \equiv w_{t2}$. ∎

Example 4.2: Suppose that $y_t \geq 0$, and consider the following competing models for $E(y_t|x_t)$:

$$m_t(x_t,\alpha) = w_t\alpha \tag{4.4}$$

$$\mu_t(x_t,\beta) = \exp(w_t\beta) \tag{4.5}$$

where w_t is 1xP. Again, in a time series context, assume that w_t has a lag length independent of t. Note that even though $y_t \geq 0$ the linear model (4.4) cannot be ruled out *a priori*. In contrast, a normality assumption for y_t is untenable, and so it is not imposed under either model.

If the linear model is taken as the null and homoskedasticity is maintained, the CME test is an LM-type test based on TR_u^2 from the regression

$$\hat{e}_t \text{ on } w_t, \exp(w_t\hat{\beta}_T)w_t, \qquad t=1,\ldots,T, \tag{4.6}$$

where $\hat{e}_t \equiv y_t - w_t\hat{\alpha}_T$ and $\hat{\alpha}_T$ is the OLS estimator of α_o under H_0. Under H_0 and homoskedasticity, $TR_u^2 \overset{d}{\to} \chi_P^2$. Because homoskedasticity is not always a reasonable assumption for nonnegative economic variables, the heteroskedasticity-robust approach might be particularly useful. In Procedure 3.1 simply set $\nabla_\alpha \hat{m}_t \equiv w_t$ and $\hat{\lambda}_t \equiv \exp(w_t\hat{\beta}_T)w_t$. ■

Example 4.3: Frequently researchers are interested in comparing linear and log-linear regression models. Although this is certainly in the spirit of comparing linear and exponential forms for $E(y_t|x_t)$, linearity of $E(\log y_t|x_t)$ need not imply that $E(y_t|x_t)$ has an exponential form, nor vice versa. To compare linear and log-linear models a further assumption is needed under the log-linear model. Many tests assume that

$$\log y_t|x_t \sim N(w_t\delta_o,\sigma_o^2), \text{ some } \delta_o \in \Delta, \text{ some } \sigma_o^2 > 0. \tag{4.7}$$

Here it suffices to make the weaker assumptions

$$E(y_t|x_t) = \exp(w_t\alpha_o) \tag{4.8}$$

$$E(y_t|x_t) = \exp[\kappa_o + E(\log y_t|x_t)] \tag{4.9}$$

for some $\kappa_o > 0$. If (4.7) holds it is well known that $\kappa_o = \sigma_o^2/2$ in (4.9).

The conditional mean of $\log y_t$ under (4.8) and (4.9) is

$$E(\log y_t|x_t) = \alpha_{o1} - \kappa_o + \alpha_{o2}w_{t2} + \dots + \alpha_{oP}w_{tP}$$
$$\equiv w_t\delta_o,$$

where it has been assumed that $w_{t1} \equiv 1$. Testing the log-linear model against the linear alternative is very simple. First, let $\hat{\delta}_T$ and $\widehat{\log y_t}$ be the OLS estimator and fitted values from the regression

$$\log y_t \text{ on } w_t \qquad t=1,\dots,T.$$

Compute an estimate of $\exp(\kappa_o)$ and predicted values of y_t from the OLS regression of y_t on $\exp[\widehat{\log y_t}]$ (without an intercept). Let $\hat{\kappa}_T$ and $\hat{y}_t$ be the estimator of κ_o and the fitted values of y_t, respectively, and define the residuals as $\hat{e}_t \equiv y_t - \hat{y}_t$. Then simply apply Procedure 3.1 with $\nabla_\alpha \hat{m}_t \equiv \hat{y}_t w_t$ and $\hat{\lambda}_t \equiv w_t$. Note that the computations required for the test can be done entirely by <u>linear</u> least squares regressions. Also, the implicit estimator for α_o, $\hat{\alpha}_T \equiv (\hat{\delta}_{T1}+\hat{\kappa}_T, \hat{\delta}_{T2}, \dots, \hat{\delta}_{TP})$, has no particular optimality properties even under conditional homoskedasticity, yet the test is asymptotically equivalent to the procedure which uses the NLS estimator of α_o.

