INDIANA UNIVERSITY
Working Papers in Economics

CONDITIONAL MOMENT TESTS
AND ORTHOGONAL POLYNOMIALS

A. Colin Cameron
Pravin K. Trivedi

December 1990

Number 90-051

Center for Econometric Model Research
Department of Economics

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: 9 December 1990

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 moment tests in a number of cases, including the leading case of linear exponential families with quadratic variance functions. 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 and Shiferaw Gurmu for their very helpful comments on a previous draft which resulted in numerous improvements. However, we remain responsible for the current version. Trivedi also acknowledges with thanks partial research support received from the Center for Research in Econometric Modelling at Indiana University.

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 (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 and their basic properties are well known and widely used. Though many excellent treatises on this subject are available, the econometric literature on specification testing has not exploited the properties of orthogonal polynomials in designing tests. Therefore, we introduce orthogonal polynomials in section 2; important general results used in this and later sections are in Appendix A. General expressions for orthogonal polynomials are given in section 2 and in Appendix B. In section 3 we consider series expansions for distributions in terms of a base distribution and related orthogonal functions. In this context specification tests based on orthogonal polynomials are shown to be score (or LM) tests of certain moment restrictions on the base density. These general results are specialized in section 4 where the discussion is narrowed to the leading case of the linear exponential family with quadratic variance function (LEF-QVF). Illustratively, several specification tests, some well known, are derived using orthogonal polynomials. Section 5 considers extensions to models involving truncation. Section 6 concludes.

2. ORTHOGONAL POLYNOMIALS: SELECTED PROPERTIES AND RESULTS

2.1 A review of some basic 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, $a \le y \le b$. 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, μ_n, defined by

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

exist. When y is unbounded, we replace $[a,b]$ by $(-\infty,\infty)$.

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 $f(y)$ for generality and more compact notation.

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

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

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, $k_n = 1$. Later in the paper we shall refer more generally to *orthogonal or orthonormal functions*, denoted by $\xi_n(y)$, which are defined analogously.

Existence: 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 $\Delta_{ij} = \mu_{i+j-2}$, where moments may be taken either about the mean or an arbitrary origin;

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

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

The determinant in (2.4) may be partitioned as follows:

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

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:

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

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 (or monic) OPS can be generated by the following relationship:

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

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

Equation (2.7) is the solution of (2.2) with $k_n=1$, which establishes that the result is an orthonormal polynomial. The derivation of (2.7) is given in Appendix A. Also see Cramer (1946, p.132).

Variance of $P_n(\theta)$: The theoretical variance of $P_n(\theta)$ is

(2.9) $$\mathbb{E}[P_n^2(y)] = \|\Delta_n\| \,/\, \|\Delta_{n-1}\| > 0$$

$$= \mu_{2n} - d'[\Delta_{n-1}]^{-1}d \,,$$

where the last line follows from (2.6).

Other properties: In addition to uniqueness and linear independence, orthonormal polynomials satisfy other attractive properties such as minimum variance and completeness which are summarized in Appendix A. Especially useful in the construction of CM tests is the following three point recurrence relation which orthogonal polynomials satisfy:

(2.10) $$P_n(y) = (y - \alpha_n)P_{n-1}(y) - \beta_n P_{n-2}(y) ,$$

where $P_{-1}(y) = 0$, $P_0(y) = 1$, and (α_n, β_n) are defined in Appendix A.

2.2 Some special cases

The first step is the construction of the OPS for a given density. This may be done by the Gram-Schmidt orthogonalization or by the direct use of (2.7) if the moments are available. More conveniently, the first two terms in the OPS may be derived directly and the higher order terms derived using the recurrence relations for a given distribution. The OPS and the recurrence relations for many "standard" distributions of applied econometrics are widely documented; see, for example, Abramovitz and Stegun (1964).

As a simple illustrative exercise consider the orthonormal polynomial

(2.11) $$P_k(y) = y^k + a_{k1}y^{k-1} + \ldots\ldots + a_{kk}.$$

It is easy to see that $P_0(y) = 1$, $P_1(y) = y + a_{11}$, which, together with the conditions $\mathbb{E}[y] = \mu_1$ and $\mathbb{E}[P_0(y)P_1(y)] = 0$, implies that

(2.12) $$P_1(y) = y - \mu_1.$$

Since

$$\begin{aligned} P_2(y) &= y^2 + a_{21}y + a_{22} \\ &= (y - \mu_1)^2 + a_{21}(y - \mu_1) + a_{22} + 2\mu_1(y - \mu_1) + a_{21}\mu_1 + \mu_1^2 , \end{aligned}$$

orthogonality conditions $\mathbb{E}[P_2(y)] = 0$ and $\mathbb{E}[P_1(y)] = 0$ imply $a_{22} + a_{21}\mu_1 = -(\mu_2 + \mu_1^2)$. Also $\mathbb{E}[P_2(y)P_1(y)] = 0 \Rightarrow a_{21} = -\dfrac{\mu_3}{\mu_2} - 2\mu_1$, where μ_2 and μ_3 are second and third central moments. The above results imply the following second order orthogonal polynomial in $(y - \mu_1)$:

$$P_2(y) \equiv P_2(y - \mu_1) = (y - \mu_1)^2 - (\mu_3/\mu_2)(y - \mu_1) - \mu_2. \tag{2.13}$$

General expressions required to derive the orthonormal polynomials up to fourth order are given in Appendix B.

