CONDITIONAL MOMENT TESTS AND ORTHOGONAL POLYNOMIALS

By

A. Colin Cameron
Department of Economics
University of California
Davis, CA, 95616, U.S.A.

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

This version: June 1993

ABSTRACT

The conditional moment (CM) tests of Newey (1985) and Tauchen (1985) are based on the asymptotic distribution of a function with zero mean. The construction of a suitable moment function is the first step in this procedure. This paper presents a unified theory for deriving the moment functions in the parametric case using known results from the theory of series expansions of distributions in terms of a baseline distribution and related orthogonal polynomials. The approach is used to construct CM tests in a number of cases, including the leading case of linear exponential families with quadratic variance functions. This includes Poisson and negative binomial models for count data, exponential for duration data and binomial for discrete data, in addition to the classical regression model under normality. Modifications of the approach when the data are truncated and connections with the score test are also considered.

Some Key Words: *CONDITIONAL MOMENT SPECIFICATION TESTS; SERIES EXPANSIONS; ORTHOGONAL POLYNOMIALS; LEF-QVF PARAMETERIZATION; GENERALIZED LINEAR MODELS; SCORE TESTS; INFORMATION MATRIX TESTS.*

Acknowledgments: We are grateful to Anil Bera, Shiferaw Gurmu, George Tauchen, and two anonymous referees for comments and discussions of previous drafts which resulted numerous improvements. However, the authors remain responsibile for the current version.

Address for correspondence: Pravin K. Trivedi, Department of Economics, Ballantine Hall, Indiana University, Bloomington, IN 47405.

1. INTRODUCTION

Consider a set-up with data $\{(y_t, X_t), t = 1,\dots,T\}$ independent across t, where the dependent variable is y_t, and explanatory variables are the vector X_t. The true data generating process (d.g.p.) for y given X is unknown, but we have a hypothesized parametric density function, denoted $f(y,X,\theta)$, $\theta \in \mathbb{R}^q$. Conditional moment tests are tests of the validity of moment conditions implied by these assumed parameterizations. In this paper we propose an approach to the construction of moment functions based on orthogonal polynomials.

By definition, a conditional moment test is any test based on an sx1 vector of functions $m(y,X,\theta)$ that satisfy the moment condition:

$$(1.1) \qquad \mathbb{E}_0[m(y_t,X_t,\theta) \mid X_t] = 0 \quad ,$$

where the subscript 0 denotes expectation with respect to the assumed distribution.

Tests based on a moment condition of the form (1.1), henceforth called CM tests, were introduced by Newey (1985) and Tauchen (1985), who also developed the associated asymptotic theory. Further results by Pagan and Vella (1989), White (1987, 1990) and Wooldridge (1990) demonstrate the unifying and simplifying power of CM tests as tests of specification. Since most specification tests can be interpreted as CM tests, there is a strong case for adopting it as the preferred general approach to specification tests.

The simplest version of a CM test based on (1.1) uses the corresponding sample moment:

$$(1.2) \qquad m_T(\theta) = T^{-1} \sum_{t=1}^{T} m(y_t,X_t,\theta) \quad .$$

To operationalize a CM test, the parameter θ in (1.2) 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.

To date most authors assume at the start that a suitable moment function for constructing the test is available. However, since such moment functions are not unique, it is desirable to avoid arbitrariness in this choice. Specifically, the chosen moment functions should satisfy some optimality criterion, and the relation between different moment functions should be clarified.

In this paper we propose an approach to the construction of CM functions based on orthogonal polynomials. The literature on orthogonal polynomials is vast, their basic properties are well known and widely used, and many excellent treatises on this subject are available. Some applications to testing of nonlinear regression models exist (Kiefer (1985), Lee (1986), Smith (1989), Cameron and Trivedi (1993)). These construct score tests against an alternative hypothesis density function that is a series expansion in terms of the orthogonal polynomials of the null hypothesis density. These examples are quite specific, and generally use the approach as a way to specify an alternative hypothesis density rather than fully utilizing the properties of orthogonal polynomials. The approach retains considerable unexploited potential as a *general* approach to specification testing within the CM framework.

A related testing procedure is to use k-th order moment functions such as $[(y - \mu)^k - \mathbb{E}[(y - \mu)^k \mid X,\theta]$ where $\mu = \mathbb{E}[y \mid X,\theta]$, as in, for example, Pagan and Vella (1989), or $y^k - \mathbb{E}[y^k \mid X,\theta]$, as in Smith (1989). These tests in general differ from those obtained by the orthogonal polynomial approach. Which approach leads to more powerful tests clearly depends on the alternative hypothesis, as any CM test can be interpreted as a score (and hence locally most powerful) test against some alternative (White (1990)). We give some conditions under which the orthogonal polynomial tests are more powerful. And in those examples that we are aware of in which score tests are functions of polynomials, they are functions of orthogonal polynomials.

We introduce orthogonal polynomials in section 2. Selected expressions for orthogonal polynomials are given in this section, while important general results used in this and later sections are given in Appendix A. In section 3 we propose a general procedure for specification tests of distributional misspecification based on orthogonal polynomials and suitable functions of exogenous variables, and analyze local asymptotic power of these tests. In section 4 the discussion is narrowed to the leading case of the linear exponential family with quadratic variance function (LEF-QVF). This includes Poisson and negative binomial models for count data, exponential for duration data and binomial for discrete data, in addition to the classical regression model under normality. Illustratively, several specification tests, some well known and some new, are very simply derived using orthogonal polynomials. Many other applications are possible, and section 5 considers a model with truncation. Section 6 concludes.

2. ORTHOGONAL POLYNOMIALS: SELECTED PROPERTIES AND RESULTS

Let F(y) denote the distribution function and let dF(y) = f(y)dy where f(y) is the density of the independently distributed scalar continuous random variable y. The density function f(y) is taken to be nonnegative and integrable on an interval [a,b] and F(y) has points of increase on a sufficiently large subset [a,b]. All arguments given below can be repeated after appropriate change of notation for the case of a discrete random variable and corresponding results for the discrete case may be reproduced.

It is assumed that finite moments of all order, denoted by μ_n, exist;

(2.1) $$\mu_n = \mathbb{E}[y^n] = \int y^n \cdot f(y)\, dy\ , \qquad n=0,1,2...$$

In general f(y) may be a marginal or a conditional density, but for the purposes of this paper f(y) will be a conditional density, usually denoted by $f(y,X,\theta \mid X)$ where θ is an unknown parameter and X is data. We use f(y) forgenerality and more compact notation. While expectations in (2.1) and elsewhere in section 2 are taken w.r.t. the assumed density f(y), this may not be the true d.g.p.

Definition: A system of orthogonal polynomials, henceforth abbreviated to OPS, $P_n(y)$ (or $P_n(y,X,\theta \mid X)$), degree $[P_n(y)] = n$, is called orthogonal with respect to f(y) (or $f(y,X,\theta \mid X)$) on the interval $a \le y \le b$ if

(2.2) $$\int P_n(y)\cdot P_m(y)\cdot f(y)\, dy = \begin{cases} k_n & \text{if } m=n \\ 0 & \text{if } m\neq n \end{cases}$$

That is, $P_n(y)$ is a polynomial of degree n, a positive integer, in y satisfying the orthogonality condition

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

where δ_{mn} is the Kronecker delta, $\delta_{mn} = 0$ if $m \neq n$, $\delta_{mn} = 1$ if $m = n$. In the special case of an *orthonormal polynomial* sequence, $k_n = 1$.

Orthogonal polynomials have several properties we exploit in the construction of tests of moment restrictions such as (2.3). These include uniqueness, linear independence, and minimum variance; they are summarized in Appendix A.

The basic idea of the paper is that conditional moment restrictions implied by models derived from parametric families of distributions can be expressed and tested using the corresponding sequences of orthogonal polynomials, as an alternative to raw (non-orthogonal) moment restrictions. A comparison between the two alternatives is considered in Section 3. Tests will in practice be based on low order orthogonal or orthonormal polynomials, rarely exceeding three or four. General conditional moment expressions required for such tests can be derived using the methods of Appendix A. Important special cases given in Table 1 are used in section 4.