This test requires only the additional assumption (4.9) to compare linear and log-linear regression models, and not the stronger assumption (4.7). If (4.7) is believed to be true then this test cannot be optimal; the only information about y_t that is used is the exponential form of the conditional expectation, so that additional information about the conditional distribution of y_t given x_t is ignored. The strength of the current approach is that it does not require distributional or second moment assumptions under either model. The Cox (1961,1962) test, which requires distributional assumptions under H_0 as well as H_1, can be quite difficult to compute (see

Aneuryn-Evans and Deaton (1980)).

The procedure that takes the linear model as the null hypothesis (and does not impose distributional or variance assumptions under the null) is the same as Example 4.2, except that $\hat{\beta}_T$ is constructed from a log-linear OLS regression as above rather than NLS. Compared with procedures that impose a plausible distribution of $y_t|x_t$ in the linear model, the current approach is more robust and computationally much easier. ■

5. Extension to Weighted Nonlinear Least Squares

The approach of section 3 extends directly to the case where one or both models are estimated by weighted NLS (WNLS). Let

$$\{m_t(x_t,\alpha): \alpha \in A\} \qquad \{h_t(x_t,\gamma): \gamma \in \Gamma\} \tag{5.1}$$

and

$$\{\mu_t(x_t,\beta): \beta \in B\} \qquad \{\eta_t(x_t,\delta): \delta \in \Delta\} \tag{5.2}$$

be the "competing" models, where h_t and η_t are weighting functions such that $h_t(x_t,\gamma) > 0$, $\eta_t(x_t,\delta) > 0$. It is important to stress that the null hypothesis is the <u>same</u> as in section 3, i.e.

$$H_0: E(y_t|x_t) = m_t(x_t,\alpha_o) \quad \text{for some } \alpha_o \in A,\ t=1,2,\ldots. \tag{5.3}$$

It is <u>not</u> assumed that $h_t(x_t,\gamma)$ is a correctly specified parameterized version of the conditional variance of y_t given x_t under H_0 (i.e. it is not assumed that $h_t(x_t,\gamma_o)$ is proportional to $V(y_t|x_t)$ for some $\gamma_o \in \Gamma$). Instead, assume that there are estimators of the nuisance parameters γ and δ such that

$$\sqrt{T}(\hat{\gamma}_T - \gamma_T^o) = O_p(1), \qquad \sqrt{T}(\hat{\delta}_T - \delta_T^o) = O_p(1) \tag{5.4}$$

where $\{\gamma_T^o\}$ and $\{\delta_T^o\}$ are nonstochastic sequences. First suppose that $\hat{\alpha}_T$ is the WNLS estimator that solves

$$\min_{\alpha \in A} \sum_{t=1}^{T} (y_t - m_t(x_t,\alpha))^2/h_t(x_t,\hat{\gamma}_T). \tag{5.5}$$

The WNLS estimator based on model (5.2) solves

$$\min_{\beta \in B} \sum_{t=1}^{T} (y_t - \mu_t(x_t,\beta))^2/\eta_t(x_t,\hat{\delta}_T). \tag{5.6}$$

The solution to (5.6), again denoted $\hat{\beta}_T$, is generally such that

$$\sqrt{T}(\hat{\beta}_T - \beta_T^o) = O_p(1) \tag{5.7}$$

where β_T^o solves the nonstochastic minimization problem

$$\min_{\beta \in B} \sum_{t=1}^{T} E[(y_t - \mu_t(x_t,\beta))^2/\eta_t(x_t,\delta_T^o)].$$

Under H_0,

$$E[(y_t - \mu_t(x_t,\beta))^2/\eta_t(x_t,\delta_T^o)] \tag{5.8}$$

$$= E[(m_t(\alpha_o) - \mu_t(\beta))^2/\eta_t(\delta_T^o)] + E[(e_t^o)^2/\eta_t(\delta_T^o)].$$

