CONDITIONAL MOMENT TESTS WITH EXPLICIT ALTERNATIVES

A. Colin Cameron
and
Pravin K. Trivedi

Working Paper Series
#366

Department of Economics
University of California
Davis, California 95616-8578

CONDITIONAL MOMENT TESTS WITH EXPLICIT ALTERNATIVES

by

A. Colin Cameron
Department of Economics
University of California
Davis, California 95616-8578

Pravin K. Trivedi
Department of Economics
Indiana University
Bloomington, Indiana 47405

Working Paper Series No. 366
September 1990

ABSTRACT

The conditional moment (CM) tests of Newey (1985) and Tauchen (1985) were developed as diagnostic tests of whether or not a moment condition holds. In this paper CM tests against an explicit alternative are presented. In many testing situations this is the natural way to proceed in constructing tests, mirroring the procedure used for classical statistical tests. Furthermore, many diagnostic tests such as the information matrix test, apparently without an alternative, can be interpreted as tests against an explicit alternative. The CM tests considered here are motivated by a regression, and are accordingly called regression-based CM tests. Testing can be based directly on this regression, when the moment fundamental to the test satisfies an additional moment condition. An example where this condition holds yields new tests for misspecification of the central moments. New CM tests based on orthogonal polynomials are also presented.

Some Key Words: conditional moment specification tests; LM tests; score tests; information matrix tests; orthogonal polynomials; generalized linear models; heteroskedasticity; symmetry; kurtosis.

1. Introduction

Conditional moment (CM) tests, introduced by Newey (1985) and Tauchen (1985), offer a unifying framework for tests of parametric model misspecification, and are often simple to implement.

The richness of this approach is demonstrated in a recent book by White (1990). He draws in to this framework the classical tests - the Wald, Likelihood Ratio and Lagrange Multiplier (or Score) test principles - and non-classical tests - Hausman, Information Matrix, and Encompassing Tests. He obtains the distributions of moment test statistics under minimal assumptions about the data generating process, and proposes auxiliary regressions to compute the test statistics. White advocates a methodology of specification analysis based on this theory. Pagan and Vella (1989) emphasize the simplicity of CM testing and very strongly urge its adoption.

Nonetheless, specification testing based on general moment conditions has been slow to take hold. There are several reasons for this. First, there is a perception that the newer moment tests, i.e. those not based on the three classical test procedures, do not involve the formulation of an alternative hypothesis model beyond stating that the null does not hold, and therefore are more difficult to intuitively interpret. Second, the wide range of moment conditions on which to base CM tests leads to aversion of seemingly ad-hoc choice of moment conditions and reversion to moment tests based on classical test principles whose properties are well known. Finally, in most applications of CM tests, a density is actually specified under the null (e.g. any information matrix test requires this), and a generalization of the information matrix equality is used to justify computation of test statistics by an auxiliary regression, called the outer-product-of-the-gradient regression by Godfrey (1988) and Davidson and MacKinnon (1990), that has very poor size properties.

In this paper we propose an approach to CM testing, the regression-based CM test, that overcomes these objections. A specified conditional moment, called the "fundamental" moment, equals zero under the null hypothesis. The fundamental moment is then embedded in a more general alternative: under the alternative hypothesis the fundamental moment equals a specified function of parameters and exogenous variables. The null hypothesis is tested by performing a significance test on the coefficients from the regression of the fundamental moment on the specified function of parameters and exogenous variables. But this significance test is itself a test of a moment condition, called a "regression-based" CM test. It generalizes the work of Cameron and Trivedi (1990a), who used this test in the context of testing variance-mean equality, a property of the Poisson model.

Like any CM test, implementation of a regression-based CM test requires replacing unobserved parameters by estimates of the parameters of the null hypothesis model. In a number of leading examples, asymptotic results are unchanged by this replacement, so that the significance tests mentioned in the previous paragraph can be performed without modification. This permits direct tests for the significance of subcomponents of the alternative hypothesis. Thus the regression is more than just an auxiliary regression.

In other examples the replacement of parameters by estimates affects the asymptotic distribution. We can then appeal to the general theory of CM tests. Different estimators and different moment conditions lead to different limit distributions of the CM test statistic. The corresponding chi-square test statistics can then be computed by one or more different auxiliary regressions. White (1990) gives a very general treatment, that offers alternatives to using the outer-product-of-the-gradient regression in many common testing situations. Wooldridge (1990a) has developed a procedure to transform the CM test statistic to one, not necessarily equivalent, for which asymptotic theory is valid under relatively weak distributional assumptions.

The regression-based CM test approach decomposes CM tests into two components: the fundamental moment (zero under the null hypothesis) and its parameterization under the alternative hypothesis. The choice of fundamental moment is considered according to whether or not the conditional distribution of the dependent variable is fully parameterized.

If distributional assumptions are to be minimized, it seems natural to let the fundamental moment be the expectation of a function of the central moments of the dependent variable, or equivalently of the error term. In particular, we propose new tests of misspecification of the conditional central moments of the dependent variable, i.e. heteroskedasticity, skewness, and kurtosis.

When the conditional density under the null hypothesis is specified, standard specification tests are the lagrange multiplier (LM), Hausman (H) and information matrix (IM) tests. These tests can be interpreted as regression-based CM tests. For the LM and H tests, the fundamental moment condition is the expectation of a subcomponent of the score vector of the null hypothesis model, while for the IM test the fundamental moment is the expectation of a subcomponent of the sum of the outer product and the derivative of the score vector.

For the LM test, interpretation as a regression-based CM test should not be surprising, since there is usually a moment restriction whose imposition gives the null hypothesis density as a special case of the alternative hypothesis density. However, the regression-based CM test approach is more direct, less parametric, and often considerably easier analytically than the LM test approach. Furthermore, the two approaches frequently give asymptotically equivalent test statistics. To the extent that this happens, the regression based CM test is asymptotically equivalent not only to the LM test, but also to the Wald and Likelihood Ratio tests.

For the IM and H tests, equivalence to a regression-based CM test permits interpretation of the IM and H tests as being a test of a conditional moment against a specific alternative. This re-interpretation of these tests overcomes one of their perceived weaknesses. It also suggests a wider class of IM tests than generally used.

When the conditional density is specified, we also present a new choice of fundamental moment condition, that based on orthogonal polynomial systems. This new theory is related to LM and IM tests in the case of quasi-maximum likelihood estimation for the linear exponential family with quadratic variance function.

The general theory of regression-based CM tests is given in section 2, and contrasted with the standard formulation of CM tests. Direct regressions and auxiliary regressions to implement regression-based CM tests are presented in section 3. Examples are given in section 4. These cover many common testing situations, are easily motivated and implemented, and include some new tests. Distributional assumptions are minimized in sections 2-4. In section 5, the regression-based CM test approach is compared with other principles used to obtain CM tests, including Lagrange multiplier, Hausman and information matrix tests, for models where the conditional density under the null hypothesis is specified. A new class of CM tests, that based on orthogonal polynomial systems is also presented. An application to the linear exponential family with quadratic variance function is given in section 6. Some concluding remarks are made in section 7.

2. Regression-Based Tests for Model Specification

2.1 *Conditional Moment Tests*

In regression analysis, we are interested in explaining dependent variables, a vector y_t, conditional on explanatory variables, a vector X_t. For simplicity, this paper focuses on cross-section data, where the data $\{(y_t, X_t), t = 1,\ldots,T\}$ are independent across t. The approach can be extended to dynamic models, where X_t is the vector of current and lagged values of the exogenous variables and lagged values of the dependent variables, and (y_t,X_t) are no longer independent across t.

The true data generating process (d.g.p.) for y given X is unknown. Instead, statistical inference is based on an assumed parameterized density function (quasi-ML estimation) or an assumed parameterization of moments (GMM estimation, e.g. least squares). Conditional moment tests are tests of the validity of moment conditions implied by these assumed parameterizations.

Specifically, a conditional moment test is any test based on an $s\text{x}1$ vector of functions $m(y,X,\theta)$, where θ is a $q\text{x}1$ vector of parameters, that satisfies the moment condition:

$$(2.1)\qquad \mathbb{E}_0[m(y_t,X_t,\theta) \mid X_t] = 0 \quad ,$$

where the subscript 0 denotes expectation with respect to an assumed model, not necessarily the true d.g.p.

Tests based on a moment condition of the form (2.1), henceforth called CM tests, were introduced by Newey (1985) and Tauchen (1985). Further results are given by Pagan and Vella (1989), White (1987, 1990) and Wooldridge (1990a). The simplest version of a CM test based on (2.1) uses the corresponding sample moment:

$$(2.2) \qquad m_T(\theta) = T^{-1} \sum_{t=1}^{T} m(y_t, X_t, \theta) \quad .$$

To operationalize a CM test, the parameter θ in (2.2) is replaced by an estimator $\hat{\theta}_T$, consistent under the maintained model. CM specification tests are statistical tests of the departure of $m_T(\hat{\theta}_T)$ from zero.

The concern of this paper is CM testing. This has two dimensions: the choice of moment to use in (2.2), and the effect of replacing θ in (2.2) by an estimator. We begin with the first of these.