For most of the discussion in this paper we concentrate on tests of first and second moments. We consider two types of tests, those based on orthogonal polynomials and those based on their orthonormal counterparts. It is convenient to provide general expressions for these. They can be specialized to apply to specific distributions by substituting expressions for the relevant moments. Let $\mu_1 = \mathbb{E}(y|X)$, $\mu_2' \equiv \sigma^2 = \mathbb{E}(y - \mu_1)^2$; $\mu_3' = \mathbb{E}(y - \mu_1)^3$; $\mu_4' = \mathbb{E}(y - \mu_1)^4$; further, let $\gamma_1 = \mathbb{E}[(y - \mu_1)/\sigma]^3$ and $\gamma_2 = \mathbb{E}[(y - \mu_1)/\sigma]^4 - 3$ define the standardized skewness and excess kurtosis parameters. Then the first two orthogonal polynomials ($P_0 = 1$), expressed as deviations from the mean, are

(2.4) $$P_1(y_i) \equiv P_1(y_i - \mu_{1i}) = y_i - \mu_{1i}$$

(2.5) $$P_2(y_i) \equiv P_2(y_i - \mu_{1i}) = (y_i - \mu_{1i})^2 - (\mu_{3i}'/\mu_{2i}')(y_i - \mu_{1i}) - \mu_{2i}'.$$

To derive orthonormalized versions of these, each polynomial is standardized by the respective variance expression, derived using the methods of Appendix A. The resulting polynomials, which we denote by the symbol $Q_i(y)$ to avoid confusion, have zero mean and unit variance property; they also incorporate more information about the moment properties of the hypothesized distribution than their orthogonal counterparts.

(2.6) $$Q_1(y_i) \equiv P_{1i}/\sqrt{\text{var } P_{1i}} = (y_i - \mu_{1i})/\sigma_i$$

(2.7) $$Q_2(y_i) \equiv P_{2i}/\sqrt{\text{var } P_{2i}} = \frac{(y_i - \mu_{1i})^2 - \gamma_{1i}(y_i - \mu_{1i})\cdot\sigma_i - \sigma_i^2}{\sigma_i^4(\gamma_{2i} + 2 - \gamma_{1i})}$$

Such re-expression of orthogonal polynomials in terms of residuals, $y - \mu_1$, rather than in y alone, is natural in the regression context.

3. TESTS BASED ON ORTHOGONAL POLYNOMIALS

3.1 Conditional moment test based on orthogonal polynomials

If the assumed distribution implies testable moment restrictions, the tests can be carried out, singly or jointly, using orthogonal polynomials of the appropriate order. Since $\mathbb{E}_0[P_n(y,X,\theta)\,|X] = 0$, use of the law of iterated expectations, following Newey (1985, p.1055), suggests CM tests based on moment functions of the form

(3.1) $$\mathbb{E}_0[m_n(y,X,\theta)\,|X] = 0,$$

where

(3.2) $$m_n(y,X,\theta) = G_n(X,\theta)\cdot P_n(y,X,\theta),$$

and $G_n(X,\theta)$ is a function of X and θ, and different subsets of X may appear in the functions G_n and P_n. For a single moment restriction $G_n(\cdot)$ is a scalar function, for a vector of moment conditions it is a matrix. For example, a test of omitted variables, denoted by X_2, from the conditional mean function may be based on the orthogonality condition (3.3) and (3.4) as appropriate:

(3.3) $$\mathbb{E}_0[m_1(y,X,\theta)\,|X] = \mathbb{E}_0[X_2\cdot P_1(y,X_1,\theta)\,|X] = 0,$$

(3.4) $$= \mathbb{E}_0[X_2 \cdot (y - \mu(X_1,\theta)) \mid X] = 0 .$$

Similarly a test of misspecified variance function may be based upon

(3.5) $$\mathbb{E}_0[m_2(y,X,\theta \mid X)] \equiv \mathbb{E}_0[G_2(X,\theta)\cdot((y - \mu_1)^2 - (\mu_3'/\mu_2')(y - \mu_1) - \mu_2')] = 0,$$

where μ_1, μ_2' and μ_3' are functions of X, θ. The same general approach can be used to derive higher moment restrictions.

The approach based on orthogonal polynomials has considerable algebraic simplicity. The derivation of CM tests for distributions with finite moments (expressible in closed form) of requisite order requires no more than substitution into appropriate formulae. When dealing with data for which the first few moments are the same as those of a known distribution, the approach suggests suitable moment functions for testing.

To test an nth order moment restriction we may use the nth order orthogonal polynomial, and under the null hypothesis density the resulting test statistic will be asymptotically independently distributed of all other tests based on higher or lower order polynomials, in the absence of unknown nuisance parameters. Linear independence of moment functions is an advantage in testing when tests are likely to confound different moment misspecifications. Correlation between (say) first and second moment tests can distort the size of an individual misspecification test. When orthogonal polynomials are used, ‘portmanteau’ or simultaneous tests of several restrictions may be easily implemented when the joint test is additive in its components, as it will be in many cases. The properties of uniqueness and minimum variance (in the class of monic polynomial functions) has implications for the asymptotic local power of tests based on orthogonal polynomials, as is shown later in Section 3.4.

Polynomial tests (orthogonal or nonorthogonal) of moments of order n are based on moment assumptions up to order n. The derivation of optimal versions of such tests will involve moment assumptions to order 2n. These optimal versions can be made robust by methods similar to Koenker (1981) so as to depend on moment assumptions up to order n, though for high order n tests may be numerically unstable unless the sample is very large.

3.2 Score tests based on orthogonal polynomial expansions for densities

Let f(y) be a continuous density function and let $\{P_0(y),\ldots\}$ be the corresponding set of orthonormal polynomials; Let g(y) be another density assumed to be ϕ^2-bounded in the sense that $\phi^2 + 1 = \int_{-\infty}^{\infty}\{g(y)/f(y)\}^2 f(y)dy < \infty$, then the following series expansion is formally valid (Ord (1972)):

$$g(y) = f(y)\cdot\left[a_0P_0(y) + a_1P_1(y) + \ldots\right]. \tag{3.6}$$

Multiplying (3.6) by $P_n(y)$ and integrating term by term, and noting $P_0(y) = 1$,

$$a_n = \int P_n(y)g(y)dy, \quad a_0 = 1 \tag{3.7}$$

$$\phi^2 = \sum_{n=1}^{\infty} a_n^2. \tag{3.8}$$

The coefficients $\{a_n\}$ in the expansion are linear combinations of the moments of g(y).

Consider whether a finite number of terms in the series expansion provides an adequate approximation to g(y), the unknown true data generating process, the simplest case being the one in which we truncate the expansion after the first term. Then, f(y) is some baseline density and we wish to test its adequacy as an approximation to g(y). This is equivalent to the null hypothesis

(3.9) $$H_0\colon a_1 = a_2 = \ldots. = 0.$$

Omitting the observation subscript, from (3.6) we have

(3.10) $$\log g(y) = \log f(y) + \log[1 + \Sigma\, a_n P_n(y)]$$

(3.11) $$\nabla_{a_n} \log g(y)\Big|_{a_n=0} = P_n(y) \quad , \qquad i=1,2,\ldots.$$

where $\nabla_a = \partial/\partial a$. We wish to test H_0 without estimating a_n, that is to follow the score test approach. The score test will be based on $\mathbb{E}_0[\nabla_{a_n} \log g(y)\,|_{a_1=a_2=\ldots=a_n=0}] = 0$, which implies that

(3.12) $$\mathbb{E}_0[P_n(y)] = 0.$$

Thus, if the unknown true density g(y) admits a formal series expansion in terms of the baseline density f(y) and the corresponding orthonormal functions $P_n(y)$, then a test of the null hypothesis may be based on the formulation $\mathbb{E}_0[P_n(y)] = 0$, n=1,2, ..; that is, the expectation of the orthogonal functions under the null density is zero. While the preceding argument derives this test as a score test, note that (3.12) is implied by (3.7) under H_0. A comparison of (3.12) with (1.1) shows that any test based on an orthogonal polynomial is a CM test. The analysis leading up to (3.11) shows that every specification test based on an orthogonal polynomial is a score test against some alternative. A test based on the nth order orthonormal function is a test of the nth order moment restriction on the null density.