The appropriate first order condition for β_T^o is

$$T^{-1}\sum_{t=1}^{T} E[\nabla_\beta \mu_t(\beta_T^o)'(m_t(\alpha_o) - \mu_t(\beta_T^o))/\eta_t(\delta_T^o)] = 0$$

and the relevant statistic is

$$= T^{-1}\sum_{t=1}^{T} [\nabla_\beta \mu_t(\hat{\beta}_T)/\eta_t(\hat{\delta}_T)]'\hat{e}_t$$

$$= T^{-1}\sum_{t=1}^{T} [h_t(\hat{\gamma}_T)^{-1/2}(h_t(\hat{\gamma}_T)/\eta_t(\hat{\delta}_T))\nabla_\beta \mu_t(\hat{\beta}_T)]'h_t(\hat{\gamma}_T)^{-1/2}\hat{e}_t$$

$$\equiv T^{-1}\sum_{t=1}^{T} [\hat{h}_t^{-1/2}\hat{\lambda}_T]'\hat{h}_t^{-1/2}\hat{e}_t \tag{5.9}$$

where $\hat{e}_t \equiv y_t - m_t(\hat{\alpha}_T)$ and

$$\hat{\lambda}_t \equiv (\hat{h}_t/\hat{\eta}_t)\nabla_\beta\hat{\mu}_t. \tag{5.10}$$

If the models are nested and $\hat{h}_t = \hat{\eta}_t$ then (5.9) leads to a statistic that is asymptotically equivalent to the usual LM statistic in the context of WNLS. More generally, (5.9) suggests basing a test on the correlation of the weighted indicator $\hat{h}_t^{-1/2}\hat{\lambda}_t$ and the weighted residuals under H_0, $\hat{h}_t^{-1/2}\hat{e}_t$. Note that the indicator $\hat{\lambda}_t$ is a particular weighting of the gradient of the alternative regression function. If instead of H_0 the null hypothesis is

$$H_0'': H_0 \text{ holds, and for some } \gamma_o \in \Gamma,\ \sigma_o^2 > 0, \tag{5.11}$$
$$V(y_t|x_t) = \sigma_o^2 h_t(x_t,\gamma_o), \quad t=1,2,\ldots$$

then, assuming now that $\hat{\gamma}_T$ is a $\sqrt{T}$-consistent estimator of γ_o, a simple regression test using the weighted residuals as the dependent variable is available. Let $\tilde{e}_t \equiv \hat{h}_t^{-1/2}\hat{e}_t$, $\nabla_\alpha\tilde{m}_t \equiv \hat{h}_t^{-1/2}\nabla_\alpha\hat{m}_t$, and $\tilde{\lambda}_t \equiv \hat{h}_t^{-1/2}\hat{\lambda}_t = (\hat{h}_t^{1/2}/\hat{\eta}_t)\nabla_\beta\hat{\mu}_t$. The LM-like test is obtained by running the regression

$$\tilde{e}_t \text{ on } \nabla_\alpha\tilde{m}_t,\ \tilde{\lambda}_t \qquad t=1,\ldots,T \tag{5.12}$$

and using TR_u^2 as asymptotically χ_Q^2 under H_0'' (assuming no redundancies in $\tilde{\lambda}_t$). The following procedure is valid whether or not $\{h_t(x_t,\gamma): \gamma \in \Gamma\}$ contains a version of $V(y_t|x_t)$ under H_0.

PROCEDURE 5.1:

(i) Let $\hat{\alpha}_T$ be any $\sqrt{T}$-consistent estimator of α_o under H_0, and let $\hat{\beta}_T$ be an estimator such that $\sqrt{T}(\hat{\beta}_T - \beta_T^o) = O_p(1)$. Compute $\hat{\alpha}_T$, $\hat{h}_t$, $\nabla_\alpha\hat{m}_t$, $\hat{e}_t$, $\hat{\beta}_T$, $\hat{\eta}_t$, $\nabla_\beta\hat{\mu}_t$ and $\hat{\lambda}_t \equiv (\hat{h}_t/\hat{\eta}_t)\nabla_\beta\mu_t(\hat{\beta}_T)$. Define $\tilde{e}_t \equiv \hat{h}_t^{-1/2}\hat{e}_t$, $\nabla_\alpha\tilde{m}_t \equiv \hat{h}_t^{-1/2}\nabla_\alpha\hat{m}_t$, and $\tilde{\lambda}_t \equiv \hat{h}_t^{-1/2}\hat{\lambda}_t$;