3. TESTS BASED ON ORTHOGONAL POLYNOMIALS

3.1 Score tests based on orthogonal function expansions for densities

In this paper we present CM tests based upon orthogonal polynomials. Before doing so, we discuss the more general case of orthogonal *functions* that are not necessarily *polynomials.*

Consider the case of a continuous random variable y with density function f(y) and let $\{\xi_0(y),\ldots\}$ be the corresponding set of complete orthonormal functions, not necessarily polynomials; see Appendix A for definition of completeness. 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)):

$$g(y) = f(y)\cdot\left[a_0\xi_0(y) + a_1\xi_1(y) + \ldots\right]. \tag{3.1}$$

Multiplying (3.1) by $\xi_n(y)$ and integrating term by term, and noting $\xi_0(y) = 1$,

$$a_n = \int \xi_n(y)g(y)dy, \quad a_0 = 1, \tag{3.2}$$

(3.3) $$\phi^2 = \Sigma_{n=1}^{\infty} a_n^2.$$

That is, the coefficients $\{a_n\}$ in the formal expansion are linear combinations of the moments of g(y). An analogous result holds more generally, including discrete distributions.

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

(3.4) $$H_0: a_1 = a_2 = \dots = 0.$$

From (3.1) we have

(3.5) $$\log g(y) = \log f(y) + \log[1 + \Sigma a_n \xi_n(y)]$$

(3.6) $$\nabla_{a_n} \log g(y)\Big|_{a_n=0} = \xi_n(y) \quad , \qquad i=1,2,\dots.$$

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 $\mathbb{E}_0[\nabla_{a_n} \log g(y)] = 0$, which implies that

(3.7) $$\mathbb{E}_0[\xi_n(y)] = 0.$$

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 $\xi_n(y)$, then a test of the null hypothesis may be based on the formulation $\mathbb{E}_0[\xi_n(y)] = 0$, n=1,2, ..; that is, the expectation of the orthogonal functions under the null density is zero. A comparison of (3.7) with (1.1) shows that any test based on an orthogonal function is a CM test. The analysis leading up to (3.6) shows that every specification test based on an orthonormal function is also 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. Special cases of score tests of such moment restrictions, without exploiting in detail the properties of the orthogonal functions, have sometimes appeared in the literature, e.g. Lee (1986).

3.2 Score tests based on orthogonal polynomials for densities

In this section we restrict $\xi_n(y)$ to be an nth order orthonormal polynomial - a valid restriction if $\int y^{2n} f(y)dy$ exists (Ord (1972), p.198). Then the CM specification tests will be based on orthonormal polynomials.

Two well known examples of orthogonal series expansions for densities are Gram-Charlier Type A and B (Cramer (1946), Kendall and Stuart, Vol. 1 (1969)) which take the form

$$g(y) = f(y)\cdot[1 + \Sigma_{n=1}^{\infty} a_n P_n(y)/ n!] \tag{3.8}$$

based on the normal and Poisson baseline densities. In Type A expansion f(y) is the standard normal density and $P_n(y)$ are Hermite orthogonal polynomials:

$$\begin{cases} f(y) = (2\pi)^{-1/2}\exp(- y^2/2), \quad -\infty < y < \infty \\ P_n(y) = (-1)^n \dfrac{d^n f(y)/dy^n}{f(y)} \end{cases} \tag{3.9}$$

In Type B expansion $f(y)$ is the Poisson density with parameter λ and $P_n(y)$ are Poisson-Charlier orthogonal polynomials:

$$(3.10)\qquad \begin{cases} f(y) = e^{-\lambda}\lambda^{y}/y! \quad , & y = 0, 1, .. \\ P_n(y) = \dfrac{d^n f(y)/d\lambda^n}{\partial\lambda^n} \quad . & \end{cases}$$

Note the use of the Rodrigues formula of Appendix A in (3.9) and (3.10).

Such series expansions are the basis of specification tests discussed in Section 4.

3.3 Implementation of tests based on orthogonal polynomials

The results given above were stated in terms of $P_n(y,X,\theta)$, $n \ge 0$, orthogonal polynomials in y, given X and θ. In some cases, it will be possible also to express the polynomials in terms of centered observations $(y-\mu)$ where $\mu = \mu(X, \theta)$.