2.2 *Regression-Based Conditional Moment Tests*

It is assumed that the process generating the data is such that the $n \times 1$ vector of functions $r(y,X,\theta)$, satisfies the "fundamental moment condition":

$$(2.3) \qquad H_0: \mathbb{E}_0[r(y_t, X_t, \theta) \mid X_t] = 0 \quad ,$$

where the subscript 0 denotes expectation with respect to the null hypothesis model. This moment condition is "fundamental" in the sense that the dimension of $r(\cdot)$ in (2.3) is generally considerably less than the dimension of $m(\cdot)$ in (2.1), as should be clear from the ensuing discussion. In fact, $n = 1$ in most commonly used tests, but for completeness we give the theory for the more general case. Sections 4 and 5 will discuss at some length the choice of fundamental moment condition.

Suppose that were the expectation in (2.3) to be taken with respect to the true distribution, (2.3) would no longer hold. In particular, we wish to test against the alternative that for the j-th moment condition:

$$H_1: \mathbb{E}_1[r_j(y_t, X_t, \theta) \mid X_t] = g_j(X_t, \theta) \cdot \alpha_j \quad , \qquad j = 1, \ldots, n,$$

where the subscript 1 denotes expectation with respect to the true d.g.p., $g_j(X_t,\theta)$ = is a specified $1xp_j$ vector function, and α_j is a p_jx1 vector of additional unknown parameters. Continuing the earlier example, the alternative moment condition may be that the error has non-zero mean due to omitted variables $g(X_t,\theta)$. Combining all n moment conditions we test H_0 against the alternative that:

$$(2.4) \qquad H_1\colon \mathbb{E}_1[r(y_t,X_t,\theta) \mid X_t] = G(X_t,\theta)\cdot\alpha \quad ,$$

where $G(X_t,\theta)$ is a nxp matrix whose j-th row has g_j in columns $(p_1+ .. + p_{j-1}+ 1)$ to $(p_1+ .. + p_j)$ and zeroes elsewhere, $p = (p_1+ .. + p_n)$, and $\alpha = (\alpha_1', .. ,\alpha_n')'$ is a px1 parameter vector.

Tests of the moment condition under H_0 against that under H_1 are tests of whether $\alpha = 0$. The obvious basis for such a test is the estimated coefficient of α in the following multivariate regression:

$$(2.5) \qquad r(y_t,X_t,\theta) = G(X_t,\theta)\cdot\alpha + \varepsilon_t,$$

where the sx1 heteroscedastic error term ε_t is defined by

$$(2.6) \qquad \varepsilon_t = r(y_t,X_t,\theta) - E[r(y_t,X_t,\theta) \mid X_t] \quad .$$

Weighted least squares estimation of (2.5) with sxs symmetric weighting matrix $W(X_t,\theta)$ yields the usual weighted least squares estimator:

$$(2.7) \quad \hat{\alpha} = \{ \sum_{t=1}^{T} G(X_t,\theta)'\cdot W(X_t,\theta)\cdot G(X_t,\theta) \}^{-1}\cdot \sum_{t=1}^{T} G(X_t,\theta)'\cdot W(X_t,\theta)\cdot r(y_t,X_t,\theta) \; .$$

For the purposes of statistical inference, tests based on $\hat{\alpha}$ are determined by the distribution of:

$$(2.8) \qquad m_{\alpha T}(\theta) = \sum_{t=1}^{T} G(X_t,\theta)'\cdot W(X_t,\theta)\cdot r(y_t,X_t,\theta) \quad .$$

From section 2.1, this is a CM test based on the moment condition:

$$(2.9) \qquad \mathbb{E}_0[G(X_t,\theta)'\cdot W(X_t,\theta)\cdot r(y_t,X_t,\theta) \mid X_t] = 0 \quad .$$

This test is called a "regression-based" CM test, or RBCM test, since this form of the CM test is motivated by the regression of the fundamental moment on to its parameterization under the alternative. This should not be confused with the implementation of a CM test by an auxiliary regression. It is in this latter sense that other authors have used, or perhaps misused, the term "regression-based" in the context of testing. In section 3 we show that sometimes the RBCM test can be implemented directly from the regression (2.5), while at other times an auxiliary regression may be necessary.

RBCM tests were introduced by Cameron and Trivedi (1990a), in the context of testing variance-mean equality. Then in (2.5), y_t is scalar, $r(y_t,X_t,\theta) = \{(y_t - \mu_t(X_t,\theta))^2 - y_t\}$, where $\mu_t(X_t,\theta) = \mathbb{E}_0[y_t \mid X_t]$, and $G(X_t,\theta) = g(\mu_t(X_t,\theta))$ for specified scalar function $g(\cdot)$.

2.3 *Optimal Regression-Based CM Test*

Different choices of the weighting matrix lead to different test statistics. The optimal test within the class of RBCM tests will be that based on the most efficient estimator of α. This is the generalized least squares estimator:

$$(2.10) \qquad m_{\alpha T,opt}(\theta) = \sum_{t=1}^{T} G(X_t,\theta)' \cdot \Sigma(X_t,\theta)^{-1} \cdot r(y_t,X_t,\theta) \quad ,$$

where we make the additional assumption:

$$\begin{aligned}(2.11) \qquad & \mathbb{E}_0[\varepsilon_t \varepsilon_t' \mid X_t] \\ &= \mathbb{E}_0[r(y_t,X_t,\theta) \cdot r(y_t,X_t,\theta)' \mid X_t] \\ &= \Sigma(X_t,\theta) \quad ,\end{aligned}$$

for specified variance function $\Sigma_t = \Sigma(X_t,\theta)$. So the optimal RBCM test for (2.3) against alternatives of the form (2.4) is a CM test based on:

$$(2.12) \qquad \mathbb{E}_0[G(X_t,\theta)' \cdot \Sigma(X_t,\theta)^{-1} \cdot r(y_t,X_t,\theta) \mid X_t] = 0.$$

Specification of Σ_t usually requires distributional assumptions for the null hypothesis model additional to the minimal assumptions needed for (2.3). For example, in tests of omitted regressors in the classical regression model, the usual additional assumption is constancy of the error variance. Such additional distributional assumptions are not critical, since it is possible to construct tests that are asymptotically valid even if Σ_t is not fully specified..

2.4 *Discussion*

The RBCM tests differ from conventional CM specification tests in that they are tests against an explicit alternative, given in (2.4). In this subsection the two approaches are compared.

It follows immediately from section 2.3 that any CM test based on the moment condition

$$(2.13) \qquad \mathbb{E}_0[G^*(X_t,\theta)'\cdot r^*(y_t,X_t,\theta) \mid X_t] = 0 ,$$

where G^* is a n^*xp^* matrix and r^* is a n^*x1 vector, is a RBCM test of:

$$(2.14) \qquad H_0\text{: } \mathbb{E}_0[r^*(y_t,X_t,\theta) \mid X_t] = 0 ,$$

against the alternative hypothesis

$$(2.15) \qquad H_1\text{: } \mathbb{E}_1[r^*(y_t,X_t,\theta) \mid X_t] = (W^*(X_t,\theta))^{-1}\cdot G^*(X_t,\theta)\cdot\alpha^* ,$$

where $W^*(X_t,\theta)$ is the n^*xn^* weighting matrix used in the regression, and α^* is a n^*x1 parameter vector. (2.13) corresponds to the optimal RBCM test of (2.14) against (2.15) when $W^*(X_t,\theta) = \mathbb{E}_0[r_t^*\cdot r_t^{*\prime} \mid X_t]$.

Note that (2.13) is exactly of the form of the CM test of Newey (1985). In Newey's framework $G^*(X_t,\theta)$ is an arbitrarily chosen matrix of functions, and $r^*(y_t,X_t,\theta)$ satisfies (2.14). The motivation is that if (2.14) holds, then by the law of iterated expectations, (2.13) holds for a wide range of choices of $G^*(X_t,\theta)$. Little guidance is given for the choice of $G^*(X_t,\theta)$, except in the case where the distribution of y_t is specified under both H_0 and H_1, and the score under H_0 is multiplicative in $r^*(y_t,X_t,\theta)$. In this case, Newey gives the optimal choice of $G^*(X_t,\theta)$. By contrast, the optimal RBCM test interpretation imposes less structure, and provides a direct approach to choosing the optimal $G^*(X_t,\theta)$, i.e. specify the variance of $r^*(y_t,X_t,\theta)$ under H_0 and the mean of $r^*(y_t,X_t,\theta)$ under H_1.

Pagan and Vella (1989) also use the form (2.13). Tauchen (1985) considers the simplest case where $G^{*}(X_t,\theta) = 1$. White (1987, 1990) and Wooldridge (1990a) call r_t^{*} "generalized residuals" and G_t^{*} "misspecification indicators". The terminology generalized residual is used following Cox and Snell (1968), though the examples below suggest that a better terminology might be functions of the residual. The RBCM framework provides a natural way to select, or to interpret, the misspecification indicators. The choice of generalized residual, or equivalently of fundamental moment, is still an open issue.

Wooldridge (1990a) factorizes (2.13) further to:

$$(2.16) \qquad \mathbb{E}_0[G^{*}(X_t,\theta)'\cdot W^{*}(X_t,\theta)\cdot r^{*}(y_t,X_t,\theta)] \;=\; 0 \;.$$

Wooldridge specializes to this form because many specification tests are of this form. From section 2.3, (2.16) is the optimal RBCM test of (2.14) against

$$(2.17) \qquad H_1\colon\; \mathbb{E}_1[r^{*}(y_t,X_t,\theta)' \mid X_t] \;=\; G^{*}(X_t,\theta)\cdot\alpha^{*} \;,$$

if $W^{*}(X_t,\theta) \;=\; \mathrm{Var}(r^{*}(y_t,X_t,\theta) \mid X_t)$. So the RBCM test approach additionally provides an interpretation of the weighting function in (2.16).