(ii) Run the regression

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

and save the residuals, say $\tilde{\xi}_t$;

(iii) Run the regression

$$1 \quad \text{on} \quad \tilde{e}_t \tilde{\xi}_t \qquad t=1,...,T$$

and use $TR_u^2 = T - SSR$ as asymptotically χ_Q^2 under H_0. Again, delete any redundant elements in $\tilde{\lambda}_t$ and reduce the degrees of freedom as needed. ∎

The weighted NLS extension allows simple robust tests for a wide variety of models. In particular, quasi-maximum likelihood estimation (QMLE) of a linear exponential family is accomodated because the QMLE is asymptotically equivalent to a particular WNLS estimator with estimated weights (see Gourieroux, Monfort, and Trognon (1984)).

Being agnostic about whether the family $\{h_t(x_t,\gamma): \gamma \in \Gamma\}$ contains the conditional variance of y_t under H_0 allows for possible improvements over NLS (although this is in no way guaranteed!) while guarding against inference with incorrect asymptotic size due to a misspecified variance. Moreover, nothing is lost in terms of local power if the weighting function happens to be correctly specified for the conditional variance. In other words, the robust procedure is optimal (in the class of WNLS procedures) if $h_t(x_t,\gamma)$ is a correctly specified version of $V(y_t|x_t)$, and so $h_t(x_t,\gamma)$ should reflect the researcher's best guess for $V(y_t|x_t)$.

As an illustration, consider the analysis of count data. One might believe that a conditional Poisson distribution provides a better approximation to the second moment of y_t than, say, the assumption of homoskedasticity. However, the assumption that the conditional mean and variance are equal (or proportional) is not one on which conditional mean specification testing should rely. And if the mean and variance do happen to be equal, nothing is lost asymptotically by using the robust procedure.

If the nonrobust tests had the ability to systematically detect violation of the conditional variance assumption then the nonrobustness criticism would be somewhat mitigated. However, the nonrobust conditional mean tests (and the robust forms proposed in this paper) are inconsistent against the alternative

$$H_1'': H_0 \text{ holds but } V(y_t|x_t) \neq \sigma^2 h_t(x_t,\gamma) \text{ for all } \gamma \in \Gamma,\ \sigma^2 > 0. \qquad (5.13)$$

Consequently, one should not expect to detect departures from the conditional variance assumption by using nonrobust conditional mean tests. Under H_0, the actual size of the nonrobust test can be larger or smaller than the nominal size, and it is difficult if not impossible to determine *a priori* which is likely to be the case.

The asymptotic local distribution of the CME test for WNLS is analogous to the NLS case. The indicator $\lambda_t^o = (h_t^o/\eta_t^o)\nabla_\beta\mu_t^o$ now replaces $\nabla_\beta\mu_t^o$. With this modification the same calculation works if e_t^* and ξ_t^o are simply weighted by $h_t^{o-1/2}$.

Turning now to an example of an CME test for a weighted NLS problem, again consider testing an exponential versus linear regression model.

<u>Example</u> <u>5.1</u>: The competing models are

$$m_t(x_t,\alpha) \equiv \exp(w_t\alpha) \qquad (5.14)$$

$$\mu_t(x_t,\beta) \equiv w_t\beta. \qquad (5.15)$$

Suppose that the test which takes the exponential model as the null is to be based on a weighted sum of squared residuals. In particular, let the weighting function be the square of the regression function: $h_t(x_t,\gamma) \equiv [\exp(w_t\alpha)]^2$. It is important to stress that $[\exp(w_t\alpha_o)]^2$ is not assumed to be proportional to $V(y_t|x_t)$ under H_0, although this of course is not ruled out.