3.3 Implementation of tests

The conditional moment test based on the orthogonal polynomial will be based on

$$(3.13) \qquad m_{n,T}(\hat{\theta}_T) = T^{-1} \sum_{t=1}^{T} m_{n,t}(y, X, \hat{\theta} \mid X_T).$$

The asymptotic distribution of $m_{n,T}(\hat{\theta}_T)$ may vary with the estimator $\hat{\theta}_T$. Treatments are given in Newey (1985), Tauchen (1985), White (1987, 1990) and Pagan and Vella (1989), the last reference giving a particularly accessible presentation. In the special case where $\hat{\theta}_T$ is the ML estimator (see also Pierce (1982)) a $\chi^2(\dim(m_n))$ test statistic can be conveniently computed as T times the uncentered R^2 from the auxiliary regression of 1 on $\hat{m}_{n,t}$ and $\hat{s}_{\theta,t}$, where $\hat{m}_{n,t} \equiv m_n(y_t, X_t, \hat{\theta}_T)$ and $\hat{s}_{\theta,t}$ denotes the likelihood based score, $\partial \log L_t(\theta)/\partial\theta|_{\theta=\hat{\theta}}$. A second special case is where the following condition:

$$(3.14) \qquad \mathbb{E}_0[\nabla_\theta m_n(y, X, \theta \mid X)] = 0$$

holds. Then the asymptotic distribution of $m_{n,T}(\hat{\theta}_T)$ is the same as for $T^{1/2} m_{n,T}(\theta)$ (Newey (1985)) despite the substitution of $\hat{\theta}_T$ for θ, simplifying computation.

3.4 Local asymptotic power analysis

Valid CM tests of k-th order moment restrictions can also be derived using nonorthogonal polynomial functions. A leading example is testing for overdispersion in the Poisson regression model. A second central moment test may be based on $\{(y - \mu)^2 - \mu\}$, whereas the second order orthogonal polynomial is $\{(y - \mu)^2 - (y - \mu) - \mu\}$, i.e. $\{(y - \mu)^2 - y\}$. When premultiplied by a function $G_2(X,\theta)$ as in (3.2), these lead to two different test statistics. We emphasize that these different test statistics have different distributions under a given sequence of local alternatives. This is demonstrated analytically and by simulation in Cameron and Trivedi (1990a). Which CM test is more powerful? Clearly the answer depends on the alternative hypothesis, as any given CM test can be interpreted as a score (and hence locally most powerful) test against some alternative (White (1990)). Under the standard alternative hypothesis that the mean is correctly specified but y is negative binomial with a more general variance, the CM test based on the orthogonal polynomial coincides with the score test, and is therefore more powerful than the CM test based on the second central moment.

Similar results hold in more general settings. CM tests based on orthogonal polynomials differ from CM tests based on nonorthogonal polynomials. In a leading case which includes many examples where tests are based on polynomials, orthogonal polynomial tests are locally more powerful than tests based on nonorthogonal polynomials. In general the comparison leads to an ambiguous conclusion.

To demonstrate these results, consider testing of the n-th moment, given correct specification of the first (n-1) moments. Specifically, test:

$$H_0:\ \mathbb{E}_0[y^k \mid X,\theta] = \mu_k\ , \qquad k = 1,\ldots,n$$

against:

$$H_L:\ \mathbb{E}_L[y^k \mid X,\theta] = \mu_k\ , \qquad k = 1,\ldots,n-1$$
$$\mathbb{E}_L[y^n \mid X,\theta] = G_n(X,\theta)\cdot\alpha_n\ ,$$

where μ_k denotes moments under H_0, $\alpha_n = \delta/\sqrt{T}$ is a column vector of the same dimension (g) as the row vector G_n, and δ is a constant. Let $R_n(y) = R_n(y,X,\theta)$ be any polynomial function of y of degree n, with leading coefficient normalized to unity, such that $\mathbb{E}_0[R_n(y) \mid X,\theta] = 0$. Then under H_L given above:

$$(3.15) \qquad \mathbb{E}_L[R_n(y) \mid X,\theta] = G_n(X,\theta)\cdot\delta/\sqrt{T} .$$

The obvious CM test is based on:

$$(3.16) \qquad m_{n,T}(\hat{\theta}_T) = T^{-1}\sum_{t=1}^{T} \hat{G}_t{}'\hat{W}_t\hat{R}_t ,$$

where $G_t = G(X_t,\theta)$, $R_t = R(y_t,X_t,\theta)$, $W_t = W(X_t,\theta)$ is a scalar weight, and $\hat{G}_t$, $\hat{R}_t$, and $\hat{W}_t$ are evaluated at $\hat{\theta}_T$. We analyze the asymptotically equivalent quantity:

$$(3.17) \qquad \hat{\alpha}_{n,T} = (\sum_{t=1}^{T} \hat{G}_t{}'\hat{W}_t\hat{G}_t)^{-1}\cdot\sum_{t=1}^{T} \hat{G}_t{}'\hat{W}_t\hat{R}_t ,$$

which can be interpreted as the coefficient from regression of $\hat{R}_t$ on $\hat{G}_t$ with weights $\hat{W}_t$.

3.4.1 Power when derivative condition (3.14) is satisfied.

When (3.14) is satisfied, under H_L:

$$(3.18) \qquad T^{1/2}\hat{\alpha}_{n,T} \overset{d}{\to} \mathbb{N}[\delta,\ \Sigma_{GWG}^{-1}\cdot\Sigma_{GW\Omega WG}\cdot\Sigma_{GWG}^{-1}\ ],$$

where $\Sigma_{GWG} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t{}'W_tG_t$; $\Sigma_{GW\Omega WG} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t{}'W_t\Omega_tW_tG_t$; and $\Omega_t = \mathbb{E}_L[R_tR_t' \mid X_t]$ is the unspecified conditional variance of R_t. It follows that the test statistic:

$$(3.19) \qquad \tau = [\sum_{t=1}^{T} \hat{R}_t\hat{W}_t\hat{G}_t]\cdot[\sum_{t=1}^{T} \hat{G}_t{}'\hat{W}_t\hat{\Omega}_t\hat{W}_t\hat{G}_t]^{-1}\cdot[\sum_{t=1}^{T} \hat{R}_t\hat{W}_t\hat{G}_t]$$

has limiting chisquare distribution with q degrees of freedom under H_0, and limiting noncentral chisquare distribution under H_L with noncentrality parameter λ where

$$\lambda = \delta'\Sigma_{GWG}\cdot\Sigma^{-1}_{GW\Omega WG}\cdot\Sigma_{GWG}\delta. \tag{3.20}$$

Power is maximized when the noncentrality parameter, which depends upon the misspecification indicators G_t, the weights W_t and the variance Ω_t of the polynomial R_t, is minimized. Given Ω_t, which requires knowledge of the first 2n moments, the noncentrality parameter is maximized when $\lambda = \delta'\Sigma_{GWG}\delta$, where $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} G_t'\Omega_t^{-1}G_t$. This in turn is maximized when Ω_t is minimized. But from the uniqueness and minimum variance properties of orthogonal polynomials (appendix B), $\Omega_t = \mathrm{Var}(R_t)$ is uniquely minimized amongst monic polynomials when R_t is the orthonormal polynomial. Hence orthonormal polynomials lead to locally most powerful CM tests based on n-th order polynomials.