Standard CM tests can be interpreted as RBCM tests. Strictly speaking, hypothesis tests merely reject or do not reject the null hypothesis. However, specifying an alternative hypothesis provides a very direct way to construct CM tests for particular forms of misspecification of a fundamental moment, as illustrated in section 4. The relationship between (2.13) and (2.15) or (2.14) and (2.17) can be used to interpret and contrast many standard tests, as done in section 5. And for a narrow but widely-used class of models, CM tests can be implemented directly by running the regression (2.5) of the RBCM test.

2.5 *Locally Equivalent Alternatives*

In principle, under the alternative hypothesis the right hand side of (2.4) may be nonlinear in α and g. For simplicity, consider the standard case of one fundamental moment condition:

$$(2.18) \qquad H_1\colon\ \mathbb{E}_1[r(y_t,X_t,\theta) \mid X_t] = h(X_t,\theta,\beta) \quad ,$$

where $h(X_t,\theta,\beta{=}\beta^*) = 0$. Then by first-order Taylor series expansion about $\beta = \beta^*$, $h(X_t,\theta,\beta^*) = 0 + \nabla_\beta h(X_t,\theta,\beta^*)\cdot(\beta - \beta^*)$, where $\nabla_\beta h(X_t,\theta,\beta^*)$ denotes the derivative of $h(X_t,\theta,\beta)$ w.r.t. β, evaluated at $\beta = \beta^*$. But this is of the form (2.4). The remainder term in the Taylor series expansion will disappear asymptotically for $\beta - \beta^* = O_p(T^{-1/2})$. Thus at least to local alternatives the linear form (2.4) nests nonlinear alternatives of the form (2.18). We use (2.4) as it is simpler, but could use (2.18) and estimate the corresponding regression by nonlinear, rather than linear, least squares.

Locally equivalent tests are discussed in Godfrey (1988). An example given in section 4 is the use of locally equivalent alternatives to transform specification tests for the conditional mean to an omitted variables problem.

3. IMPLEMENTATION OF REGRESSION-BASED CONDITIONAL MOMENT TESTS

The RBCM test approach can always be used to motivate and/or interpret the moment condition tested in a CM test. To implement RBCM tests, the parameter vector θ in (2.5) needs to be replaced by an estimate. The RBCM test is simplest to use when this replacement does not effect the asymptotic distribution of the test.

3.1 Tests when the RBCM Test Regression can be directly implemented

The estimator $\hat{\alpha}$ in (2.7) can be viewed as maximizing, with respect to α, the quasi-likelihood function:

$$(3.1) \qquad -Q_T(\alpha,\theta) = \sum_{t=1}^{T} W(X_t,\theta)\cdot\{r(y_t,X_t,\theta) - G(X_t,\theta)\cdot\alpha\}^2 .$$

To implement RBCM tests, we need to replace θ by an estimate. It is known, e.g. Pagan (1986) or White (1990), that minimization of $Q_T(\alpha,\hat{\theta})$, where $\hat{\theta}$ is consistent for θ_0, the true value of θ, yields an estimator $\hat{\alpha}$ which has the same asymptotic properties as the estimator $\tilde{\alpha}$ which minimizes $Q_T(\alpha,\theta_0)$, if

$$(3.2) \qquad E_0\left[\frac{\partial^2 Q_T(\alpha_0,\theta_0)}{\partial\alpha\cdot\partial\theta} \mid X_t\right] = 0 .$$

For the quasi-likelihood function (3.1), this condition will be satisfied under H_0 (or local alternatives to H_0), if in addition to the fundamental moment condition (2.3), the following moment condition is satisfied:

$$(3.3) \qquad \mathbb{E}_0[\nabla_\theta r(y_t,X_t,\theta) \mid X_t] = 0 .$$

This condition implies zero asymptotic covariance between the moment criterion function $T^{1/2}r(y_t,X_t,\hat{\theta})$ and $T^{1/2}(\hat{\theta} - \theta_0)$ which are assumed to be asymptotically jointly normally distributed. (Pierce (1982, p.478). Models for which this condition holds are presented in sections 4 and 6.

When (3.3) holds, the asymptotic theory of section 2 is unchanged by replacing θ_0 by $\hat{\theta}_T$. Therefore we can test H_0 in (2.3) against H_1 in (2.4) by testing the significance of the weighted least squares estimator of α in the regression:

$$(3.4) \qquad r(y_t,X_t,\hat{\theta}_T) = G(X_t,\hat{\theta}_T)\cdot\alpha + u_t,$$

with weighting matrix $W_t(X_t,\hat{\theta}_T)$, and apply the usual theory of weighted least squares treating $r(y_t,X_t,\hat{\theta}_T)$ as a regular vector dependent variable and $G(X_t,\hat{\theta}_T)$ as a regular matrix of regressors. We have:

$$(3.5) \qquad \hat{\alpha} = \{ \sum_{t=1}^{T} \hat{G}_t'\hat{W}_t\hat{G}_t \}^{-1}\cdot \sum_{t=1}^{T} \hat{G}_t'\hat{W}_t\hat{r}_t \quad .$$

where $\hat{G}_t = G(X_t,\hat{\theta}_T)$, $\hat{W}_t = W(X_t,\hat{\theta}_T)$, and $\hat{r}_t = r(y_t,X_t,\hat{\theta}_T)$. We consider the limit distribution of $T^{1/2}\cdot\hat{\alpha}_W$ under local alternatives H_L: $\alpha = T^{-1/2}\gamma$, or more formally

$$(3.6) \qquad H_L:\ E[r(y_t,X_t,\theta) \mid X_t] = G(X_t,\theta)\cdot(T^{-1/2}\gamma),$$

where γ is a finite vector. Then under H_L:

$$(3.7) \qquad T^{1/2}\cdot\hat{\alpha}_W \xrightarrow{d} N(\gamma,\ [\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t'W_tG_t]^{-1}\cdot [\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t'W_t\Omega_tW_tG_t] \cdot [\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t'W_tG_t]^{-1} \quad ,$$

where $G_t = G(X_t,\theta_0)$, $W_t = W(X_t,\theta_0)$, $r_t = r(y_t,X_t,\theta_0)$, and $\Omega_t = \mathbb{E}[r_tr_t' \mid X_t]$ is the unspecified conditional variance of r_t.

Specializing to H_0, (3.7) yields t-tests for individual components of α equalling zero. A consistent estimate of the variance-covariance matrix in (3.7) replaces G_t by $\hat{G}_t$, W_t by $\hat{W}_t$, and Ω_t by $\hat{\Omega}_t$, where $\hat{\Omega}_t$ is such that $\operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T} \hat{G}_t'\hat{W}_t\hat{\Omega}_t\hat{W}_t\hat{G}_t = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t'{}_t\Omega_tW_tG_t$. We use the obvious estimator $\hat{\Omega}_t = \sum_{t=1}^{T} \hat{r}_t\hat{r}_t'$.

A joint test of whether all components of α equal zero is given by the chi-square test statistic:

$$(3.8) \quad d_w = [\sum_{t=1}^{T} \hat{r}_t'\hat{W}_t\hat{G}_t] \cdot [\sum_{t=1}^{T} \hat{G}_t'\hat{W}_t\hat{\Omega}_t\hat{W}_t\hat{G}_t]^{-1} \cdot [\sum_{t=1}^{T} \hat{G}_t'\hat{W}_t\hat{r}_t] .$$

Under H_0, d_w is asymptotically chi-square distributed with p degrees of freedom.

Implementation of these tests in principle requires a multivariate regression package and matrix multiplication routines. When $n = 1$, however, we need only use an instrumental variables package along the lines suggested by Domowitz (1983) for inference based on heteroskedastic consistent estimators of the variance-covariance matrix of the least squares estimator.

Under H_L, d_w is asymptotically distributed as non-central chi-square with noncentrality parameter:

$$(3.9) \qquad \lambda_w = \gamma^2 \cdot V_{G*G*}V_{G**G**}V_{G*G*}$$

where $V_{G*G*} = [\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t'W_tG_t]^{-1}$, $V_{G**G**} = [\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t'W_t\Omega_tW_tG_t]$.

The optimal RBCM test based on (2.5) uses weights W_t that maximize the local power by maximizing λ_w in (3.9). As expected, this is $W_t = \Omega_t^{-1}$, the generalized least squares estimator. Thus, we assume that under H_0, $\Omega_t = \Sigma_t$, defined in (2.11), which entails stochastic assumptions in addition to (2.3) and (3.3). We obtain the test statistic:

$$(3.10) \quad d_w^{opt} = [\sum_{t=1}^{T} \hat{r}_t'\hat{\Sigma}_t^{-1}\hat{G}_t] \cdot [\sum_{t=1}^{T} \hat{G}_t'\hat{\Sigma}_t^{-1}\hat{G}_t]^{-1} \cdot [\sum_{t=1}^{T} \hat{G}_t'\hat{\Sigma}_t^{-1}\hat{r}_t] .$$

This optimal test is easily computed as the square root of the explained sum of squares from the OLS regression

$$(3.11)\quad \hat{\Sigma}_t^{-1/2}\hat{r}_t = \hat{\Sigma}_t^{-1/2}\hat{G}_t \cdot \alpha + u_t \ .$$

This regression can also be used to test whether individual components of α equal 0, using the usual t-tests. It is the regression that is directly suggested by the theory of section 2, and is viewed as the regression that is the basis for the test, rather than an auxiliary regression used to compute a test statistic.