The estimator $\hat{\alpha}_T$ can be the NLS estimator, or the WNLS estimator which solves

$$\min_{\alpha\in A} \sum_{t=1}^{T} (y_t - \exp(w_t\alpha))^2/\hat{h}_t.$$

In the context of Example 4.3 under (4.8) and (4.9), a computationally convenient estimator is obtained from the log-linear regression. For any $\sqrt{T}$-consistent estimator $\hat{\alpha}_T$ define the weighted residuals and weighted gradient as $\tilde{e}_t \equiv \hat{h}_t^{-1/2}\hat{e}_t$, $\nabla_\alpha \tilde{m}_t \equiv \hat{h}_t^{-1/2}\nabla_\alpha \hat{m}_t = \exp(-w_t\hat{\alpha}_T)\exp(w_t\hat{\alpha}_T)w_t = w_t$. The indicator is $\hat{\lambda}_t \equiv w_t$, and the weighted indicator is $\tilde{\lambda}_t \equiv \hat{h}_t^{-1/2}w_t = \exp(-w_t\hat{\alpha}_T)w_t$. These quanitities are then used in Procedure 5.1.

Note that in the setup of Example 4.3 all computations can be carried out by OLS. Also, the weights can be easily computed as $\hat{h}_t = [\exp(\widehat{\log} y_t)]^2$. ■

Consideration of weighted NLS introduces a possibility not allowed in the framework of section 3. Provided that h_t and η_t are sufficiently different one can take $\alpha = \beta$ and $\mu_t(x_t,\beta) = m_t(x_t,\alpha)$. That is, suppose one does not have a particular alternative to m_t in mind (either nested or nonnested), but instead another WNLS estimator is used to detect misspecification of m_t. This application of the Durbin-Wu-Hausman methodology has been considered by White (1980) in the context of NLS on cross section data. It can be shown that, when $h_t(x_t,\gamma)$ is correctly specified for $V(y_t|x_t)$, the statistic obtained from the regression (5.12) is asymptotically equivalent to the DWH statistic that compares the difference of the two WNLS estimators and exploits the relative efficiency of the WNLS estimator based on $\hat{h}_t$ (for a similar result, see Ruud (1984)). The robust approach obtained by setting $\mu_t = m_t$ in Procedure 5.1 does not require either estimator to be relatively efficient under H_0, but it is still asymptotically equivalent to the usual DWH test in the event that $h_t(x_t,\gamma)$ is correctly specified for $V(y_t|x_t)$.

6. Comparison with Other Related Nonnested Hypotheses Tests

Davidson and MacKinnon (1981) (DM) suggested a method for testing nonnested, nonlinear regression models which has proven to be useful in practice. Their approach can be derived from the general framework of Cox (1961,1962) under normality and homoskedasticity.

In the notation of this paper, the DM statistic is obtained by testing $\delta_o = 0$ in the artificial model

$$y_t = (1-\delta_o)m_t(x_t,\alpha_o) + \delta_o\mu_t(x_t,\hat{\beta}_T) + \text{error}_t. \qquad (6.1)$$

The LM form of the test is particularly convenient since it requires only NLS estimation of each model and then one auxiliary OLS regression. Let $\hat{e}_t \equiv y_t - m_t(x_t,\hat{\alpha}_T)$ be the residuals from the model under H_0. Then the LM approach is to compute R_u^2 from the regression

$$\hat{e}_t \text{ on } \nabla_\alpha\hat{m}_t,\ \hat{\mu}_t - \hat{m}_t \quad t=1,\ldots,T \qquad (6.2)$$

and use TR_u^2 as asymptotically χ_1^2 under H_0. Thus, the DM test is simply an omitted variables test of $\hat{\mu}_t - \hat{m}_t$ in the nonlinear model

$$y_t = m_t(x_t, \alpha_o) + e_t^o. \tag{6.3}$$

The standard DM test, as well as the LM form in (6.2), is invalid in the presence of heteroskedasticity. A robust version can be computed by modifying the misspecification indicator in Procedure 3.1: simply set $\hat{\lambda}_t \equiv \hat{\mu}_t - \hat{m}_t$ (see Wooldridge (1987b)). Because the robust version allows $\hat{\alpha}_T$ and $\hat{\beta}_T$ to be any $\sqrt{T}$-consistent estimators, a DM test for the log-linear versus linear model can be computed entirely with OLS along the lines of Example 4.3.