3.4.2 Power when derivative condition (3.14) is not satisfied

In some situations the derivative condition (3.14) will not be satisfied. To compare the asymptotic variances of $\hat{\alpha}_T$ in the case of orthogonal and nonorthogonal moment functions, we treat it as a sequential estimator and follow the approach of Newey (1984). We consider the case of p moment restrictions and (p×1) parameter α. Let $\hat{\theta}_T$ be the solution to the first order conditions $\sum_{t=1}^{T} s(y_t,X_t,\hat{\theta}_T)= 0$. The joint estimating equations for $\hat{\theta}_T$ and $\hat{\alpha}_T$ are:

$$\sum_{t=1}^{T} s(y_t,X_t,\hat{\theta}_T) = 0. \tag{3.21}$$

$$\sum_{t=1}^{T} G(X_t,\hat{\theta}_T)'\, W(X_t,\hat{\theta}_T)\{R(y_t,X_t,\hat{\theta}_T) - G(X_t,\hat{\theta}_T)\cdot\hat{\alpha}_T\} = 0. \tag{3.22}$$

This is a specialization of the estimator $\sum_{t=1}^{T} q(y_t, X_t, \hat{\beta}_T) = 0$, where $\beta' = (\theta'\ \alpha')$, is ((q+p)×1) vector with components $q_{1t}(\beta) = s(y_t, X_t, \theta)$ and $q_{2t}(\beta) = G(X_t,\theta)'\ W(X_t,\theta)\cdot\{m(y_t,X_t,\theta) - G(X_t,\theta)'\cdot\alpha\}$. For the estimator which satisfies the first order conditions the asymptotic distribution of $\hat{\beta}_T$ is given by:

$$(3.23) \qquad T^{1/2}(\hat{\beta}_T - \beta) \xrightarrow{d} \mathbb{N}[0,\ A(\beta)^{-1}\cdot B(\beta)\cdot A(\beta)^{-1\prime}]$$

where

$$(3.24) \qquad A(\beta) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}[\nabla_\beta q(y_t, X_t, \beta)]$$

$$(3.25) \qquad B(\beta) = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \mathbb{E}[q(y_t, X_t, \beta)\cdot q(y_t, X_t, \beta)'].$$

Partition A and B conformably with q: $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$; $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$.

Then using partitioned inverse, together with the simplification $A_{12} = 0$, it can be established that under H_L:

$$(3.25) \qquad T^{1/2}\hat{\alpha}_T \xrightarrow{d} \mathbb{N}[\upsilon,\ V]$$

where

$$\upsilon = \delta - A_{21}A_{11}^{-1}A_{11}^{-1} \lim_{T\ \infty} T^{-1} \sum_{T=1}^{T} \mathbb{E}_L[s(y_t, X_t, \theta_0) \mid X_t]$$

$$V = A_{22}^{-1}(B_{22} + A_{21}A_{11}^{-1}B_{11}A_{11}^{-1}A_{12} - A_{21}A_{11}^{-1}B_{12} - B_{21}A_{11}^{-1}A_{12})A_{22}^{-1}$$

$$A_{11} = \lim_{T\ \infty} T^{-1} \sum_{T=1}^{T} \mathbb{E}_L[\nabla_\theta\ s(y_t, X_t, \theta_0) \mid X_t]$$

$$A_{12} = 0$$

$$A_{21} = \lim_{T\ \infty} T^{-1} \sum_{T=1}^{T} \mathbb{E}_L[G(X_t,\theta_0)'W(X_t,\theta_0)\ \nabla_\theta R(y_t, X_t, \theta_0) \mid X_t]$$

$$A_{22} = \lim_{T\ \infty} T^{-1} \sum_{T=1}^{T} G(X_t,\theta_0)'W(X_t,\theta_0)G(X_t,\theta_0)$$

$$B_{11} = \lim_{T \to \infty} T^{-1} \sum_{T=1}^{T} \mathbb{E}_L[s(y_t,X_t,\theta_0)\cdot s(y_t,X_t,\theta_0)' \mid X_t]$$

$$B_{12} = B_{21}'$$

$$B_{21} = \lim_{T \to \infty} T^{-1} \sum_{T=1}^{T} \mathbb{E}_L[G(X_t,\theta_0)'W(X_t,\theta_0)R(y_t,X_t,\theta_0)s(y_t,X_t,\theta_0) \mid X_t]$$

$$B_{22} = \lim_{T \to \infty} T^{-1} \sum_{T=1}^{T} \mathbb{E}_L[G(X_t,\theta_0)'W(X_t,\theta_0)R(y_t,X_t,\theta_0) R(y_t,X_t,\theta_0)'W(X_t,\theta_0)G(X_t,\theta_0) \mid X_t] ,$$

and θ_0 is now the pseudo-true value under H_L (White (1982)).

Consider tests based on different polynomial functions. Under H_L, these converge to noncentral chi-square distributions with noncentrality parameters $\lambda^{(1)} = \upsilon^{(1)\prime}V^{(1)}\upsilon^{(1)}$ and $\lambda^{(2)} = \upsilon^{(2)\prime}V^{(2)}\upsilon^{(2)}$ using (3.25), where the superscripts (1) and (2) denote respectively tests based on orthogonal polynomial function $R_1(y_t,X_t,\theta)$ and nonorthogonal polynomial function $R_2(y_t,X_t,\theta)$. Let $B_{ij}^{(k)}$ (i,j,k =1,2) denote the partitions of the B matrix. Note that $B_{11}^{(1)} = B_{22}^{(2)} = B_{11}$.

We first assume $\upsilon^{(1)} = \upsilon^{(2)}$. Two special cases are considered.

Case 1: s(•) is the likelihood score function and first order orthogonal polynomial; $\hat{\theta}$ is MLE; R(•) is second or higher order orthogonal polynomial for testing higher second or higher moment specification of the model. Assume $\mathbb{E}_L[R(\cdot)s(\cdot)] = 0$, which implies $B_{21}^{(1)} = 0$. For a nonorthogonal moment function $B_{21}^{(2)} \neq 0$; also $(B_{22}^{(2)} - B_{22}^{(1)})$ is positive definite because of the minimum variance property of orthogonal polynomials. However, since

$$V^{(1)} = A_{22}^{-1}[B_{22}^{(1)} + A_{21}A_{11}^{-1} B_{11}A_{11}^{-1}A_{12}]A_{22}^{-1} ,$$

and $$V^{(2)} = A_{22}^{-1}[B_{22}^{(2)} + A_{21}A_{11}^{-1}B_{11}A_{11}^{-1}A_{12} - A_{21}A_{11}^{-1}B_{12}^{(2)} - B_{21}^{(2)} A_{11}^{-1}A_{12}]A_{22}^{-1},$$

the difference $(V^{(2)} - V^{(1)})$ is indeterminate without additional structure.

Case 2: Suppose that the density assumed under H_0 obeys regularity conditions such that, in addition to the assumptions of Case 1, the generalized information equality applies so that $A_{21} = B_{21}$, and $A_{11} = B_{11}$.

Then $\quad V^{(1)} = A_{22}^{-1} B_{22}^{(1)} A_{22}^{-1}$

and $\quad V^{(2)} = A_{22}^{-1}[B_{22}^{(2)} - B_{21}^{(2)} B_{11}^{-1} B_{12}^{(2)}]A_{22}^{-1}$.

Once again without additional assumptions the difference $(V^{(2)} - V^{(1)})$ is indeterminate. Ranking in terms of power of tests based on orthogonal and nonorthogonal moment functions is therefore difficult.

In general, the means of the moment functions under H_L will differ, i.e. $\upsilon^{(1)} \neq \upsilon^{(2)}$. Analysis will be even less conclusive in this case. To the extent that the power of CM tests has been analyzed, stronger results are obtained when attention is confined to the choice of misspecification indicators rather than the underlying function of y (or generalized residual). See Newey (1985), Bierens (1990).

The preceding analysis demonstrates that orthogonal polynomials lead to more powerful tests than nonorthogonal polynomials when (3.14) is satisfied, as in the case of LEF-QVF discused in section 4. Even when (3.14) is not satisfied examples can be found in which orthogonal polynomials are again superior. For example, the score test of $N(\mu, \mu)$ against $N(\mu, \mu+\alpha\cdot g(\mu))$ is the same as a CM test based on the second-order orthonormal polynomial that is obtained by application of (3.5).

We now consider joint tests of the first n moments, rather than testing the n-th order moment conditional on correct specification of the first (n-1). Since a *set* of moment functions have more than one parameterization, they may be equivalent for testing purposes. To the extent that an orthogonal set may be transformed to an equivalent nonorthogonal set, there is no theoretical advantage in using the latter. This additionally requires transformation of misspecification indicators, however, which in practice is not done. Investigators typically perform CM tests with a given set of misspecification indicators. Furthermore, sequential rather than joint tests are the norm in applied work. In sequential testing, orthogonal polynomial tests have the advantage that they do not require a fixed n. For testing an individual or a subset of moment restrictions, with n not fixed, different moment functions do not have the same size and power properties.