The Pitman relative efficiency of d_w relative to d_w^{opt} is given by

$$\mathcal{E}ff\left[\frac{d_w}{d_w^{opt}}\right] = \frac{V_{G*G*}V_{G**G**}V_{G*G*}}{V_{G^+G^+}}$$

where $V_{G^+G^+} = [\lim T^{-1} \sum_{t=1}^{T} G_t'\Omega_t^{-1}G_t]^{-1}$. This shows the reduction in Pitman relative effciency due to the use of the robust rather than the optimal version of the test.

The optimal RBCM test does not require specification of the conditional distribution of y under H_0. However, it does require sufficient assumptions on the distribution of y to determine the second moment of $r(y_t,X_t,\theta)$, while the less powerful tests based on the weighted least squares regression (3.4) can be implemented using only (3.3) and the initial assumption that $r(y_t,X_t,\theta)$ has zero first moment.

To guard against the possibility of misspecification of Σ_t, we can of course compute asymptotically valid versions of the individual t-tests , using the result (3.7) with $\hat{W}_t = (\hat{\Sigma}_t)^{-1}$.

Furthermore it can be shown that (3.8) is chi-square distributed if $\hat{W}_t$

$= \hat{\Omega}_t$ (a non-trivial result since $\hat{W}_t$ then depends on r_t directly as well as indirectly via $\hat{\theta}_T$). Therefore, T times the uncentered R^2 from the univariate regression of the Tx1 vector of ones on $\hat{G}_t' \hat{W}_t \hat{r}_t$ is asymptotically distributed as chi-square under H_0. This regression, unlike regressions (3.4) or (3.11), is an auxiliary regression.

3.2 Tests when the RBCM Test Regression cannot be directly implemented

Most of the tests in this paper deliberately choose $r(y_t, X_t, \theta)$ so that (3.3) is satisfied. This is a departure from previous studies, aside from that of White (1990), who also exploits the simplification that arises when a condition similar to (3.3) is satisfied.

For tests of misspecification of the conditional mean of y_t, given in section 4.1, (3.3) will not hold. A separate theory for RBCM tests of the conditional mean can be developed, permitting inference on the estimated coefficients from the regression of $r(y_t, X_t, \hat{\theta}_T)$ on $G(X_t, \hat{\theta}_T)$, assuming that $\hat{\theta}_T$ is a weighted least squares estimator or a maximum likelihood estimator. Instead, we consider the joint test of significance of all coefficients, and appeal to the general theory of CM tests.

The general CM test (2.2) is implemented by using an estimator $\hat{\theta}_T$ of θ, consistent under H_0, to form $m_T(\hat{\theta}_T)$. Under appropriate assumptions, $T^{1/2} m_T(\hat{\theta}_T)$ has a limit distribution that is multivariate normal, which is the basis for a chi-square test of the null hypothesis moment condition. This test statistic can often be computed by an auxiliary regression.

The relevant theory is given in great detail in White (1990). A nice exposition is given in Pagan and Vella (1989). For any moment condition, under suitable conditions, by a first-order Taylor series expansion:

$$(3.12) \qquad T^{1/2} m_T(\hat{\theta}_T) = T^{1/2} m_T(\theta_0) + B_0 \cdot T^{1/2}(\hat{\theta}_T - \theta_0) + o_p(1) \; .$$

where $\theta_0 = \text{plim } \hat{\theta}_T$ and $B_0 = \text{plim } \nabla_\theta m_T(\theta)$. The implementation of CM tests differs according to whether or not the following condition is satisfied:

$$(3.13) \qquad \mathbb{E}_0[\nabla_\theta m(y_t, X_t, \theta) \mid X_t] = 0 \quad .$$

When the conditional moment $m_t \equiv m(y_t, X_t, \theta)$ is chosen so that (3.13) is satisfied, under H_0, $B_0 = 0$, so that the asymptotic distribution of $m_T(\hat{\theta}_T)$ coincides with that of $m_T(\theta_0)$. Since $\mathbb{E}_0[m_t \mid X_t] = 0$, and m_t are assumed independent over t, a central limit theorem yields under H_0,

$$(3.14) \qquad T^{1/2} m_T(\hat{\theta}_T) \overset{d}{\rightarrow} N(0, \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} m_t m_t') \quad ,$$

so that under H_0,

$$(3.15) \qquad (\sum_{t=1}^{T} \hat{m}_t') \cdot [\sum_{t=1}^{T} \hat{m}_t \hat{m}_t']^{-1} \cdot (\sum_{t=1}^{T} \hat{m}_t) \overset{d}{\rightarrow} \chi^2(n) \quad ,$$

where $\hat{m}_t \equiv m(y_t, X_t, \hat{\theta}_T)$.

Three things should be noted about this test statistic. First, it can be conveniently computed as T times the uncentered R^2 from the auxiliary regression of 1 on $\hat{m}_t$. Second, it is relatively robust in that the only assumptions are that (2.2) and (3.13) hold under H_0. Third, we note that (3.3) implies (3.13), and that the test statistic (3.15) and the corresponding auxiliary regression are the same as those for the RBCM test mentioned at the end of the previous subsection.

The condition (3.3) may not be satisfied when, for example, there are nuisance parameters in the p.d.f. If (3.3) does not hold, implying (3.13) also does not hold, the asymptotic distribution of $m_T(\hat{\theta}_T)$ can still be obtained, but will in general differ from that of $m_T(\theta_0)$ and will vary with the choice of consistent estimator $\hat{\theta}_T$. Furthermore, the corresponding chi-square statistic can no longer be simply estimated from an auxiliary regression, though other alternatives similar to (3.4) may be available in certain special cases; see White (1990).

By restricting attention to certain classes of estimators and/or certain classes of conditional moments, a chi-square statistic can again be computed from a convenient auxiliary regression, even if (3.13) does not hold. White (1990) presents a number of auxiliary regressions. These regressions are applicable to a narrow range of situations, e.g. the Gauss-Newton regression, or are more general but have poor empirical performance, e.g. the Outer Product of the Gradient (OPG) regression used extensively by Tauchen (1985), Newey (1985) and Pagan and Vella (1989). White's results do indicate, however, that these authors sometimes use the OPG regression when other, potentially better, auxiliary regressions might be used.

When the sample moment $m_T(\theta)$ is of the form (2.8) but does not satisfy (3.13), Wooldridge (1990a) has proposed a quite general method to transform $m_T(\theta)$ to a sample moment $m_T^*(\theta)$ which does satisfy this condition, so that we can then appeal to the simpler theory. His tests are "robust", in the sense that they are asymptotically valid provided only that (2.3) holds, and "regression-based" in the sense that they can be computed from one or two auxiliary regressions. His results are especially useful for CM tests of the conditional mean in models estimated by quasi-maximum likelihood using a density in the linear exponential family with nuisance parameters; see Wooldridge (1990b). In other cases, however, tests using $m_T^*(\theta)$ may differ, even asymptotically, from tests using $m_T(\theta)$. In the next section, with the exception of conditional mean tests, we focus on examples where (3.3) holds, in which case the transformation proposed by Wooldridge is not necessary.

4. Application to Specification Tests of Central Moments

The regression-based CM test approach essentially splits the choice of moment condition into two pieces: choice of the fundamental moment and choice of its expectation under the alternative. A corresponding decomposition into generalized residual and misspecification indicator is used in the standard CM test approach. In either case, choice of the fundamental moment condition is discretionary.

When the conditional density is specified, testing principles such as the LM and IM determine the fundamental moment. This case is considered in section 5. In this section, we instead consider the case where distributional assumptions are to be minimized. It is then natural to focus on tests of the first few conditional moments of the dependent variable, since inference, such as OLS and GMM estimation, is typically based on these. Where possible, the fundamental moment is chosen to satisfy (3.3). For simplicity, the dependent variable is scalar.

The examples given here can be viewed as generalizing the tests of misspecification of the mean and variance in the linear regression model with constant variance given in Pagan and Hall (1983) and Pagan (1984). However, in generalizing their work, choosing the fundamental moment to satisfy (3.3) leads to different CM tests than does using moment conditions to be the direct analogues of those for the linear regression model with constant variance.

4.1 *Tests for Misspecified Conditional Mean*

In the null hypothesis model, $\mathbb{E}_0[y_t \mid X_t] = \mu(X_t,\theta)$, for some scalar function μ. We therefore consider RBCM tests of:

$$(4.1) \qquad H_0\colon\ \mathbb{E}_0[(y_t - \mu(X_t,\theta)) \mid X_t] = 0$$

against

$$(4.2) \qquad H_1\colon\ \mathbb{E}_1[(y_t - \mu(X_t,\theta)) \mid X_t] = g(X_t,\theta)\cdot\alpha \ .$$

Here the choice of fundamental moment, $r(y_t,X_t,\theta) = (y_t - \mu(X_t,\theta))$, is obvious. Tests for different types of misspecification of the mean correspond to different choices of its expectation, i.e. the additional variables in $g(X_t,\theta)$, under the alternative hypothesis. Two standard forms of misspecification are omitted variables and misspecified functional form. To illustrate, we specialize to the common case where $\mu(X_t,\theta)$ equals $\mu(X_t'\theta)$.