A robust DM test based on WNLS estimation (and therefore for QMLE in a linear exponential family) is also easy to obtain. Let the mean and weighting functions be given by (5.1) and (5.2), with the null hypothesis taken to be (5.3). Then the indicator $\hat{\lambda}_t$ for the DM test is the scalar $\hat{\lambda}_t \equiv (\hat{h}_t/\hat{\eta}_t)(\hat{\mu}_t - \hat{m}_t)$. Note that the same reweighting of the indicator that appears in the CME test also shows up in the DM test for weighted nonlinear regressions. This misspecification indicator is used in Procedure 5.1 in place of $(\hat{h}_t/\hat{\eta}_t)\nabla_\beta \hat{\mu}_t$.

Even though the DM test is only a one degree of freedom test, it is always consistent against the alternative H_1. In the case of unweighted NLS, consistency of the test follows if it can be shown that, under H_1,

$$\liminf_{T\to\infty} \left| T^{-1} \sum_{t=1}^{T} E[(\mu_t(\beta_o) - m_t(\alpha_T^o))e_t(\alpha_T^o)] \right| > 0. \tag{6.4}$$

But $e_t(\alpha_T^o) = y_t - m_t(\alpha_T^o) = \epsilon_t(\beta_o) + \mu_t(\beta_o) - m_t(\alpha_T^o)$, where $\epsilon_t(\beta) \equiv y_t - \mu_t(x_t, \beta)$. Under H_1, $E(\epsilon_t^o | x_t) = 0$, so that

$$E[(\mu_t(\beta_o) - m_t(\alpha_T^o))e_t(\alpha_T^o)] = E[(\mu_t(\beta_o) - m_t(\alpha_T^o))^2],$$

and (6.4) holds except in uninteresting degenerate situations.

The CME test has degrees of freedom that depend on the dimension of β in the alternative model; without redundancies, the degrees of freedom equals the dimension of β. The condition for consistency of the CME test against H_1 is

$$\liminf_{T\to\infty} \left| T^{-1} \sum_{t=1}^{T} E[\nabla_\beta \mu_t(\beta_o)' e_t(\alpha_T^o)] \right| > 0, \tag{6.5}$$

where $|\cdot|$ now denotes Euclidean norm. Under H_1,

$$E[\nabla_\beta \mu_t(\beta_o)' e_t(\alpha_T^o)] = E[\nabla_\beta \mu_t(\beta_o)'(\mu_t(\beta_o) - m_t(\alpha_T^o))]$$

and the condition for consistency reduces to

$$\liminf_{T\to\infty} \left| T^{-1} \sum_{t=1}^{T} E[\nabla_\beta \mu_t(\beta_o)'(\mu_t(\beta_o) - m_t(\alpha_T^o))] \right| > 0. \tag{6.6}$$

For general μ_t and m_t it is possible for (6.6) to fail. Nevertheless, for linear models, (6.6) holds except in degenerate situations. Also, when m_t and μ_t are linear or exponential functions, (6.6) holds provided that the regressors contain a constant. Consistency is easy to establish for the LM-type tests that employ the NLS estimators since the regressors in the DM auxiliary regression are linear combinations of the regressors in the CME test regression. To verify consistency of the CME test for the more general robust procedure, (6.6) can be demonstrated directly for all of the examples in section 4.

The dimension of the space of alternatives against which the CME test is consistent is greater than the corresponding dimension for the DM test. When the alternatives of interest consist only of the specified competing model then the one degree of freedom test may be adequate. However, as a general model diagnostic, the DM test may not have sufficient power against certain alternatives of interest. For more on the issue of the "implicit null hypothesis" of a test the reader is referred to Pesaran (1982), MacKinnon (1983, with discussion), Mizon and Richard (1986, section 4), and Davidson and MacKinnon (1987). Characterizing the implicit null in a useful manner for the CME test in nonlinear, dynamic models is difficult and is necessarily done on a case by case basis. Mizon and Richard (1986, section 4) find the implicit null of the DM and CPE tests for competing linear models with strictly exogenous regressors; the same calculation works for the CME test in this case.