4. APPLICATION TO SPECIFICATION TESTS IN THE LEF-QVF

4.1 Conditional Moment tests based on orthogonal polynomials for the LEF-QVF

To illustrate the use of orthogonal polynomials as the basis for the choice of moment function, we consider linear exponential families (LEF) with quadratic variance functions (QVF). This covers many commonly used econometric models: regression models under normality with constant variance; discrete choice models such as probit and logit; 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. To keep the focus on essentials the detailed statement of the LEF-QVF class is given in Appendix B.

In regression applications of the LEF, regressors X_t are introduced via the mean parameter, $\mu_t = \mu(X_t,\theta)$, and possibly via the parameters v_0, v_1 and v_2 of the QVF, $V(\mu_t) = v_0 + v_1\mu_t + v_2\mu_t^2$ defined in (B.4), which may be parameterized in terms of μ_t and a nuisance parameter α. The function μ is such that the parameters θ can be identified (McCullagh and Nelder (1983)). Note that some or all of v_0, v_1 and v_2 will be known. As discussed in section 3.3, the procedure is to progressively test for n = 1,2,..

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

$$(4.2) \qquad m_n(y_t,X_t,\theta) = G_n(X_t,\theta)\cdot P_n(y_t,\mu(X_t,\theta)),$$

for some chosen function $G_n(X_t,\theta)$, where for simplicity we have suppressed the nuisance parameter α. The recurrence relation (B.6) generates $P_n(y_t,\mu(X_t,\theta))$. The variance of $P_{n,t}$ for the optimal test is easily obtained using (B.8) and (B.10). Tests based on different degrees of polynomial are orthogonal by (B.9). The first three orthonormal polynomials, denoted by Q, for several members of the LEF-QVF are given in Table 1; in the binomial and the negative binomial cases we give orthogonal polynomials.

The orthogonal polynomials can be directly used to generate specification tests for the associated distributions, which span most of the standard parametric models for the various types of cross-section data. In some cases these tests are new, while in other cases these tests coincide with existing tests, as indicated below in the discussion for specific examples. In all cases these tests are simpler to obtain than the usual approach of embedding the distribution of a specific LEF-QVF member, e.g. Poisson, in a more general parametric family, e.g. negative binomial, and algebraically obtaining the score test. And by property (B.10) CM tests based on the orthogonal polynomials for LEF-QVF without nuisance parameter α satisfy the derivative condition (3.14).

4.2 Discussion for specific members of the LEF-QVF

In discussing results for specific members of the LEF-QVF, we focus on comparison with tests obtained by other methods. This is done to illustrate that basing CM tests on orthogonal polynomial is not only simple, but is capable of leading to standard tests obtained by other methods. The orthogonal polynomial approach may of course be applied in any testing situation where necessary moments are defined, including situations where no standard tests exist. Two of many such examples are given towards the end of section 4.2.

Since the tests here are based on parameter estimates under the null hypothesis only, it is natural to compare the tests with score tests, where such tests have appeared in the literature, rather than Wald or likelihood ratio tests. We also compare the tests with CM tests based on the k-th order moments $y^k - \mathbb{E}[y^k | X,\theta]$ or $(y-\mu)^k - \mathbb{E}[(y-\mu)^k | X,\theta]$, and with the information matrix test for the normal.

We say that tests coincide if y_t appears in the corresponding moment condition only via the function $P_n(y_t, \mu(X_t,\theta))$. In particular, consider a score test based on the alternative hypothesis density $g(y_t,X_t,\theta,\gamma)$ such that $g(y_t,X_t,\theta,\gamma=\gamma^*)$ yields the null hypothesis LEF-QVF density. If the factorization

$$\nabla_\gamma \ln g(y_t,X_t,\theta,\gamma)\big|_{\gamma=\gamma^*} = G_n^*(X_t,\theta)\cdot P_n(y_t, \mu(X_t,\theta))$$

occurs, for some $G_n^*(X_t,\theta)$, then the score test coincides with the CM test based on the nth order orthogonal polynomial.[1]

Tests based on $P_{1,t} = (y_t - \mu_t)$ coincide with score tests of omitted variables from the conditional mean function in an LEF-QVF model (Cameron and Trivedi (1990b)). They also coincide with the score test for misspecified functional form of the conditional mean, where the alternative hypothesis model is embedded in an LEF-QVF, (Gurmu and Trivedi (1990b)).

Tests based on $P_{2,t} = \{(y_t - \mu_t)^2 - a(\mu_t)(y_t - \mu_t) - \mu_t\}$, where the function $a(\mu_t)$ differs for different members of the LEF-QVF, correspond in some cases to score tests for misspecification of the conditional variance. These tests differ from tests based on the second central moment, i.e.

---

[1] We note that often the score tests that appear in the literature lead to a very specific choice of $G_n^*(X_t,\theta)$, whereas we permit quite general choice, as in (3.2). For example, see the discussion for the normal distribution.

$\{(y_t - \mu_t)^2 - \mu_t\}$, unless $a(\mu_t) = 0$ which is the case for the normal but not for any other LEF-QVF member.

We consider some examples of tests based on second and higher order polynomials in each of the LEF-QVF families.

Example 1 - *Normal*: CM tests based on the order two polynomials are equivalent to the standard score test for heteroskedasticity. CM tests based on order two, three and four orthogonal polynomials are tests of heteroskedasticity, skewness and non-normal kurtosis identical to score tests against the Pearson system.

Versions of these score tests given by Bera and Jarque (1982) differ in the following way. For testing symmetry, the orthogonal polynomial approach leads to a CM test of $\mathbb{E}[G_3(X_t,\theta)\{(y_t - \mu_t)^3 - 3\sigma^2(y_t - \mu_t)\}]$, whereas Bera and Jarque (1982), and many others following, instead use the modification $\mathbb{E}[G_3(X_t,\theta)\cdot(y_t - \mu_t)^3]$. Similarly for non-normal kurtosis the test of $\mathbb{E}[G_4(X_t,\theta)\{(y_t - \mu_t)^4 - 6\sigma^2(y_t - \mu_t)^2 + 3\sigma^4\}]$ is usually modified to the fourth central moment $\mathbb{E}[G_4(X_t,\theta)\{(y_t - \mu_t)^4 - 3\sigma^4\}]$. These lead to tests which, in general, are not asymptotically equivalent under the alternative hypothesis. For the specific alternative considered by Bera and Jarque, however, they do coincide.[2]

Hall (1987) has shown that for the general linear regression model with normal errors and correctly specified conditional mean function $(X_t'\beta)$, the information matrix test of White (1982) can be decomposed into three

---

[2] For particular choices of $G_3(.)$, $G_4(.)$ and estimator, the sample moments corresponding to the above moment functions may coincide. This will be true when the following conditions are satisfied: $\Sigma_t\ G_{3,t}(.)\cdot(y_t-\hat{\mu}_t)= 0$, and $\Sigma_t\ G_{4,t}(.)(y_t-\hat{\mu}_t)^2 = \hat{\sigma}^2\Sigma_t G_{4,t}(.)$. In Bera and Jarque (1982), $G_{3,t} = X_t$, $G_{4,t} = 1$, $\hat{\beta}$ and $\hat{\sigma}^2$ are ML estimates, so these conditions are satisfied as they are the first-order conditions for the MLE.

components which are the Bera-Jarque tests for heteroskedasticity, skewness and non-normal kurtosis. The OPS approach suggests a wider range of simultaneous ("portmanteau") tests of homoskedasticity, zero skewness and non-normal kurtosis by using different linear combinations of $P_{2,t}$, $P_{3,t}$ and $P_{4,t}$, rather than linear combinations of P_{2t}, $(y_t - \mu_t)^3$, and $(y_t - \mu_t)^4 - 3\sigma^4$. Linear dependence of the orthogonal polynomials implies additivity property of the former simultaneous test.

Example 2 - *Poisson*: The Poisson density is the benchmark model for count data, where y_t takes values 0,1,2,... A common feature of count data is that, conditional on regressors, the variance exceeds the mean (overdispersion), whereas the Poisson imposes variance-mean equality. Tests for overdispersion are the analogues of tests of heteroskedasticity in the normal case. CM tests of overdispersion in the Poisson may be based on the second order polynomial $P_{2,t} = (y_t - \mu_t)^2 - y_t$. As noted in section 3.4, this leads to different CM tests than those based on the more obvious $(y_t - \mu_t)^2 - \mu_t$, the difference between the second-order central moment, $(y_t - \mu_t)^2$, and its expectation under the null hypothesis. The usual score test for Poisson versus the negative binomial (or more generally the Katz system) coincides with the test based $P_{2,t}$.

