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 $\{(y_t, X_t), t = 1,\ldots,T\}$ 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:

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

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:

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

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

Similarly a test of misspecified variance function may be based upon

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

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 $\{P_0(y), \dots\}$ 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)):

$$g(y) = f(y) \cdot \left[a_0 P_0(y) + a_1 P_1(y) + \dots \right] . \tag{3.6}$$

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) $$H_0: a_1 = a_2 = \ldots. = 0.$$

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:

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

The obvious CM test is based on:

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

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:

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

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:

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

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:

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

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

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

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 $((q+p)\times 1)$ 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

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

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

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

$$A_{12} = 0$$

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

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

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

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

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

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

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: $s(\cdot)$ is the likelihood score function and first order orthogonal polynomial; $\hat{\theta}$ is MLE; $R(\cdot)$ 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) $$H_0\colon\ \mathbb{E}_0[m_n(y_t,X_t,\theta) \mid X_t] = 0 \quad ,$$

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