As mentioned several times above, Mizon and Richard (1986) develop the notion of complete parametric encompassing tests and discuss how they can be applied to testing nonnested hypotheses. The CPE tests are closely related to the tests of Gourieroux, Monfort, and Trognon (1983): both approaches rely on the notion of a pseudo-true value in the alternative model. The tests derived by MR and GMT lead to well-known tests in nested situations, and are similar in spirit to the tests derived here. To compare the CPE tests and the CME tests the CPE principle must be extended to the case where the conditional distribution of y_t given x_t and joint distribution of $(y_1,z_1),\ldots,(y_T,z_T)$ are not completely specified. Letting $b_T(\alpha_o)$ denote the pseudo-true value of β under H_0, the Wald encompassing test (WET) is based on

$$T^{1/2}(\hat{\beta}_T - b_T(\hat{\alpha}_T)). \qquad (6.6)$$

It can be shown that when the regressors are treated as nonrandom (or strictly exogenous), (6.6) and (3.6) are asymptotically equivalent up to multiplication by a sequence of uniformly positive definite matrices. Thus, in this case, the CPE and CME tests are asymptotically equivalent, and the CME test can be viewed as a computationally simple robust version of the CPE tests of Mizon and Richard (1986) and GMT (1983). The equivalence breaks down for general dynamic models, partly because the pseudo-true value function becomes a complicated function of the parameters of the distribution of y_t (including α). In fact, strictly speaking, the CPE tests as developed by Mizon and Richard (1986) cannot be computed for all cases considered here because the null hypothesis in this paper only specifies $E(y_t|x_t)$, whereas the derivative of the pseudo-true value function can depend on the joint as well as the conditional distribution of y_t given x_t. As this function is needed to compute the limiting distribution of the CPE statistic, one must specify more than $E(y_t|x_t)$ under H_0. Nevertheless, in some cases the "natural" way of operationalizing a CPE test leads to a test asymptotically equivalent to the corresponding CME test. If the alternative model is linear, e.g. $\mu_t(x_t,\beta) = w_t\beta$, it is sensible to choose

$$b_T(\alpha_o) \equiv \left(\sum_{t=1}^{T} w_t'w_t\right)^{-1} \sum_{t=1}^{T} w_t'm_t(x_t,\alpha_o)$$

(but note that $b_T(\alpha_o)$ is generally random). In this case it is straightforward to show that (6.6) and (3.6) lead to asymptotically equivalent tests.

When the CPE test and the CME test are not asymptotically equivalent it is difficult to determine analytically which has superior power properties. It is unlikely that one test uniformly dominates the other in terms of asymptotic local power. Comparing the powers of the tests in situations where they are not asymptotically equivalent requires a detailed study that is beyond the scope of the current paper. But it is useful to note that when the two tests are not asymptotically equivalent the CME tests have significant computational advantages, as well as being easy to "robustify."

As exposited by GMT (1983) and Mizon and Richard (1986), the CPE principle has broad applicability. However, the current application of the encompassing principle also applies to situations more general than WNLS. The approach used in sections 3 and 5 can be invoked in any setting where estimators are defined through optimization problems, including the general maximum likelihood setting. But in more general settings the resulting tests suffer from one of the same drawbacks as the CPE tests: calculation of the test statistic requires estimation of the asymptotic variance of $\sqrt{T}(\hat{\beta}_T - \beta_T^o)$ under H_0. Nevertheless, the approach of this paper never requires one to find or even to characterize in any way the pseudo-true value function. The extension to general (quasi) maximum likelihood estimation is left to future work, primarily because the the statistics would no longer be very easy to compute.

7. Conclusions

The conditional mean tests developed in this paper are applicable to testing nested and nonnested hypotheses for cross section or dynamic conditional means. There are several attractive features of these test. First, they can be computed by using linear least squares regressions after the original estimation. Second, they do not require homoskedasticity or other second moment assumptions, and can be computed using any $\sqrt{T}$-consistent estimators. Finally, the CME tests are asymptotically equivalent to well known tests in special cases, such as the LM test for nested models and the Durbin-Wu-Hausman test for comparing two WNLS estimators of the same parameters.