CM tests for non-Poisson skewness may be based on the third order polynomial given in Table 1. Lee (1986, equation (5.12)) derives essentially the same test as a score test of the Poisson against a truncated Gram-Charlier series expansion. A considerably simpler derivation of Lee's test results is possible if one directly exploits knowledge of the Poisson-Charlier orthogonal polynomials together with the results of section 3, as has been done here.

Example 3 - Exponential: CM tests of the exponential may be readily obtained using Laguerre polynomials in Table 1. A special case, the unit exponential arises in diagnostic tests for any uncensored parametric hazard model, where the generalized residuals ε_t are defined by the integrated hazard function which has a unit exponential distribution. Then $\mathbb{E}_0[\varepsilon^j] = j!$, $j \geq 0$.

In duration models a likely source of misspecification is neglected heterogeneity, which leads to generalized residuals ε_t having non-unitary variance. The CM test based on $P_{2,t}$ can be used to test for zero neglected heterogeneity. Lancaster (1985) considers the score test against a non-unitary constant variance. Then the CM test based on $P_{2,t} = \varepsilon_t^2 - 2\varepsilon_t - 1$ reduces to a second moment test based on $\varepsilon_t^2 - 1$ since $\Sigma_t \varepsilon_t = T$. However, for a test against a heteroskedastic variance, the CM test based on $P_{2,t}$ will differ from the second moment test.

The interested reader may readily construct several new specification tests using the results in Table 1, along the lines of Cameron and Trivedi (1990c). Two examples are given.

Example 4 - Negative Binomial. The negative binomial is the standard fully parametric model for overdispersion in count data. To test whether the negative binomial, with mean μ and variance $\mu + \mu^2/\alpha$ adequately models overdispersion, the results in this paper suggest using a CM test based on $P_{2,t}(y_t) = (y_t - \mu_t)^2 - (1+(\mu_t/\alpha))y_t$.

Example 5 - Binomial. For the binomial with n trials,[3] use CM tests based on $P_{2,t}(y_t) = (y_t - \mu_t)^2 + (2(\mu_t/n)-1)y_t - \mu_t(1-(\mu_t/n))$.

Implementation of the above tests for LEF-QVF examples is usually straight-forward since by (B.10), $\mathbb{E}[\nabla_\mu P_n] = \mathbb{E}[(-a_n/a_{n-1})P_{n-1}] = 0$, so (3.14) holds. To the extent that no nuisance parameters are present, the computation of the asymptotic variance of the moment function and the test is simpler.

---

[3] In the case of the binomial density with n=1, e.g. probit or logit models, it is meaningless to make tests based on polynomials of order higher than one because the density has only two support points.

This is also the case for the normal density, in spite of the presence of the nuisance parameter σ^2.

Thus to implement all of the CM tests given in this section aside from the negative binomial example, but additionally including tests for the binomial with known number of trials (n>1), we need simply regress 1 on $G_{n,t}(X_t, \hat{\theta}_T) \cdot P_{n,t}(X_t, \hat{\theta}_T)$. T times the uncentered R^2 from this regression is χ^2 with degrees of freedom $\dim(G_{n,t})$ under H_0.

Furthermore, note that the orthogonal polynomials approach leads exactly to tests for which (3.14) holds, whereas this is generally not the case when k-th order moments alone are used. For example, for skewness in example 1 we have $P_{3,t} = (y_t - \mu_t)^3 - 3\sigma^2(y_t - \mu_t)$, whereas most authors use $P_t = (y_t - \mu_t)^3$ which does not lead to tests for which (3.14) holds.

5. ORTHOGONAL POLYNOMIAL TESTS FOR TRUNCATED MODELS

Truncated regression models provide a useful illustration of the differences between OPS based CM tests and conventional score tests. Truncated distributions feature widely in applied econometric work but there is little consensus on the use of appropriate diagnostic tools. The application of diagnostic tests in such cases is especially desirable since the failure of common distributional assumptions such as homoskedasticity of the latent dependent variable in a Tobit type model can have serious implications for consistency, not just efficiency (Amemiya (1985)). The diagnostic tests for these models are often cumbersome to derive and to compute as evidenced by, *inter alia,* Bera, Jarque and Lee (1984), Lee and Maddala (1985), Robinson, Bera and Jarque (1985), Gurmu and Trivedi (1992). Computation of the CM test may be simplified using the popular OPG variant of the information matrix, but this frequently has unsatisfactory properties.

In this section we consider CM tests derived using orthogonal polynomials when the baseline sample density is obtained by restricting the set of support points for the parent distribution. Kiefer (1985) for the censored exponential and Smith (1989) for some limited dependent variable models use orthogonal polynomial expansions, but in terms of the underlying latent variable. Smith actually bases tests on k-th order moments of the latent variable; see also Pagan and Vella (1989). In our approach we consider the series expansion (3.6) around a truncated null density:

$$g(y) = f^*(y)\cdot[1 + \sum_{n=1}^{\infty} a_n^* P_n^*(y)]. \tag{5.1}$$

$$f^*(y_t | y_t \in \mathbb{Y}) = \frac{f(y_t)}{1 - F(y_t | y_t \in \mathbb{Y})}, \tag{5.2}$$

where $f^*(y)$ is a truncated density with support points restricted to set $\mathbb{Y}$, $P_n^*(y)$ are corresponding orthogonal polynomials, and a_n^* are functions of the moments of $g(y)$. Specification tests of the truncated distribution based on $P_n^*(y)$ are tests of the null H_0: $a_1^* =..= a_n^* = 0$. The results given in section 2.2 can be used to derive orthogonal polynomials for the truncated case after interpreting all relevant moments as those of the truncated distribution. Since the OPS for a given baseline distribution is unique, the OPS based CM criteria will be different from those in the regular (untruncated) case. But the approach to the construction of CM tests is unchanged.

Compare the above strategy with that used in the construction of a score test where the starting point is likely to be the selection of a baseline truncated model and a *truncated* alternative. (For example, Gurmu and Trivedi (1992) derive a score test of overdispersion for the truncated Poisson regression as the null model and the truncated negative binomial as the alternative.) Though the OPS based CM test is a score test, it will be based on an implicit alternative density $g(y)$, which in general will be different from that used in the derivation of a score test against a specific alternative. Therefore, OPS based CM tests may differ for truncated distributions even when they coincide for the untruncated counterparts.

Further, conventionally designed score tests of moment restrictions in truncated models are generally not independent. This feature of score tests appears in the context of some non-truncated models where the parameterization of the model does not lead to a block diagonal information matrix. The OPS based CM tests have the independence property by design, but the test may be based on an implied direction of departure from the null different from that of another conventionally designed score test.

As an illustration we reconsider the example of left truncated Poisson distribution analyzed in Gurmu and Trivedi (1992). Let the untruncated Poisson pdf be $h(y_t,\psi_t)= \exp(\psi_t)\psi_t{}^{y}t/y_t!$ where ψ_t is the untruncated mean, usually specified to be log-linear in a set of exogenous variables. Consider the positive Poisson (Poisson distribution without zeroes). This has the pdf $h(y_t)/(1 - h(0))$, or

$$(5.3) \qquad f(y_t,\psi_t|y_t \geq 1) = \frac{\psi_t^{y_t}}{(\exp(\psi_t) - 1)\cdot y_t!}.$$

The first three (truncated) moments of the positive Poisson are as follows: $\mu_{1t} = \psi_t + \delta_t$; $\mu'_{2t} = \psi_t - \delta_t(\mu_{1t} - 1)$; $\mu'_{3t} = \psi_t + 2\mu_{1t}(\mu_{1t}^2 - \psi_t^2) + \delta_t(3\psi_t + 1)$; where $\delta_t = \psi_t/[(\exp(\psi_t) -1)]$.