If the assumed distribution implies a testable moment restriction, the test can be carried out using an orthogonal polynomial of the appropriate order. Since $\mathbb{E}[P_n(y,X,\theta) \mid X] = 0$, use of the law of iterated expectations, following Newey (1985, p.1055), suggests CM tests based on moment functions of the form

$$(3.11)\qquad \mathbb{E}[m_n(y,X,\theta) \mid X] = 0,$$

where

$$(3.12)\qquad m_n(y,X,\theta) = G_n(X,\theta)\cdot P_n(y,X,\theta),$$

and $G_n(X,\theta)$ is a matrix of functions of X and θ, and different subsets of X may appear in the functions G_n and P_n. For example, a test of omitted variables, denoted by X_2, from the conditional mean function may be based on the orthogonality condition (3.13) and (3.14) as appropriate:

$$\mathbb{E}[m_1(y,X,\theta) \mid X] = \mathbb{E}[X_2 \cdot P_1(y,X_1,\theta) \mid X] = 0, \tag{3.13}$$

$$= \mathbb{E}[X_2 \cdot P_1(y - \mu(X_1,\theta)) \mid X] = 0 \ . \tag{3.14}$$

Similarly a test of misspecified variance function may be based upon

$$\mathbb{E}[m_2(y,X,\theta) \mid X] \equiv \mathbb{E}[G_2(X,\theta) \cdot ((y - \mu_1)^2 - (\mu_3/\mu_2)(y - \mu_1) - \mu_2)] = 0, \tag{3.15}$$

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 attraction of orthonormal polynomials as the basis for CM tests is founded on their properties of uniqueness, minimum variance (in the class of polynomial functions) and linear independence. To test an nth order moment restriction we may use the nth order polynomial, and under the null hypothesis density the resulting test statistic will be independently distributed of all other tests based on higher or lower order polynomials, in the absence of unknown nuisance parameters. 'Portmanteau' or simultaneous tests of several restrictions may also be devised by linear combination of several individual restrictions. But the joint test will be additive in its components.

For a test of nth order moment restriction 2n moments must exist; see for example, equation (2.4). Implementation of the nth order test will use knowledge of the moments up to order 2n-1, see (2.7). The variance of the nth order polynomial will involve moments up to order 2n. It is clear that high order moment restriction tests are likely to be numerically unstable unless the sample is very large.

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

(3.16) $$m_{n,T}(\theta) = T^{-1} \sum_{t=1}^{T} m_{n,t}(y,X,\theta).$$

To operationalize the CM test based on OPS, θ is replaced by the estimator $\hat{\theta}_T$ consistent under the null, yielding $m_{n,T}(\hat{\theta}_T)$. If this substitution leads to $m_{n,T}(\hat{\theta}_T)$ which has the same asymptotic distribution as $m_{n,T}(\theta_0)$, then the test statistic is easily constructed. Under appropriate assumptions $T^{1/2}m_{n,T}(\hat{\theta}_T)$ has a limiting standard normal distribution; see Pagan and Vella (1989) and White (1990). For any moment condition based on the nth order orthogonal polynomial, under suitable conditions, by a first-order Taylor series expansion:

(3.17) $$T^{1/2}m_{n,T}(\hat{\theta}_T) = T^{1/2}m_{n,T}(\theta_0) + B_0 \cdot T^{1/2}(\hat{\theta}_T - \theta_0) + o_p(1),$$

where $\theta_0 = \text{plim } \hat{\theta}_T$ and $B_0 = \text{plim } \nabla_\theta m_{n,T}(\theta)$. Implementation of the CM test differs according to whether or not the following condition is satisfied:

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

From (3.12), this will be the case if the orthogonal polynomial satisfies $\mathbb{E}_0[\nabla_\theta P_n(y_t,X_t,\theta) \mid X_t] = 0$.

When the conditional moment $m_{n,t}(y_t,X_t,\theta)$ is chosen so that (3.18) is satisfied, under H_0, $B_0 = 0$, so that the asymptotic distribution of $m_{n,T}(\hat{\theta}_T)$ coincides with that of $m_{n,T}(\theta_0)$. Since $\mathbb{E}_0[m_{n,T} \mid X_t] = 0$, and $m_{n,t}$ are assumed independent over t under H_0, a central limit theorem yields

(3.19) $$T^{1/2}m_{n,T}(\hat{\theta}_T) \overset{a}{\sim} N(0, \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} m_{n,t} \cdot m_{n,t}') ,$$

so that under H_0, the test statistic

$$(3.20) \qquad \tau_n \equiv \sum_{t=1}^{T} \hat{m}'_{n,t} \cdot [\sum_{t=1}^{T} \hat{m}_{n,t} \cdot \hat{m}'_{n,t}]^{-1} \cdot \hat{m}_{n,t} ,$$

where $\hat{m}_{n,t} \equiv m_n(y_t, X_t, \hat{\theta}_T)$, is $\chi^2(\dim(m_n))$. This test statistic can be conveniently computed as T times the uncentered R^2 from the auxiliary regression of 1 on $\hat{m}_{n,t}$.