For omitted variables we have:

$$H_1\colon\ \mathbb{E}_1[y_t \mid X_t] = \mu(X_t'\theta + Z_t'\alpha) \ .$$

By a first order Taylor series expansion of the right-hand side about $\alpha = 0$, and rearranging, locally equivalent tests are based on the moment condition:

$$(4.3) \qquad H_1\colon\ \mathbb{E}_1[y_t - \mu(X_t'\theta) \mid X_t] = \nabla_\alpha\mu(X_t'\theta)\cdot Z_t\cdot\alpha \quad ,$$

where $\nabla_\alpha\mu(\cdot)$ denotes the derivative of $\mu(\cdot)$. In this example we choose $g(X_t'\theta) = \nabla_\alpha\mu(X_t'\theta)\cdot Z_t$.

For incorrect functional form it is typically assumed that:

$$(4.4) \qquad H_1\colon \ \mathbb{E}_1[y_t \mid X_t] \ = \ \mu_\alpha(X_t'\theta,\alpha) \ ,$$

where $\mu(X_t'\theta,\alpha)$ evaluated at $\alpha = 0$, say, equals $\mu(X_t'\theta)$. By a first order Taylor series expansion about $\alpha = 0$, locally equivalent tests are based on the moment condition:

$$(4.5) \qquad H_1\colon \ \mathbb{E}_1[y_t - \mu(X_t'\theta) \mid X_t] \ = \ \nabla_\theta\mu_\alpha(X_t'\theta)\cdot\alpha \ ,$$

where $\nabla_\theta\mu_\alpha(X_t'\theta)$ denotes the derivative of $\mu_\alpha(X_t'\theta,\alpha)$ w.r.t. θ evaluated at α=0.

For any test of the conditional mean, the optimal RBCM test (2.10) will require a second moment assumption on y_t.

The above moment conditions for CM tests of the conditional mean are motivated without appeal to a conditional density. When a density is specified, examples of CM tests of misspecification of the mean are given in Pagan and Hall (1983) for the linear regression model under normality, in Newey (1985) for the probit model, and in Pagan and Vella (1989) for the tobit, probit and some duration models. These CM tests are often presented as tests of the assumed density, but in fact may only be tests of the conditional mean. In particular, any model for discrete choice data must have the property that all moments are uniquely determined by the the mean, so specification tests will always be equivalent to tests of the mean. Thus so-called tests for "heteroscedasticity" and "nonnormality" in the distribution of the underlying latent variable for the probit model are best viewed as tests for a misspecified mean. For discrete choice models, the optimal RBCM test should always be used as it entails no additional assumptions: the variance of y_t is always the product of the conditional mean and unity less the conditional mean. Furthermore, the tests of Newey (1985, p.1062) can be very simply derived as optimal RBCM tests.

For conditional mean tests, an auxiliary regression is always needed to implement the joint test of $\alpha = 0$, since

$$(4.6) \qquad \mathbb{E}_0[\ \nabla_\theta(y_t - \mu(X_t,\theta))\ |\ X_t] \ = \ \nabla_\theta\mu(X_t,\theta) \ \neq \ 0 \quad ,$$

so that (3.3) will not be satisfied. Implementation of CM tests will therefore vary according to the estimator of θ. When estimation is based on a quasi-likelihood in the linear exponential family, in which case a weighted least squares estimator of θ is used, the results of Wooldridge (1990b) are directly applicable.

4.2 *Tests for Misspecified Variance*

We assume that (4.1) holds under both the null and alternative hypotheses, i.e. the conditional mean is correctly specified. However, under the alternative hypothesis the conditional variance is misspecified. We consider the case where the variance under the null hypothesis is a function of the mean, not necessarily constant. We have H_0: $\mathbb{E}_0[(y_t - \mu(X_t,\theta))^2 \ |\ X_t] = V(\mu(X_t,\theta))$. It should be understood that throughout this section, additionally $\mathbb{E}[y_t - \mu(X_t,\theta)\ |\ X_t] = 0$ under both H_0 and H_1.

The obvious fundamental moment condition is:

$$(4.7) \qquad H_0\text{: } \mathbb{E}_0[(y_t - \mu(X_t,\theta))^2 \ - \ V(\mu(X_t,\theta))\ |\ X_t] \ = \ 0 \quad ,$$

and to test this against the alternative that the expectation equals $g(X_t,\theta)\cdot\alpha$.

This is the fundamental moment used in tests of heteroskedasticity in the regression model with constant variance. Then $V(\mu(X_t,\theta)) = \sigma^2$, and tests of heteroskedasticity are presented as tests against H_1: $\mathbb{E}_1[(y_t - \mu(X_t,\theta))^2 \mid X_t] = h(X_t,\alpha)$, where $h(X_t,\alpha=0) = \sigma^2$. But a locally equivalent alternative is H_1: $\mathbb{E}_1[(y_t - \mu(X_t,\theta))^2 - \sigma^2 \mid X_t] = \nabla_\alpha h(X_t,\alpha=0)\cdot\alpha$, and indeed one way to compute the Breusch-Pagan test for heteroskedasticity is as the test of significance in the regression of $(y_t - \mu(X_t,\theta))^2$ on $\nabla_\alpha h(X_t,\alpha=0)$.

When $V(\mu(X_t,\theta))$ is constant, (4.7) is a convenient choice of fundamental moment since then (3.3) is satisfied. But when $V(\mu(X_t,\theta))$ varies with θ, (3.3) will no longer hold. For example, for the Poisson regression model, where $V(\mu(X_t,\theta)) = \mu(X_t,\theta)$, Cameron and Trivedi (1990a), propose RBCM tests based on the fundamental moment condition $\mathbb{E}_0[(y_t - \mu(X_t,\theta))^2 - y_t \mid X_t] = 0$. This coincides with the leading examples of LM tests in this situation.

More generally, Cameron (1990) proposes regression-based CM tests for the variance-mean relationship (4.7) based on the fundamental moment condition:

$$(4.8)\qquad H_0\text{: } \mathbb{E}_0[\{(y_t - \mu_t)^2 - V(\mu_t)\} + \nabla_\mu V(\mu_t)\cdot(y_t - \mu_t) \mid X_t] = 0 \quad ,$$

where $\mu_t = \mu(X_t,\theta)$. The motivation is that we can add a term $h(\mu_t)\cdot(y_t - \mu_t)$, since $\mathbb{E}_0[y_t - \mu_t \mid X_t] = 0$, and the choice $h(\mu_t) = \nabla_\mu V(\mu_t)$ ensures that (3.3) is satisfied. This nests both the constant variance and Poisson examples, since $\nabla_\mu V(\mu_t)$ equals 0 for the normal and -1 for the Poisson.

$$(4.9)\ H_1\text{: } \mathbb{E}_1[\{(y_t - \mu_t)^2 - V(\mu_t)\} + \nabla_\mu V(\mu_t)\cdot(y_t - \mu_t) \mid X_t] = g(X_t,\theta)\cdot\alpha \ ,$$

and it is understood that additionally $\mathbb{E}_1[y_t - \mu_t \mid X_t] = 0$.

The optimal RBCM test requires assumptions up to the fourth moments of y_t, though RBCM tests using assumptions on the first and second moments can also be constructed.

The regression-based tests are valid for any null hypothesis model for y_t in which the mean is the specified function $\mu(X_t,\theta)$ and the variance is the specified function $V(\mu(X_t,\theta))$. However, if considerably more structure is placed on y_t, by specifying both a null hypothesis and alternative hypothesis distributions satisfying (4.8) and (4.9) respectively, Cameron (1990) gives a number of examples, beyond the normal and Poisson, for which the LM test of $\alpha = 0$ is the optimal regression based test using the fundamental moment condition (4.9). However, the LM test is not always a function of the fundamental moment (4.9).

The RBCM test is easily implemented, since (3.3) is satisfied.

4.3 *Tests for Misspecified Third and Higher Central Moments*

The tests for the variance in the preceding section can be generalized to tests of higher order moments. We illustrate this for the third moment. Begin with the moment condition:

$$(4.10)\ H_0\text{: } \mathbb{E}_0[\{(y-\mu)^3 - v_3(\mu)\} + h_2(\mu)\cdot\{(y-\mu)^2 - v_2(\mu)\} + h_1(\mu)\cdot(y-\mu) \mid X] = 0 .$$

where we suppress the subscript t, and $v_j(\mu) \doteq \mathbb{E}_0[(y-\mu)^j \mid X]$. Then (3.3) is satisfied if and only if the $h_j(\mu)$ are chosen so that:

$$(4.11)\quad \{-3v_2(\mu) - \nabla_\mu v_3(\mu)\} + h_2(\mu)\cdot\{-\nabla_\mu v_2(\mu)\} - h_1(\mu) = 0.$$

Different null hypothesis models for y_t assume different moments $v_2(\mu)$ and $v_3(\mu)$. For example, consider a test for symmetry of the distribution of y_t when the variance of y_t is constant, i.e. $v_3(\mu) = 0$ and $v_2(\mu) = \sigma^2$. These are of course moment restrictions implied by the normal regression model. Then (4.11) implies $-3\sigma^2 - h_1(\mu) = 0$. So CM tests are most easily implemented if based on the fundamental moment condition:

$$(4.12) \qquad H_0\colon \mathbb{E}_0[\{(y-\mu)^3 + h_2(\mu)\cdot\{(y-\mu)^2 - \sigma^2\} - 3\sigma^2\cdot(y-\mu) \mid X] = 0 .$$

Note that a range of functions $h_2(\mu)$ may be used. One criteria is to choose choose $h_2(\mu)$ so that the test of third moments based on (4.12) is orthogonal to the second moment test based on (4.8). This orthogonality will require assumptions on up to the fifth moment of y_t. The theory of orthogonal polynomial systems, presented in section 5, uses this criteria.