We may construct an OPS based CM test of the second moment restriction using (2.15), which yields

$$(5.5) \qquad P_2(y_t,X_t,\theta) = \varepsilon_t^2 - (\mu'_{3t}/\mu'_{2t})\varepsilon_t - \mu'_{2t},$$

where $\varepsilon_t = (y_t - \mu_{1t})$. By contrast the score function given in Gurmu and

Trivedi (1992) is the sum of terms that are a multiple (not depending on y_t) of the polynomial

$$(5.10) \qquad P_{score}(y_t, X_t, \theta) = (\varepsilon_t^2 - y_t) + (\varepsilon_t - y_t)\delta_t .$$

Evidently, unlike the case of untruncated Poisson considered in example 2 of the last section, in the truncated case the OPS-based CM test is different from that based on the score function.

6. CONCLUDING REMARKS

Orthogonal functions offer a new and convenient approach to specifying CM functions and deriving CM tests. Formulae given in this paper permit construction of orthogonal polynomials, particularly of low order such as in (3.5), in general settings. Even simpler formulae are presented in Table 1 for members of the LEF-QVF families, which subsume a wide range of commonly used econometric models.

For the LEF-QVF examples, to the extent that CM tests based on orthogonal polynomials coincide with existing tests, these tests are score tests. By contrast, in the example of truncated models, CM tests based on orthogonal polynomials differ from existing score tests. These examples are a small subset of the possible applications of the orthogonal polynomial approach.

OPS based CM tests are designed to be orthogonal in a specific sense. CM tests based on sequential k-th order moments, and the conventional score approach in which one examines the departure from the null in one direction at a time, do not in general ensure orthogonality of tests. The linear independence of tests based on OPS is an important advantage in some situations. For example, separate tests of homoskedasticity and normality in

Tobit type models are correlated. Yet even in high level applied work investigators sometimes apply diagnostic tests one at a time, ignoring possible correlation. When the tests are not independent, the interpretation of the test outcome is problematic since the tests will not then have the nominal asymptotic size. The orthogonal polynomial approach may have an advantage in such cases.

Table 1: Orthonormal polynomial functions for selected members of LEF-QVF

Normal $f(y) = (2\pi\sigma^2)^{-1/2}\exp(-\frac{1}{2\sigma^2}\cdot(y-\mu)^2)$ $-\infty < y < \infty \quad E[y] = \mu, \ V(\mu) = \sigma^2$	*Hermite* $Q_1(y) = (y-\mu)/\sigma$ $Q_2(y) = \{(y-\mu)^2 - \sigma^2\}/\sqrt{2}\sigma^2$ $Q_3(y) = \{(y-\mu)^3 - 3\sigma^2(y-\mu)\}/\sqrt{6}\sigma^3$
Poisson $f(y) = \lambda^y e^{-\lambda}/y!$, $y=0,1,... \quad E[y] = \mu = \lambda, \quad V(\mu) = \mu$	*Poisson-Charlier* $Q_1(y) = (y-\mu)/\sqrt{\mu}$ $Q_2(y) = \{(y-\mu)^2 - y\}/\sqrt{2}\mu,$ $Q_3(y) = \{(y-\mu)^3 - 3(y-\mu)^2 -$ $(3\mu-2)(y-\mu) + 2\mu\}/\sqrt{6}\mu^{3/2}$
Gamma $f(y) = \alpha\exp(-\alpha y)$ $\mathbb{E}[y] = \alpha; \ \mathrm{var}[y] = \alpha^2$	*Generalized Laguerre* $Q_1(y) = (y-\alpha)/\alpha$ $Q_2(y) = ((y-\alpha)^2 - 2\alpha(y-\alpha) - \alpha^2)/2\alpha^2$ $Q_3(y) = ((y-\alpha)^3 - 6\alpha(y-\alpha)^2 +$ $3\alpha^2(y-\alpha) + 4\alpha^3)/(6\alpha^3)$
Binomial $f(y) = \binom{n}{y} p^y(1-p)^{n-y}, \quad y = 0,1,..,n$ $\mu = \mathbb{E}[y] = np; \ V(\mu)=\mu(1-\mu)/n$	*Krawtchouk* $P_1(y) = (y-np)$ $P_2(y) = (y-np)^2+(2p-1)y - np(1-p)$ $P_3(y) = (y-np)^3+ (6p-3)y^2 +$ $+\{3(n+2)p^2 - 3(n+2)p + 2\}y$ $+ 2np(2p^2 - 3p + 2)$
Negative Binomial $f(y) = (1+\theta)^{-\alpha-y} \theta^y \binom{y+\alpha-1}{y},$ $y=0,1,..; \ \mu = \alpha\theta, \ V(\mu) = \mu + \mu^2/\alpha$	*Meixner* $P_1(y) = (y-\mu)$ $P_2(y) = \{(y-\mu)^2 - (1 + (\mu/\alpha))y\}$ $P_3(y) = \{(y-\mu)^3 - 3(2\theta+1)(y-\mu)^2$ $+ \{3(2-\alpha)\cdot(\theta^2+\theta) + 2\}(y-\mu)$ $+ \{2\alpha\theta\cdot(2\theta^2+3\theta+1)\}$

Appendix A

We shall review a number of important results on orthogonal polynomials. No proofs are given and the interested reader may wish to consult Cramer (1946), Lancaster (1969), and Szegö (1975) for proofs and further details.

Existence: Given a random variable y with distribution function F(y) and density dF(y)=f(y)dy, for an arbitrary real moment sequence $\{\mu_n\}$ to give rise to an OPS unique up to an arbitrary constant, a necessary and sufficient condition is that the determinants $\|\Delta_n\|$ are positive where $\Delta = [\Delta_{ij}]$ and $\Delta_{ij} = \mu_{i+j-2}$, where moments may be taken either about the mean or an arbitrary origin;

(A.1)
$$\|\Delta_n\| \equiv \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{n+1} \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \mu_{n+2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \mu_n & \mu_{n+1} & \cdot & \cdots & \mu_{2n} \end{vmatrix} > 0, \quad n=0,1,2,\ldots$$

For proof see Cramer (1946, chapter 12.6) or Szegö (1975, chapter II).

The determinant in (A.1) may be partitioned as follows:

(A.2)
$$\|\Delta_n\| = \left\| \begin{array}{c:c} \Delta_{n-1} & d \\ \hdashline d' & \mu_{2n} \end{array} \right\| .$$

For a positive definite Δ_n, Δ_n^{-1} exists $\forall$ n, and the application of the bordered determinant theorem yields the following alternative representation:

(A.3)
$$\|\Delta_n\| = \mu_{2n}\|\Delta_{n-1}\| - d' \, Adj\, (\Delta_{n-1})\, d$$

where
$$d' = (\mu_n \; \mu_{n+1} \cdots \mu_{2n-1}); \quad \|\Delta_{-1}\| = \|\Delta_0\| = 1.$$

The above discussion has assumed an infinite number of points of increase but the results will apply to finite discrete distributions if only polynomials of degree less than the number of points of increase are considered.

Derivation of the orthonormal polynomial: For a given moment sequence $\{\mu_n\}$ the orthonormal OPS, with leading coefficient one (i.e. monic), can be generated by the following relationship:

$$P_n(y) = [\|\Delta_{n-1}\|]^{-1} \cdot \|D_n(y)\| \qquad \text{where} \tag{A.4}$$

$$\|D_n(y)\| = \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{n+1} \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \mu_{n+2} \\ \vdots & & & & \vdots \\ \mu_{n-1} & \cdots & & \cdots & \mu_{2n-1} \\ 1 & y & y^2 & \cdots & y^n \end{vmatrix} \neq 0, \quad n = 0,1,2,\ldots \tag{A.5}$$

Equation (A.4) is the solution of (2.2) with k_n=1, which establishes that the result is an orthonormal polynomial.

Derivation of the result (A.4): First partition $\|D_n(y)\|$ as follows:

$$\|D_n(y)\| = \left\| \begin{array}{c:c} \Delta_{n-1} & d \\ \hdashline c'(y) & y^n \end{array} \right\| . \tag{A.6}$$

The partitioned bordered determinant theorem yields

$$\|D_n(y)\| = y^n \|\Delta_{n-1}\| - c'(y)\, \mathit{Adj}\, (\Delta_{n-1})\, d \tag{A.7}$$

$$\frac{\|D_n(y)\|}{\|\Delta_{n-1}\|} = y^n - c'(y)\, [\Delta_{n-1}]^{-1} d \tag{A.8}$$

where $c'(y) = (1 \quad y \quad y^2 \quad y^3 \ldots\ldots y^{n-1})$.