The condition (3.18) may not be satisfied when, for example, there are nuisance parameters, say α, in the p.d.f. If (3.18) does not hold the asymptotic distribution of $m_{n,T}(\hat{\theta}_T, \hat{\alpha}_T)$ can still be obtained, but will in general differ from that of $m_{n,T}(\theta_0, \alpha_0)$ and will vary with the choice of consistent estimator $\hat{\theta}_T$. Pierce (1982) and Newey (1985) have given the appropriate derivations and proofs for this later case which can be specialized to the present case as follows.

Let $(\hat{\theta}_T\ \hat{\alpha}_T)' \equiv \hat{\theta}^{*\prime}_T$ be a consistent estimator of the true vector $\theta^{*\prime}_0$. Following Pierce (1982) and Newey (1985) we assume that under the appropriate regularity conditions,

$$(3.21) \qquad \begin{bmatrix} \sqrt{T} \cdot m_n(\theta^*_0) \\ \sqrt{T} \cdot (\hat{\theta}^*_T - \theta^*_0) \end{bmatrix} \sim N \left[0, \begin{pmatrix} V_{mm} & V_{m\theta^*} \\ V_{\theta^* m} & V_{\theta^*\theta^*} \end{pmatrix} \right] .$$

Then using the results of Pierce and Newey it can be shown that

$$(3.22) \qquad T \cdot m_n(\hat{\theta}^*_T)' \cdot [\hat{V}_{mm} - \hat{V}_{m\theta^*} \hat{V}^{-1}_{\theta^*\theta^*} \hat{V}_{\theta^* m}]^{-1} \cdot m_n(\hat{\theta}^*_T) \sim \chi^2(\dim(m_n)),$$

which will be the basis of the OPS based CM test in the presence of nuisance parameters in the pdf. As emphasized by Pagan and Vella (1989, p.S33-S34), the computational implementation of this version of the test is partially simplified by the use the OPG (outer product of gradient vector) estimators for components of the asymptotic variance of $\surd T \cdot m_n(\hat{\theta})$. Specifically, consistent estimators of $\hat{V}_{m\theta*}$ and $\hat{V}_{\theta*\theta*}$ may be obtained as follows: $\hat{V}_{m\theta*} = T^{-1} \Sigma_t \hat{m}_{n,t} \cdot \hat{s}_{\theta,t}$; $\hat{V}_{\theta*\theta*} = T^{-1} \Sigma_t \hat{s}_{\theta,t} \hat{s}'_{\theta,t}$, where s_θ denotes the likelihood based score, $\partial \log L(\theta)/\partial\theta$. For theoretical arguments underlying these results see Pierce (1982, p.478), Newey (1985) or Pagan and Vella (1989). Asymptotically equivalently, Neyman's $C(\alpha)$ approach can be used to transform (orthogonalize) $m_{n,T}(\theta)$ to a sample moment $m^*_{n,T}(\theta)$ which does satisfy condition (3.18), so that we can subsequently apply the simpler theory.

4. APPLICATION TO SPECIFICATION TESTS IN THE LEF-QVF

To illustrate the use of orthogonal polynomials as the basis for the choice of moment function, we consider linear exponential families with quadratic variance functions (LEF-QVF). This covers many commonly used econometric models: regression models 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.

The LEF is defined by

$$f(y,\psi) = \exp\{y\psi - \varphi(\psi) + k(y)\} , \tag{4.1}$$

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

$$\mathbb{E}[y] \equiv \mu = \nabla_\psi \varphi(\psi) \tag{4.2}$$

$$\text{var}[y] \equiv V(\mu) = \nabla^2_\psi \varphi(\psi) \tag{4.3}$$

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 (4.2) and (4.3) 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

$$V(\mu) = v_0 + v_1\mu + v_2\mu^2 \quad , \tag{4.4}$$

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))

$$P_n(y,\mu) = V^n\{\nabla_\mu^n f(y,\psi)/f(y,\psi)\}, \qquad n = 0,1,2,\ldots \tag{4.5}$$

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

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

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

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

(4.8) $$\mathbb{E}_0 P_n = 0 , \quad n \geq 1 ;$$

(4.9) $$\mathbb{E}_0 P_m P_n = \delta_{mn} a_n V^n, \quad m,n \geq 0 ;$$

(4.10) $$\nabla_\mu^r P_n = (-1)^r (a_n/a_{n-r}) P_{n-r}, \quad n \geq 1 , \; r = 1,\ldots,n .$$

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 (4.4) which may be parameterized in terms of μ_t and the 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. We propose using (4.8) as the basis for CM specification tests of such regression models. As discussed in section 3.3, the procedure is to progressively test for $n = 1,2,..$

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

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

for some chosen function $G_n(X_t,\theta)$, where for simplicity we have suppressed the nuisance parameter. The recurrence relation (4.6) generates $P_n(y_t, \mu(X_t, \theta))$.