The analysis can obviously be extended to higher central moments, such as tests for homokurtosis. When a density for y under the null hypothesis is specified, and is in the linear exponential family with quadratic variance funstion, analysis is simplified by using orthogonal polynomial systems. This is done in section 6.

The analysis can also be extended to factorial moments, rather than central moments. For example, the optimal RBCM test for misspecification of the third factorial moment is much simpler to derive than the equivalent LM test of Lee (1986) the testing the Poisson against a system of discrete distributions generated by the Pearson difference system.

4.4 *Tests for Serial Correlation*

For tests of serial correlation, the choice of fundamental moment condition is uncontentious, being:

$$(4.13) \qquad H_0\colon E_0[(y_t - \mu(X_t,\theta))\cdot((y_{t-j} - \mu(X_{t-j},\theta)) \mid X_t\,] = 0 \quad .$$

Under the alternative hypothesis:

$$(4.14) \qquad H_1\text{: } \mathbb{E}_1[(y_t - \mu(X_t,\theta))\cdot((y_{t-j} - \mu(X_{t-j},\theta)) \mid X_t] = g(X_t,\theta)\cdot\alpha ,$$

The simplest tests are for errors that are autocorrelated but not conditionally heteroskedastic. Then $\mathbb{E}_0[(y_t - \mu(X_t,\theta))^2 \mid X_t] = \sigma^2$, and $g(X_t,\theta) = \rho\sigma^2$. The optimal RBCM test, assuming constant fourth moment, regresses $(y_t - \mu(X_t,\theta)) \cdot(y_{t-j} - \mu(X_{t-j},\theta))$ on σ^2, which is equivalent to the usual LM test against AR(j) or MA(j).

For this time series example, X_t potentially includes lagged endogenous variables. If lagged endogenous variables are excluded, and the mean $\mu(X_t,\theta)$ is correctly specified, (4.16) satisfies (3.3), and the RBCM tests can be directly implemented. When lagged endogenous variables are present, (3.3) is no longer satisfied, and implementation is accordingly more complicated. This is essentially the observation of Durbin (1970).

5. Application to Specification Tests Based on Quasi-ML Estimation

When the conditional density under the null hypothesis is specified, standard specification tests are the lagrange multiplier (LM), Hausman (H) and information matrix (IM) tests. These tests can be interpreted as regression-based CM tests.

Many parametric single equation econometric models are based on one canonical parameter, typically the mean in some underlying model, which in turn is a function of several parameters. Sometimes, an additional parameter is needed, along with the canonical parameter, to determine the variance, but even then this may be a nuisance parameter.

In such parsimonious models, various test principles lead to specification tests based on a fundamental moment condition of low dimension, often scalar. Depending on the form of this fundamental moment condition, tests can be performed under assumptions considerably weaker than the assumption that the assumed conditional density is correctly specified.

Specifically, quasi-ML inference is based on an assumed density. For the case where y_t is i.i.d., there is a large menu of density functions of the form $f(y_t,\eta)$, where η is a hx1 vector. In regression analysis, the dependence of y_t on variables X_t is captured by letting $\eta = \eta(X_t,\theta)$, where θ is a qx1 vector of unknown parameters, so that the assumed density is of the form

$$(5.1) \qquad f(y_t,X_t,\theta) = f(y_t,\eta(X_t,\theta)).$$

Typically h is considerably less than q. For example, in the Poisson regression model η is a scalar, the mean, usually specified to equal $\exp(X_t'\theta)$.

Inference is based on the score vector:

$$(5.2) \qquad \nabla_\theta \log f(y_t,X_t,\theta) = \{\nabla_\eta \log f(y_t,\eta(X_t,\theta))\}' \cdot \nabla_\theta \eta(X_t,\theta) \quad .$$

5.1 *Lagrange Multiplier (LM) Tests and CM Tests*

In classical testing, the alternative hypothesis density is of the form (5.1), and the null hypothesis density is a specialization of (5.1) obtained by placing restrictions on θ. For ease of exposition, we suppose these restrictions are of the form $\theta_2 = 0$, where θ_2 is a q_2x1 subcomponent of θ. We have:

$$(5.3) \qquad H_0: f(y_t,X_t,\theta_1) = f(y_t,\eta(X_t,\theta_1,\theta_2=0)) \quad ,$$

$$(5.4) \quad H_1\!: f(y_t,X_t,\theta) = f(y_t,\eta(X_t,\theta)) , \quad \theta' = (\theta_1' \; ,\theta_2')' \; .$$

The LM, Wald and likelihood ratio tests are different tests of these restrictions that are asymptotically equivalent and locally most powerful if the true d.g.p. is the null hypothesis density. Specification tests focus on the LM test, since its implementation only requires estimation of parameters of the null hypothesis density.

The LM test, or score test, is a test of whether the score vector for the alternative hypothesis density, evaluated under the null, is zero, i.e. it is a test of the moment condition:

$$(5.5) \quad E_0[\, \{\nabla_\eta \log f(y_t,\eta(X_t,\theta_1,0))\}' \cdot \nabla_{\theta_1} \eta(X_t,\theta_1,0) \mid X_t] = 0 \; .$$

using (5.2), where $\nabla_\eta \log f(y_t,\eta(X_t,\theta_1,0)) = \nabla_\eta \log f(y_t,\eta(X_t,\theta))\Big|_{\theta_2=0}$ and $\nabla_{\theta_1} \eta(X_t,\theta_1,0) = \nabla_{\theta_1} \eta(X_t,\theta)\Big|_{\theta_2=0}$.

Therefore an LM test is a CM test of the fundamental moment

$$(5.6) \quad H_0\!: \; E_0[\nabla_\eta \log f(y_t,\eta(X_t,\theta_1,0)) \mid X_t] = 0 \quad ,$$

and is the optimal RBCM test against the alternative

$$(5.7) \quad H_1\!: E_1[\nabla_\eta \log f(y_t,\eta(X_t,\theta_1,0)) \mid X_t]' = \Sigma(X_t,\theta)^{-1} \cdot \nabla_{\theta_1} \eta(X_t,\theta_1,0) \cdot \alpha$$

where

$$(5.8) \quad \Sigma(X_t,\theta) = E_0[\nabla_\eta \log f(y_t,\eta(X_t,\theta_1,0)) \cdot \nabla_\eta \log f(y_t,\eta(X_t,\theta_1,0))' \mid X_t].$$

This decomposition makes clear that the LM test relies on distributional assumptions on the first two moments of $\nabla_\eta \log f(y_t,\eta(X_t,\theta_1,0))'$, a $h\times1$ vector, rather than on the entire conditional density $f(y_t,\eta(X_t,\theta_1,0))$. This explains why many commonly-used LM tests are appropriate under assumptions weaker than those used to derive the test.

If tests under even weaker null hypothesis assumptions are desired, a CM test assuming only correct specification of the first moment can easily be constructed, as in section 2.2.

A wide range of RBCM tests based on (5.6) and alternatives of the form (2.4) can be developed. In particular, many null hypothesis models depend on only one parameter, say η_1, or there are additional parameters but these can be treated as nuisance parameters. In tests of omitted variables, $\eta = \eta_1$ and we wish to test whether θ_2 in $\eta(X_t,\theta_1,\theta_2)$ equals zero. In this case the fundamental moment is simply $E_0[\nabla_{\eta_2} \log f(y_t,\eta(X_t,\theta_1) \mid X_t]$. In tests of a richer model $\eta = (\eta_1, \eta_2)'$ is a 2×1 vector, and η_1 depends on θ_1 alone and η_2 depends on θ_2 alone. Then the fundamental moment is $E_0[\nabla_{\eta_2} \log f(y_t,\eta(X_t,\theta_1) \mid X_t]$

Hausman tests can be considered as tests based on the score vector, for example see Ruud (1984) and White (1990, section 9.4.d). Given such a representation, Hausman tests can also be represented as RBCM tests along lines similar to that of the preceding section.

5.2 *Information Matrix (IM) Tests and CM Tests*

We consider tests of models with null hypothesis densities of the form (5.1), and hence score vector (5.2). The IM test is based on the information matrix equality, that the sum of the outer product of the score vector and the derivative of the score vector is zero under the null hypothesis. Define

$$(5.9)\quad R_{IM}(y_t,X_t,\theta) = \nabla^2_\eta \log f(y_t,\eta(X_t,\theta)) + \nabla_\eta \log f(y_t,\eta(X_t,\theta))\cdot\nabla_\eta \log f(y_t,\eta(X_t,\theta))' \quad ,$$

where $\nabla^2_\eta \equiv \partial^2/\partial\eta^2$. Then the IM test is based on the moment condition:

$$(5.10)\quad E_0[\nabla_\theta\eta(X_t,\theta)'\cdot R_{IM}(y_t,X_t,\theta)\cdot\nabla_\theta\eta(X_t,\theta) + \sum_{i=1}^{h} \nabla_{\eta_i} \log f(y_t,\eta(X_t,\theta))'\cdot\nabla^2_\theta\eta_i(X_t,\theta) \mid X_t] = 0 \quad ,$$

where η_i is the i-th component of η. Since $E_0[\nabla_\eta \log f(y_t) \mid X_t] = 0$, the term on the second line is dropped. Vectorizing, the IM test is equivalent to a CM test of:

$$(5.11)\quad E_0[(\nabla_\theta\eta(X_t,\theta)'\otimes\nabla_\theta\eta(X_t,\theta)')\cdot \text{vec}(R_{IM}(y_t,X_t,\theta)) \mid X_t] = 0 \quad .$$