Uniqueness: The orthonormal (monic) polynomial sequence $\{P_n(y)\}$ is unique. If $\{Q_n(y)\}$ is also an OPS, then there exist constants $c_n \neq 0$ such that $Q_n(y) = c_n P_n(y)$, n=0,1,2,..

Completeness: An orthonormal polynomial sequence is complete if any function $\psi(y)$ has $\text{var}[\psi(y)] = \sum_{i=1}^{\infty} a_i^2 < \infty$ where $a_i = \mathbb{E}[\psi(y)P_i(y)]$; for proof see Lancaster (1969, chapter 4.4).

Covariance properties: For an OPS $\{P_n(y)\}$ and for every polynomial $R_m(y)$, $m \leq n$, (i) $\mathbb{E}[R_m(y)P_n(y)] = 0$ for $m < n$; (ii) $\mathbb{E}[R_m(y)P_n(y)] \neq 0$ for $m = n$; (iii) $\mathbb{E}[y^m P_n(y)] = k_n \delta_{mn}$, $k_n \neq 0$, for $m \leq n$.

Let $P_n(y)$ be an orthonormal polynomial, and $\pi_n(y)$ be any other orthonormal polynomial. Then

(A.9) $$\mathbb{E}[\pi_n(y)P_n(y)] = \|\Delta_n\| \,/\, \|\Delta_{n-1}\| \ , \ \|\Delta_{-1}\| = 1.$$

Minimum variance property: (i) If $\|\Delta_n\| > 0$ $(n \geq 0)$, then the orthonormal polynomial $P_n(y)$ satisfies the following property for every non-orthonormal orthogonal polynomial $\pi_n(y) \neq P_n(y)$: $\text{Var}(P_n(y)) < \text{Var}(\pi_n(y))$. (ii) In the monic class of polynomials of degree n the orthogonal polynomial has the smallest variance.

Appendix B

A particularly helpful reference for orthogonal polynomials in the LEF-QVF is Morris (1982). The LEF is defined by

(B.1) $$f(y,\psi) = \exp\{y\psi - \varphi(\psi) + k(y)\} ,$$

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

(B.2) $$\mathbb{E}[y] \equiv \mu = \nabla_{\psi}\varphi(\psi)$$

(B.3) $$\text{var}[y] \equiv V(\mu) = \nabla^2_{\psi}\varphi(\psi)$$

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 (B.2) and (B.3) are satisfied. Here the natural parameterization of the LEF is used, which Morris (1982) called the natural exponential family. These are just two different parameterizations, using the mean μ or the natural parameter ψ, of the same family of densities.

An important subclass of LEF is one with quadratic variance functions, meaning the variance is a quadratic function of the mean so that $V(\mu)$ satisfies the relationship

(B.4) $$V(\mu) = v_0 + v_1\mu + v_2\mu^2 ,$$

where various possible choices of the coefficients v_0, v_1 and v_2 lead to six exponential families, five of which, the normal, Poisson, binomial, gamma, and negative binomial families constitute the Meixner class (Meixner

(1934)). Thus the restriction to QVF leaves a wide range of commonly used models.

The following results are useful in deriving the fundamental moment restrictions for the LEF-QVF class.

(i) For the LEF-QVF the orthogonal polynomial system $P_n(y,\mu)$ is defined by the Rodrigues formula (see Morris (1982))

(B.5) $$P_n(y,\mu) = V^n\{\nabla_\mu^n f(y,\psi)/f(y,\psi)\}, \qquad n = 0,1,2,\ldots$$

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

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

(B.6) $$P_{n+1} = (P_1 - n\nabla_\mu V(\mu))P_n - n(1 + (n-1)v_2)V(\mu)P_{n-1}, \quad n \geq 1.$$

(iii) Let $a_0 = 1$, and define

(B.7) $$a_n = n! \prod_{i=0}^{n-1} (1 + iv_2), \quad n \geq 1 .$$

Then

(B.8) $$\mathbb{E}_0 P_n = 0 , \qquad n \geq 1 ;$$

(B.9) $$\mathbb{E}_0 P_m P_n = \delta_{mn} a_n V^n, \qquad m,n \geq 0 ;$$

(B.10) $$\nabla_\mu^r P_n = (-1)^r (a_n/a_{n-r}) P_{n-r}, \qquad n \geq 1 , \ r = 1,\ldots,n .$$

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.

Bera, A., C. Jarque and L. F. Lee, 1984, Testing the normality assumption in limited dependent variable models, *International Economic Review* 25, 563-578.

Cameron, A.C. and P.K. Trivedi, 1990a, Regression-based tests for overdispersion in the Poisson model, *Journal of Econometrics* 46,347-364.

Cameron, A.C. and P.K. Trivedi, 1990b, Conditional moment tests with explicit alternatives, Working Paper #366, Department of Economics, U.C.-Davis.

Cameron, A.C. and P.K. Trivedi, 1990c, Conditional moment tests and orthogonal polynomials, working paper #90-51, Department of Economics, Indiana University.

Cameron, A.C. and P.K. Trivedi, 1993, Tests of independence in parametric models, with applications and illustrations, *Journal of Business and Economic Statistics,* 11, 29-43.

Cramer, H., 1946, *Mathematical methods of statistics*, Princeton: Princeton University Press.

Gallant, A. R., and D. W. Nychka, 1987, Semi-nonparametric maximum likelihood estimation, *Econometrica* 55, 2, 363-390.

Gallant, A. R., and G. Tauchen, 1989, Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications, *Econometrica* 57, 5, 1091-1120.

Gourieroux, C., A. Monfort and A.Trognon, 1984, Pseudo maximum likelihood methods: theory, *Econometrica* 52, 3, 681-700.

Gurmu, S. and P.K. Trivedi, 1992, Overdispersion tests for truncated Poisson regression models, *Journal of Econometrics*, 54, 347-370.

Gurmu, S. and P.K. Trivedi, 1993, The variable augmentation specification tests in the exponential family, *Econometric Theory*, 7, 94-114.

Hall, A., 1987, The information matrix Test in the linear model, *Review of Economic Studies* 54, 257-265.

Hall, A., 1990, Lagrange multiplier tests for normality against semi-nonparametric alternatives, *Journal of Business and economic Statistics* 417-426.

Kiefer, N., 1985, Specification diagnostics based on Laguerre alternatives for econometric models of duration, *Journal of Econometrics* 28, 135-154.

Koenker, R., 1981, A note on Studentizing a test for heteroskedasticity, *Journal of Econometrics*, 17, 107-12.

Lancaster, H. O., 1969, *The Chi-squared distribution*, New York: John Wiley.

Lancaster, T., 1985, Generalised residuals and heterogeneous duration models: with application to the Weibull model, *Journal of Econometrics* 28, 155-169.

Lee, L.-F, 1986, Specification test For Poisson regression models, *International Economic Review* 27, 3, 689-706.

Lee, L.-F and G. S. Maddala, 1985, The common structure of tests for selectivity bias, serial correlation and non-normality in the Tobit model, *International Economic Review* 26, 1-20.

McCullagh, P. and J. A. Nelder, 1983, *Generalized linear models, London*: Chapman and Hall.

Meixner, J., 1934, Orthogonale polynomsysteme mit einer besonderen gestalt der erzeugenden fuktion, *Journal of London Mathematical Society* 9, 6-13.

Morris, C. N., 1982, Natural exponential families with quadratic variance functions, *Annals of Statistics* 10, 65-80.

Newey, W. K., "A Method of moment interpretation of sequential estimators", *Economics Letters,* 14, 201-206.

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

Ord, J. K., 1972, *Families of frequency distributions,* London: Charles Griffin & Co.

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.

Robinson, P., A. Bera and C. Jarque, 1985, Tests for serial dependence in limited dependent variable models, *International Economic Review* 26, 629-638.

Smith, R. J., 1989, On the use of distributional mis-specification checks in limited dependent variables models, *Economic Journal* 99 (Conference 1989), 178-192.

Szegö, 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., 1990, A unified approach to robust, regression-based specification tests, *Econometric Theory* 6, 17-43.