$P_{n,t}$ is a polynomial of degree n in y_t and μ_t, so the distributional assumptions used in performing the test are that the first 2n moments of y are correctly specified. The variance of $P_{n,t}$ for the optimal test is easily obtained using (4.8) and (4.10). Tests based on different degrees of polynomial are orthogonal by (4.9).

In comparing tests based on these orthogonal polynomials with other tests, such as score and information matrix tests, we say the tests are identical if y_t appears in the corresponding moment condition only via the function $P_n(y_t, \mu(X_t, \theta))$. 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 is identical to a CM test based on the nth order orthogonal polynomial.

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 $P_{2,t}$ coincide with a number of score tests of conditional variance; see Cameron (1990). For tests based on higher order polynomials, we consider in turn some examples in each of the LEF-QVF families.

Example 1 - *Normal Family:* For 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 (Hermite) 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 $24\sigma^8$.

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 the score tests of Bera and Jarque (1982) against the Pearson system.

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 (IM) test of White (1982) can be decomposed into three components which individually test for heteroskedasticity, skewness and non-normal kurtosis. These three components are respectively $m_{2,t} = (\text{vec}^*(X_t X_t')/\sigma^4)\cdot P_{2,t}$, where $\text{vec}^*(X_t X_t')$ denotes the unique elements of vectorization of $(X_t X_t')$; $m_{3,t} = (\text{vec}(X_t)/2\sigma^2)\cdot P_{3,t}$; and $m_{4,t} = (1/4\sigma^8)\cdot P_{4,t}$. 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 $P_{4,t}$ to those used in the IM test. Linear dependence of the orthogonal polynomials implies additivity property of the joint test.

Example 2 - Poisson Family: 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 (Poisson-Charlier) polynomials of order two and three are $\{(y-\mu)^2 - y\}$ and $\{(y-\mu)^3 - 3(y-\mu)^2 - (3\mu-2)(y-\mu) + 2\mu\}$, with variances $2\mu^2$ and $6\mu^3$, respectively.

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, whereas the Poisson imposes variance-mean equality. Tests for variance-mean equality, given correct specification of the mean, are called tests of overdispersion (or underdispersion). They 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 $P_{2,t} = (y_t - \mu_t)^2 - y_t$. Note that this leads to different CM tests than those based on the more obvious polynomial $P_{2,t} = (y_t - \mu_t)^2 - \mu_t$, the difference between $(y_t - \mu_t)^2$ and its expectation under the null hypothesis. CM tests based on $P_{2,t}$ are identical to the score tests of overdispersion in the Poisson model against alternatives that $\mathrm{Var}(y_t) - \mu_t$ equals a non-zero constant, considered in Cameron and Trivedi (1986, 1990a) and Lee (1986). The test statistic for this reduces to

$$(4.13) \qquad T_2 = \frac{\Sigma_t (y - \hat{\mu}_t)^2 - y_t}{(2\ \Sigma_t (\hat{\mu}_t^2))^{1/2}} \overset{a}{\sim} N(0,1).$$

CM tests for non-Poisson skewness may be based on the third order polynomial defined above. The test statistic is

$$(4.14) \qquad T_3 = \frac{\Sigma\ ((y_t - \hat{\mu}_t)^3 - \hat{\mu}_t) - 3((y_t - \hat{\mu}_t)^2 - y_t)}{\sqrt{2\ \Sigma\ (\hat{\mu}_t^3)}}.$$

This derivation and the test may be compared with the corresponding test procedure of Lee (1986, equation (5.12)) who derives essentially the same test as a score test by taking the null model as the Poisson and the alternative as a truncated Gram-Charlier series expansion in terms of discrete orthogonal polynomials and Poisson baseline density.

Example 3 - Exponential Family. The unit exponential is a member of the gamma family with mean and variance unity, $\nabla_{\mu}V(\mu)=0$, $v_1 = v_2 = 0$. The recurrence relation for the orthogonal polynomials is $P_{m+1} = P_1P_m - mP_{m-1}$. The second and third orthogonal (Laguerre) polynomials are $\{(y-1)^2 - 1\}$ and $\{(y-1)^3 - 3(y-1)\}$, with variances 1 and 6. (In the more general gamma family, the orthogonal polynomials are the generalized Laguerre polynomials).

The unit exponential arises in diagnostic tests for any parametric hazard model. Following convention t denotes uncensored survival times with pdf $f(t|X)$ and distribution function $F(t|X)$, survival function $S(t|X) \equiv 1 - F(t|X)$ and hazard rate $h(t|X) = f(t|X)/S(t|X) = \nabla_t \log S(t|X)$. Let $H(t|X) = \int h(s|X)ds$ denote the integrated hazard function. Diagnostic checks for any parametric survival model may be based on the generalized residuals ε_i, $i=1,..T$, defined as

$$\varepsilon_i = H(t_i|X_i). \qquad (4.15)$$

The generalized residuals are easily shown to have a unit exponential distribution, irrespective of the parametric form of $f(t|X)$, if the null model is correctly specified. Hence $\mathbb{E}_0[\varepsilon^j] = j!$, $j \geq 0$. Therefore CM diagnostic test may be based on the departure of the sample moments of ϵ from the corresponding theoretical moments of the unit exponential distribution.