For notational simplicity we ignore the obvious point that because of the symmetry of second derivatives, we would only select the subset of unique elements of in $\nabla_\theta\eta(X_t,\theta)'\otimes\nabla_\theta\eta(X_t,\theta)'$ and $\text{vec}(R_{IM}(y_t,X_t,\theta))$. (5.11) is the optimal RBCM test of

$$(5.12)\quad E_0[\text{vec}(R_{IM}(y_t,X_t,\theta)) \mid X_t] = 0 \quad ,$$

against

$$(5.13)\quad E_1[\text{vec}(R_{IM}(y_t,X_t,\theta)) \mid X_t] = \Sigma(X_t,\theta)^{-1}\cdot(\nabla_\theta\eta(X_t,\theta)'\otimes\nabla_\theta\eta(X_t,\theta))\cdot\alpha \quad ,$$

where

$$(5.14)\quad \Sigma(X_t,\theta) = E_0[\text{vec}(R_{IM}(y_t,X_t,\theta))\cdot\text{vec}(R_{IM}(y_t,X_t,\theta))' \mid X_t] \; .$$

Therefore the IM test can be interpreted as a test of a null hyothesis moment condition against one for the alternative hypothesis. This overcomes one of the perceived weaknesses of the IM test. The above results also explain why IM tests can have poor power. The IM test as usually used does not vary with possible alternative hypotheses. It is clear from (5.13) that the IM test can be generalized to test (5.12) against alternatives $G(X_t,\theta)\cdot\alpha$ other than that given in (5.13), thereby increasing the power of the IM test in certain directions.

5.3 *CM Tests based on orthogonal polynomials*

Another possible choice of the fundamental moment is that based on orthogonal polynomials. A very brief presentation of orthogonal polynomials is given; for more details see, for example, Szegö (1975).

Let f(y) be a function of the scalar variable y, $a \leq y \leq b$. The function f(y) is taken to be nonnegative and integrable on an interval (a,b) and taken to be positive on a sufficiently large subset (a,b). It is further assumed that finite moments of the variable y,

$$\mu_n = \mathbb{E}[y^n] = \int_a^b y\cdot f(y)\ dy\ , \qquad n = 0,1,2\ldots$$

exist up to the required order.

A system of orthogonal polynomials $P_n(y)$, degree $[P_n(y)] = n$, is called orthogonal with respect to f(y) on the interval $a \leq y \leq b$ if

$$\int_a^b [P_n(y)\cdot P_m(y)\cdot f(y)]\ dy = 0\ , \quad m \neq n;\ n,m = 0,1,2\ldots.$$

By the fundamental theorem of orthogonal polynomials, a real orthogonal polynomial system (OPS) corresponding to the density function exists and is uniquely determined (Szegö, (1975), chapter II).

For the OPS, $P_n(y)$ is a polynomial of degree n in y, with

$$(5.11) \qquad \mathbb{E}[P_n(y)P_m(y)] = \delta_{mn}k_n, \quad k_n \neq 0$$

where δ_{mn} is the Kronecker delta. In the special case of an orthonormal polynomial sequence, $k_n=1$. The orthonormal polynomial sequence may be determined, using (5.11) above, in a step-by-step manner by the (Gram-Schmidt) orthogonalization process.

As an example, let $a=-\infty$ and $b=+\infty$, $f(y) = \exp(-(y-\theta)^2)$; then $P_n(y-\theta)$ is, except for a constant factor, the Hermite polynomial. As another example, let $a=0$ and $b=+\infty$, $f(y,\theta) = \exp(-(y-\theta))(y-\theta)^{\alpha}$, $\alpha > -1$; then $P_n(y-\theta)$ is, except for a constant factor, the Laguerre polynomial. See Szegö (1975, p.22).

Orthogonal polynomials are discussed in detail in Cameron and Trivedi (1990b). In applying them for testing, we replace the function f(y) by the density function $f(y_t,X_t,\theta \mid X_t)$. The orthogonal polynomials accordingly become $P_n(y_t,X_t,\theta \mid X_t)$.

They are especially useful for the choice of fundamental moments for several reasons. First, they are easily generated via a three-term recurrence relationship. This can be especially advantageous compared to LM tests which require finding, under the null hypothesis, the limiting value of the score vector for alternative hypothesis density. Second, because they are polynomials in y, distributional assumptions may conceivably be minimized. Thirdly, because of the orthogonality result (ii) given above, CM tests based on polynomials of progressively higher degree will be statistically independent. Finally, in a leading example, given in the next section, tests based on orthogonal polynomials coincide with commonly-used LM and IM tests.

6. Application to Specification Tests in the LEF-QVF

To illustrate the use of orthogonal polynomials as the basis for the choice of fundamental moment, we consider linear exponential families with quadratic variance functions, henceforth abbreviated to LEF-QVF. This covers many commonly used econometric models: regression models with constant variance; discrete choice models such as probit and tobit; poisson models for count data; and gamma models for continuous positive data. In this leading case, the fundamental moments from various testing approaches are closely related, and are the first few terms in an orthogonal polynomial system.

We use the results of Morris (1982). The LEF is defined by

$$f(y,\psi) = \exp\{y\psi - \varphi(\psi) + k(y)\}\ , \tag{6.1}$$

where ψ is a scalar parameter, and the dependence of ψ on exogenous regressors has been suppressed for notational convenience. The LEF has the property

$$\mathbb{E}[y] = \mu \equiv \mu(X,\psi) = \nabla_{\psi}\varphi(\psi) \tag{6.2}$$

$$\mathrm{var}[y] = \nabla^2_{\psi}\varphi(\psi) \tag{6.3}$$

where $\nabla^n_{\psi} \equiv \partial^n/\partial\psi^n$.

In a more general exponential family, $f(y,\psi) = \exp\{g(y,\psi) - \varphi(\psi) + k(y)\}$. The LEF is the specialization where the function $g(y,\psi)$ is linear in y, in which case y is called the natural observation, and linear in ψ, in which case ψ is called the natural parameter. Other studies, such as Gourieroux, Montfort, and Trognon (1984), use the mean parameterization of the LEF: $f(y,\mu)$

$= \exp\{A(\mu) + B(y) + C(\mu)y\}$, where the functions A, B and C are such that the density integrates to 1 and conditions corresponding to (6.2) and (6.3) are satisfied. Here the natural parameterization of the LEF is used, which Morris (1982) called the natural exponential family (NEF). These are just two different parameterizations, using the mean μ or the natural parameter ψ, of the same family of densities.

Morris (1982) restricts analysis to quadratic variance functions, meaning the variance is a quadratic function of the mean so that $V(\mu)$ satisfies the relationship

(6.4) $$V(\mu) = v_0 + v_1\mu + v_2\mu^2 \quad ,$$

where various possible choices of the coefficients v_0, v_1 and v_2 lead to six exponential families, five of which are the normal, Poisson, binomial, gamma, and negative binomial families. So the restriction to QVF still leaves a wide range of commonly used models.

The following results (Morris (1982)) are useful in deriving the fundamental moment restrictions:

(i) For the LEF-QVF the orthogonal polynomial system $P_m(y,\mu)$ is defined by

(6.5) $$P_m(y,\mu) = V^m\{\nabla_\mu^m f(y,\psi)/f(y,\psi)\}, \qquad m=0,1,2,\ldots$$

where $P_m(y,\mu)$ is a polynomial of degree m in both y and μ with leading term y^m , m=1,2,..., and $f(y,\psi)$ is the LEF-QVF density.

(ii) The polynomials $\{P_m(y,\mu)\}$ satisfy the recurrence relationship

$$(6.6)\qquad P_0 = 1$$

$$P_1 = y-\mu$$

$$P_{m+1} = (P_1 - m\nabla_\mu V(\mu))P_m - m(1 + (m-1)v_2)V(\mu)P_{m-1}\ , \qquad \text{for } m\geq 1\ .$$

(iii) Let $a_0=1$, and for $m\geq 1$,

$$(6.7)\qquad a_m = m!\prod_{i=0}^{m-1}(1 + iv_2).$$

Then

$$(6.8)\qquad \mathbb{E}_0 P_m = 0\ , \qquad m \geq 1\ ,$$

$$(6.9)\qquad \mathbb{E}_0 P_m P_n = \delta_{mn} a_m V^m, \qquad m,n \geq 0\ ,$$

$$(6.10)\qquad \nabla_\mu^r P_m = (-1)^r (a_m/a_{m-r}) P_m, \qquad m \geq 1\ ,\ r = 1,\ldots,m\ .$$

We shall use (6.8) as the fundamental moment condition for CM tests. In regression applications, ψ or equivalently μ is parameterized in terms of X_t and θ, i.e. μ above is replaced by $\mu_t = \mu(X_t,\theta)$. The procedure is to progressively test for $m = 1,2,..$

$$(6.11)\qquad H_0:\ \mathbb{E}_0[P_m(y_t,X_t,\theta)\ |\ X_t] = 0\ ,$$

against:

$$(6.12)\qquad H_1:\ \mathbb{E}_1[P_m(y_t,X_t,\theta)\ |\ X_t] = G_m(X_t,\theta)\cdot\alpha_m\ ,$$

where the recurrence relation (6.6) generates $P_m(y_t,X_t,\theta) = P_m(y_t,\mu(X_t,\theta)) = P_{mt}$ in the shorthand notation above.