Further work needs to be done to investigate the finite sample properties of the statistics proposed here. Ericsson (1983) has compared the powers of the regression F-test mentioned in Example 4.1 to the DM test, and the F-test compares favorably for many alternatives. One might expect the CME tests to perform well in more general nonlinear, dynamic models, but this remains to be seen. In addition, it would be useful to further investigate the relationship between the complete parametric encompassing tests of Mizon and Richard (1986) and the CME tests.

The nonnested tests extend easily to the case of more than one alternative regression function. One merely includes the gradients (or weighted gradients) of all competing models as indicators. The same regression procedures are still appropriate. Also, CME tests can be derived in a straightforward manner for multivariate models that are estimated by multivariate WNLS.

References

Aneuryn-Evans, G., and A. Deaton, 1980, Testing Linear Versus Logarithmic Regression Models, Review of Economic Studies 47, 275-291.

Cox, D.R., 1961, Tests of Separate Families of Hypotheses, in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 105-123.

Cox, D.R., 1962, Further Results on Tests of Separate Families of Hypotheses, Journal of the Royal Statistical Society Series B 24, 406-424.

Davidson, R., and J.G. MacKinnon, 1981, Several Tests of Model Specification in the Presence of Alternative Hypotheses, Econometrica 49, 781-793.

Davidson, R. and J.G. MacKinnon, 1985, Heteroskedasticity-Robust Tests in Regression Directions," *Annales de l'INSEE*, 59/60, 183-218.

Davidson, R. and J.G. MacKinnon, 1987, Implicit Alternatives and the Local Power of Test Statistics, Econometrica 55, 1305-1329.

Durbin, J. (1954), Errors in Variables, Review of the International Statistical Institute 22, 23-32.

Ericsson, N.R., 1983, Asymptotic Properties of Statistics for Testing Nonnested Hypotheses, Review of Economic Studies 50, 287-304.

Gourieroux, C., A. Monfort, and A. Trognon, 1983, Testing Nested or Nonnested Hypotheses, Journal of Econometrics 21, 83-115.

Gourieroux, C., A. Monfort, and A. Trognon, 1984, Pseudo Maximum Likelihood Methods: Theory, Econometrica 52, 681-700.

Hausman, J.A., 1978, Specification Tests in Econometrics, Econometrica 46, 1251-1271.

Hendry, D.F., and J.-F. Richard, 1982, On the Formulation of Empirical Models in Dynamic Econometrics, Journal of Econometrics 20, 3-33.

Mizon, G.E., 1984, The Encompassing Approach in Econometrics, in Econometrics and Quantitive Modelling, ed. D.F. Hendry and K.F. Wallis. Oxford: Basil Blackwell.

Mizon, G., and J.-F. Richard, 1986, The Encompassing Principle and Its Application to Testing Nonnested Hypotheses, Econometrica 54, 657-678.

Pesaran, M.H., 1982, Comparison of Local Power of Alternative Tests of Non-nested Regression Models, Econometrica 50, 1287-1305.

Pesaran, M.H., 1987, Global and Partial Non-nested Hypotheses and Asymptotic Local Power, Econometric Theory 3, 69-97.

Rossi, P.E., 1985, Comparison of Alternative Functional Forms in Production, Journal of Econometrics 30, 345-361.

Ruud, P.A., 1984, Tests of Specification in Econometrics, Econometric Reviews 3, 211-242.

Vuong, Q.H., 1989, Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses, Econometrica , forthcoming.

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

White, H., 1987, Specification Testing in Dynamic Models, in Advances in Econometrics, ed. T.F. Bewley. Cambridge: Cambridge University Press, 1-58.

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

Wooldridge, J.M., 1987a, A Regression-based Lagrange Multiplier Statistic that is Robust in the Presence of Heteroskedasticity, MIT Department of Economics Working Paper 478.

Wooldridge, J.M., 1987b, Specification Testing and Quasi-Maximum Likelihood Estimation, MIT Department of Economics Working Paper 479.

Wooldridge, J.M., 1988, A Unified Approach to Robust, Regression-Based Specification Tests, MIT Department of Economics Working Paper 480.

Wu, D., 1973, Alternative Tests of Independence Between Stochastic Regressors and Disturbances, Econometrica 41, 733-750.