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_{2t} can be used to test for zero neglected heterogeneity. This coincides with the score test given in Lancaster (1985). The CM tests of higher moments may be constructed using the results above.

Example 4 - Negative binomial family. The parameterization we choose is $f(y) = (1+\theta)^{-\alpha-y} \cdot \theta^y \cdot \binom{y+\alpha-1}{y}$. In this case $\mathbb{E}[y] = \alpha\theta = \mu$, $V(\mu) = \alpha\theta(1+\theta) = \mu + (1/\alpha)\mu^2$; $\nabla_{\mu}V(\mu) = 1 + (2/\alpha)\mu = 1 + 2\theta$. The first two orthogonal (Meixner) polynomials are $P_1(y) = y - \mu$, and $P_2(y) = (y - \mu)^2 - (1 + (\mu/\alpha))y$, so the condition (3.18) is not satisfied for tests based on $P_2(y)$.

The negative binomial is commonly used for count data which are overdispersed, so that the Poisson is inappropriate. CM tests based on $P_2(y)$ can be used to test the validity of the variance-mean relationship of the assumed negative binomial model. To the best of the authors' knowledge no score tests of variance-mean relationship in the negative binomial model have appeared in the literature.

A fifth example is the binomial family, in which case CM tests will be based on orthogonal (Krawtchouk) polynomials. Results are straight-forward.

Implementation of the above tests for LEF-QVF examples is usually straight-forward since by (4.10) and (4.8), $\mathbb{E}[\nabla_\mu P_n] = \mathbb{E}[(-a_n/a_{n-1})P_{n-1}] = 0$, so (3.18) holds. To the extent that no nuisance parameters are present, i.e. the only unknown parameters in LEF-QVF density for y_t appear via the mean function μ_t, we can directly compute the simpler (3.20). It can be shown that this is also the case for the normal density, the nuisance parameter σ^2 does not prevent the use of (3.20). Only in example 4 (the negative binomial with nuisance parameter α) do CM tests need to be implemented by the more general procedure discussed at the end of section 3.

Thus to implement all of the CM tests given in this section aside from example 4, but additionally including tests for the binomial with known number of trials, 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 (3.18) holds. 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 (3.18) 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 (1990a). 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. Reconsider the series expansion (3.9) which is rewritten as follows:

$$(5.1) \qquad g(y) = f^*(y)\cdot[1 + \Sigma_{n=1}^{\infty} a_n^* P_n^*(y)/n!] ,$$

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

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 $f^*(y)$. Specification tests of the truncated distribution based on

$P^*_n(y)$ are tests of the null H_0: $a^*_n = 0$. 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 (1990a) 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 may not be the same as in another conventionally designed score test.

As an illustration we reconsider the example of left truncated Poisson distribution analyzed in Gurmu and Trivedi (1990a). 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) $$f(y_t, \psi_t | y_t \geq 1) = \frac{\psi_t^{y_t}}{(\exp(\psi_t) - 1)\cdot y_t!}.$$

The first three (truncated) moments of the *positive Poisson* are as follows:

(5.4) $$\mu_{1t} = \psi_t + \delta_t$$

(5.5) $$\mu_{2t} = \psi_t - \delta_t(\mu_{1t} - 1)$$

(5.6) $$\mu_{3t} = \psi_t + 2\mu_{1t}(\mu_{1t}^2 - \psi_t^2) + \delta_t(3\psi_t + 1)$$

(5.7) $$\delta_t = \psi_t\cdot(\exp(\psi_t) + 1).$$

We may construct an OPS based CM test of the second moment restriction using (2.13), which yields

(5.8) $$P_2(y_t, X_t, \theta) = \varepsilon_t^2 - (\mu_{3t}/\mu_{2t})\varepsilon_t - \mu_{2t},$$

where $\varepsilon_t = (y_t - \mu_{1t})$. By contrast the score function given in Gurmu and Trivedi (1990a) is the sum of terms that are a multiple (not depending on y_t) of the polynomial

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

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, in quite general settings.

Even simpler formulae are presented 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 coincided with existing tests, these tests were score tests. By contrast, in the example of truncated models, CM tests based on orthogonal polynomials differed from existing score tests.

OPS based CM tests are score tests designed to be orthogonal in a specific sense. The conventional score approach in which one examines the departure from the null in one direction at a time does 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 over the standard score or CM approach in such cases. Further insights may be gained by additional work on the properties and performance of OPS based tests in settings more general than those considered here.

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.