By property (i) above, P_{mt} is a polynomial of degree m in $(y_t - \mu_t)$ and μ_t, so that the distributional assumptions used in performing the optimal RBCM test are that the first 2m central moments of y are correctly specified, and non-optimal RBCM tests can be validly used assuming that the first m central moments are correctly specified. The variance of P_{mt} for the optimal test is easily obtained using (6.7) and (6.9). Tests based on different degrees of polynomial are orthogonal by (6.9).

The tests are easily implemented, since by (6.10) condition (3.3) holds for $m > 1$, provided $E_0[P_{m-1,t}] = 0$.

In comparing tests based on these orthogonal polynomials with other tests, such as LM and IM tests, we say the tests are identical if they have the same fundamental moment condition. There is, of course, the separate issue of the choice of $G(X_t,\theta)$.

Tests based on $P_{1t} = (y_t - \mu_t)$ coincide with the tests of conditional mean given in section 4.1. The optimal RBCM test uses $\Sigma_t = V(\mu_t)$. This test is easily shown to coincide with the LM test for omitted variables in an LEF-QVF model. It also coincides with the LM test for misspecified functional form of the conditional mean, where the alternative hypothesis model is embedded in an LEF-QVF, proposed by Gurmu and Trivedi (1990).

Tests based on $P_{2t} = (y_t - \mu_t)^2 - V(\mu_t) - \nabla_\mu V(\mu_t)(y_t - \mu_t)$ coincide with the tests of conditional variance given in section 4.2. The optimal RBCM test uses $\Sigma_t = 2(1+v_2)(1+2v_2)v(\mu_t)$. Cameron (1990) gives a number of examples where this corresponds to LM tests where a specific family in the LEF-QVF is embedded in various alternative densities. Cameron shows that a variant of the IM test uses P_{2t} as the fundamental moment. Wooldridge (1990b) considers the more general multivariate LEF with nuisance parameter, and uses a different variant of the IM test that does not satisfy (3.3).

For tests based on higher order polynomials, it is easiest to consider in turn each of the LEF-QVF families in turn. This is done in Cameron and Trivedi (1990b). We consider two examples here.

For the the normal family with mean μ and variance σ^2, $V(\mu) = \sigma^2$ implies that the variance does not depend upon the mean so that $\nabla_\mu V(\mu) = 0$, $v_1 = v_2 = 0$ and the recurrence relationship for the orthogonal polynomials is $P_{m+1} = P_1 P_m - mVP_{m-1}$. The orthogonal polynomials of order two, three and four are respectively $\{(y-\mu)^2 - \sigma^2\}$, $\{(y-\mu)^3 - 3\sigma^2(y-\mu)\}$, and $\{(y-\mu)^4 - 6\sigma^2(y-\mu)^2 + 3\sigma^4\}$, and their variances are respectively $2\sigma^4$, $6\sigma^6$, and $24\sigma^8$. These lead directly to tests of heteroskedasticity, skewness and non-normal kurtosis identical to the LM tests of Bera and Jarque (1982).

For the Poisson family with mean μ, $V(\mu) = \mu$, $\nabla_\mu V(\mu) = 1$, $v_1 = 1$, $v_2 = 0$. The recurrence relation for the orthogonal polynomials is $P_{m+1} = (P_1 - m)P_m - mVP_{m-1}$. The orthogonal polynomials of order two and three are $(y-\mu)^2 - y$ and $(y-\mu)^3 - 2(y-\mu)^2 - y(y-\mu) + 2(y - \mu y + \mu^2)$, with variances $2\mu^2$ and $6\mu^3$. The test of overdispersion has already been discussed. A test of the third moments based on P_3 coincides with the LM test of Lee (1986, equation (5.12)). The derivation here is much simpler.

While these tests based on orthogonal polynomials assume a density under the null hypothesis, and are compared to LM and IM tests, we reiterate that their validity rests on correct specification of lower order moments rather than the entire density. This should be especially clear from section 5 where similar tests were obtained without specification of the density.

7. Conclusion

Parametric estimation theory has developed beyond that for maximum likelihood (ML) estimation assuming a correctly specified likelihood function to more general theories of quasi-ML estimation, White (1982), and estimation based on moment conditions, Hansen (1982). For testing of parametric models, the conditional moment testing approach of Newey (1985) and Tauchen (1985) permits a similar movement away from likelihood theory. The regression-based CM tests do just this.

Theoretically, RBCM tests provide a useful framework for contrasting various specification tests. Different CM tests correspond to different fundamental moments and parameterizations of the fundamental moment under the alternative. Various choices of the fundamental moment are discussed.

Practically, new tests for heteroscedasticity, skewness and kurtosis can be easily derived. These tests can be implemented by a convenient regression that is not auxiliary. Though not likelihood-based, these new tests can be viewed as generalizations of LM tests or variable addition tests based on the quasi-MLE of the regression model under normality to tests based on the quasi-MLE of regression models with density in the linear exponential family. Thus familiar tests for the classical regression model can be extended to a wide range of models that includes discrete choice models (probit, logit, ..), count data models (Poisson, exponential), and positive data models (gamma).

ACKNOWLEDGEMENTS: This work has benefitted from conversation with Hal White, and presentation of a related paper at University of California, San Diego.

References

Amemiya, T., 1985, *Advanced Econometrics,* Harvard University Press.

Bera, A.K. and C.M. Jarque, 1982, "Model Specification Tests: A Simultaneous Approach", *Journal of Econometrics,* 20, 59-82.

Cameron, A.C., 1990, "Regression-Based Tests of Variance-Mean Relationships", forthcoming working paper, Department of Economics, U.C.-Davis.

Cameron, A.C. and P.K. Trivedi, 1990a, "Regression-Based Tests For Overdispersion In the Poisson Model", *Journal of Econometrics,* Forthcoming.

Cameron, A.C. and P.K. Trivedi, 1990b, "Conditional Moment Tests and Orthogonal Polynomials", forthcoming working paper, Department of Economics, Indiana University - Bloomington.

Cox, D.R. and E.J. Snell, 1968, "A General Definition of Residuals", *Journal of the Royal Statistical Society*, B, 30, 248-275.

Cox, D.R., 1983, "Some Remarks on Overdispersion", *Biometrika*, 70, 269-274.

Davidson, R. and J.G. MacKinnon, 1984, "Convenient Specification Tests For Logit and Probit Models", *Journal of Econometrics*, 25, 241-262.

Davidson, R. and J. G. MacKinnon, 1990, "Specification Tests Based on Artificial Regressions", *Journal of American Statistical Association,* 85, 220-227.

Durbin, J., 1970, "Testing for Serial Correlation in Least-Squares Regression when some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410-421.

Godfrey, L.G., 1988, *Misspecification Tests In Econometrics: The Lagrange Multiplier Principle and Other Approaches*, Cambridge: Cambridge University Press.

Gourieroux, C., A. Monfort and A.Trognon, 1984, Pseudo Maximum Likelihood Methods:Theory", *Econometrica*, 52, 3, 681-700.

Gurmu, S. and P.K. Trivedi, 1990, "The Variable Augmentation Specification Tests in the Exponential Family", Indiana University Working Papers in Economics, 90-008.

Hall, A., 1987, "The Information Matrix Test in the Linear Model", *Review of Economic Studies,* 54, 257-265.

Hansen, L.P., 1982, "Large Sample Properties of Generalized Methods of Moment Estimators", *Econometrica*, 50, 1029-1054.

Hausman, J.A., 1978, "Specification Tests in Econometrics", *Econometrica,* 46, 6, 1251-1271.

Lee, L.-F, 1986, "Specification Test For Poisson Regression Models", *International Economic Review*, 27, 3, 689-706.

McCullagh, P. and J. A. Nelder, 1983, *Generalized Linear Models, London*: Chapman and Hall.

Morris, C.N., 1982, "Natural Exponential Families With Quadratic Variance Functioms", *Annals of Statistics,* 10, 65–80.

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

Pagan, A.R., 1984, "Model Evaluation by Variable Addition", in D.F.Hendry and K.F.Wallis (eds), *Econometrics and Quantitative Economics*, 103–133, Blackwell, Oxford.

Pagan, A.R., 1986, "Two Stage and Related Estimators and Their Applications", *The Review of Economic Studies,* 53, 517–538.

Pagan, A.R. and A.D.Hall, 1983, "Diagnostic Tests as Residual Analysis", *Econometric Reviews,* 2, 159–218.

Pagan, A.R. and F. Vella, 1989, "Diagnostic Tests for Models Based on Individual Data", *Journal of Applied Econometrics,* Supplement, S4, S29–S60

Pierce, D.A., 1982, The Asymptotic Effect of Substituting Estimators for Parameters in Certain Types of Statistics", *Annals of Statistics,* 10, 475–478.

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

Szego, G., 1975, *Orthogonal Polynomials,* 4th ed., Vol 23, American Mathematical Society Colloquium Publication, Providence, RI.

Tauchen, G. 1985, Diagnostic Testing and Evaluation of Maximum Likelihood Models, *Journal of Econometrics*, 30, 415–443.

White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1–25.

White, H., 1987, "Specification Testing in Dynamic Models", in T. Bewley, ed., *Advances in Econometrics, Fifth World Congress*, 1, 1–58, Cambridge University Press, New York.

White, H., 1990, *Estimation, Inference and Specification Analysis,* Cambridge University Press, forthcoming.

Wooldridge, J.M., 1990a, "A Unified Approach to Robust, Regression–Based Specification Tests", *Econometric Theory*, 6, 17–43.

Wooldridge, J.M., 1990b, "Specification testing and Quasi–likelihood Estimation", *Journal of Econometrics*, forthcoming.