Derivation of the result (2.7): First partition $\| D_n(y) \|$ as follows:

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

The partitioned bordered determinant theorem yields

(A.2)
$$\| D_n(y) \| = y^n \| \Delta_{n-1} \| - c'(y) \, Adj \, (\Delta_{n-1}) \, d$$

(A.3)
$$\frac{\| D_n(y) \|}{\| \Delta_{n-1} \|} = y^n - c'(y) \, [\Delta_{n-1}]^{-1} \, d$$

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

Uniqueness: The OPS $\{P_n(y)\}$ in which the leading coefficient is normalized to unity, i.e., the orthonormal (or monic) polynomial sequence, is unique. If $\{Q_n(y)\}$ is also an OPS, then there exist constants $c_n \neq 0$ such that $Q_n(y) = c_n P_n(y)$, $n=0,1,2,\ldots$

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

Covariance properties: For an OPS $\{P_n(y)\}$ and for every polynomial $\pi_m(y)$, $m \le n$,

(A.4i) $\mathbb{E}[\pi_m(y)P_n(y)] = 0$ for $m < n$

(A.4ii) $\mathbb{E}[\pi_m(y)P_n(y)] \neq 0$ for $m = n$

(A.4iii) $\mathbb{E}[y^m P_n(y)] = k_n \delta_{mn}$, $k_n \neq 0$, for $m \le n$.

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

(A.5) $$\mathbb{E}[\pi_n(y)P_n(y)] = \|\Delta_n\| \,/\, \|\Delta_{n-1}\| \;, \; \|\Delta_{-1}\| = 1.$$

For non-orthonormal polynomials $\pi^*_n(y)$ with leading coefficient c_n, and $P^*_n(y)$ with leading coefficient a_n,

(A.6) $$\mathbb{E}[\pi^*_n(y)P^*_n(y)] = a_n c_n \|\Delta_n\| \,/\, \|\Delta_{n-1}\|, \quad \|\Delta_{-1}\| = 1.$$

Minimum variance property: If $\|\Delta_n\| > 0$ $(n \ge 0)$, then the orthonormal polynomial $P_n(y)$ satisfies the following property for every non-orthonormal orthogonal polynomial $\pi_n(y) \neq P_n(y)$:

(A.7) $$\mathbb{E}[P^2_n(y)] < \mathbb{E}[\pi^2_n(y)].$$

Recurrence relationship: Given the positive definite sequence $\{\mu_n\}$ and an orthonormal $\{P_n(y)\}$, there exist real constants α_n and β_n, $\beta_n > 0$, such that the polynomials satisfy the following three point recurrence relationship:

(A.8) $$P_n(y) = (y - \alpha_n)P_{n-1}(y) - \beta_n P_{n-2}(y)$$

where $P_{-1}(y) = 0$, $P_0(y) = 1$, and

(A.8a) $$\beta_{n+1} = \mathbb{E}[P_n^2(y)]/\ \mathbb{E}[P_{n-1}^2(y)] = \frac{\|\Delta_{n-2}\|\ \|\Delta_n\|}{\|\Delta_{n-1}\|^2}$$

(A.8b) $$\mathbb{E}[P_n^2(y)] = \beta_1\beta_2\cdots\beta_{n+1}$$

(A.8c) $$\alpha_n = \mathbb{E}[y\cdot P_{n-1}^2(y)]/\ \mathbb{E}[P_{n-1}^2(y)]\ .$$

Note that if $\{\mu_n\}$ is symmetric then $P_n(-y) = (-1)^n P_n(y)$, and $\alpha_n = 0$, $n \geq 1$.

Rodrigues formula: In some cases, including the leading case of the classical orthogonal polynomials, the $\{P_j(y)\}$ may be conveniently generated by the Rodrigues formula

(A.9) $$P_j(y) = \frac{1}{k_j f(y)}\ \nabla^j\{f(y)\cdot\pi(y)^j\},$$

for some constants k_j where $\pi(y)$ is a polynomial in y, independent of j, and ∇ is the operator $\partial/\partial y$.

Appendix B

Orthonormal polynomials may be derived, using (2.7), in terms of moments around the origin. Below we give expressions, obtained using the computer algebra program MACSYMA Version 412.62, which can be used to derive the first four orthogonal polynomials.

(B.1) $\|\Delta_0\| = 1$

(B.2) $\|\Delta_1\| = \mu_2 - \mu_1^2$

(B.3) $\|\Delta_2\| = \mu_2\mu_4 - \mu_3^2 - \mu_2^3 - \mu_4\mu_1^2 + 2\mu_3\mu_2\mu_1$

(B.4) $\|\Delta_3\| = (\mu_2\mu_4 - \mu_3^2 - \mu_2^3)\cdot\mu_6 - \mu_2\mu_5^2 + 2(\mu_3\mu_4 + \mu_2^2\mu_3)\mu_5 - \mu_4^3 + \mu_2^2\mu_4^2$

$- 3\mu_2\mu_3^2\mu_4 + \mu_3^4 + (\mu_5^2 - \mu_4\mu_6)\mu_1^2 + 2(\mu_2\mu_3\mu_6 - (\mu_2\mu_4 + \mu_3^2)\mu_5 + \mu_3\mu_4^2)\mu_1$

(B.5) $\|D_0(y)\| = 1$

(B.6) $\|D_1(y)\| = y - \mu_1$

(B.7) $\|D_2(y)\| = \|\Delta_1\| y^2 + (\mu_2\mu_1 - \mu_3)y + (\mu_3\mu_1 - \mu_2^2)$

(B.8) $\|D_3(y)\| = \|\Delta_2\|\cdot y^3 + \left[(\mu_5\mu_1^2 - (\mu_2\mu_4 + \mu_3^2)\mu_1 - (\mu_2\mu_5 - \mu_3\mu_4 - \mu_2^2\mu_3)\right]y^2$

$+ \left[(\mu_3\mu_4 - \mu_2\mu_5)\mu_1 + \mu_3\mu_5 - \mu_4^2 + \mu_2^2\mu_4 - \mu_2\mu_3^2)\right]y + \left[(\mu_4^2 - \mu_3\mu_5)\mu_1 + \mu_2^2\mu_5\right.$

$\left.- 2\mu_2\mu_3\mu_4 + \mu_3^3\right]$

$$(B.9) \quad \|D_4(y)\| = -\|\Delta_3\| \cdot y^4 + \Big[(-\mu_2\mu_4 + \mu_3^2 + \mu_2^3)\mu_7 + (-\mu_2\mu_5 + \mu_3\mu_4 + \mu_2^2\mu_3)\cdot\mu_6$$

$$+ \mu_3\mu_5^2 + (-\mu_4^2 + 2\mu_2^2\mu_4 - \mu_2\mu_3^2)\cdot\mu_5 - 2\mu_2\mu_3\mu_4^2 + \mu_3^3\mu_4 + (\mu_4\mu_7 - \mu_5\mu_6)\mu_1^2$$

$$+ (-2\mu_2\mu_3\mu_7 + (\mu_2 m_4 + \mu_3^2)\mu_6 + \mu_2\mu_5^2 - \mu_4^3)\mu_1\Big]y^3$$

$$+ \Big[(-\mu_2\mu_5 + \mu_3\mu_4 + \mu_2^2\mu_3)\mu_7 + \mu_2\mu_6^2 - (\mu_3\mu_5 - \mu_4^2 + \mu_2^2\mu_4 - \mu_2\mu_3^2)\cdot\mu_6$$

$$+ (\mu_6^2 + \mu_5\mu_7)\mu_1^2 + ((\mu_2\mu_4 + \mu_3^2)\mu_7 - (\mu_2\mu_5 + 3\mu_3\mu_4)\mu_6 + \mu_3\mu_5^2 + \mu_4^2\mu_5)\mu_1$$

$$+ \mu_4\mu_5^2 + \mu_3^3\mu_5 \quad \mu_2\mu_4^3 - \mu_3^2\mu_4^2\Big]y^2 + \Big[(\mu_2\mu_5 - \mu_3\mu_4)\mu_7 - \mu_2\mu_6^2 + (\mu_3\mu_5$$

$$+ \mu_4^2)\mu_6 - \mu_4\mu_5^2)\mu_1 + (\mu_3\mu_5 - \mu_4^2 + \mu_2^2\mu_4 - \mu_2\mu_3^2)\cdot\mu_7$$

$$- \mu_3\mu_6^2 + ((2\mu_4 - \mu_2^2)\mu_5 + \mu_3^3)\mu_6 - \mu_5^3 + 2\mu_2\mu_3\mu_5^2 + (-\mu_2\mu_4^2 - 2\mu_3^2\mu_4)\mu_5$$

$$+ \mu_3\mu_4^3\Big]y + \Big[((\mu_2\mu_5 - \mu_4)\mu_7 - \mu_3\mu_6 + 2\mu_4\mu_5\mu_6 - \mu_5)\mu_1 +$$

$$+ (\mu_2^2\mu_5 - 2\mu_2\mu_3\mu_4 + \mu_3^2)\mu_7 - \mu_2^2\mu_6^2 + (2(\mu_2\mu_4\mu_5 + \mu_2\mu_4^2 - \mu_3^2\mu_4)\mu_6\Big]$$

REFERENCES

Abramowitz, M. and I. A. Stegun, 1964, *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.

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., 1990, "Regression-Based Tests of Variance-Mean Relationships", Working Paper, Department of Economics, U.C.-Davis.

Cameron, A.C., and P.K., Trivedi, 1986, "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators", *Journal of Applied Econometrics,* 1, 29-53.

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.

Cramer, H., 1946, *Mathematical Methods of Statistics*, Princeton: Princeton 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, 1990a, "Overdispersion Tests for Truncated Poisson Regression Models", Indiana University Working Papers in Economics, 90-007.

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

Kendall, M. G., and A. Stuart, 1969, *The Advanced Theory of Statistics*, Vol.I, London: Charles Griffin & Co.

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

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., 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.

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., 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.