Econometric Theory, **28**, 2012, 328–362.
doi:10.1017/S0266466611000168

INTEGRATED CONDITIONAL MOMENT TESTS FOR PARAMETRIC CONDITIONAL DISTRIBUTIONS

HERMAN J. BIERENS
Pennsylvania State University

LI WANG
Penn State College of Medicine

In this paper we propose consistent integrated conditional moment tests for the validity of parametric conditional distribution models, based on the integrated squared difference between the empirical characteristic function of the actual data and the characteristic function implied by the model. To avoid numerical evaluation of the conditional characteristic function of the model distribution, a simulated integrated conditional moment test is proposed. As an empirical application we test the validity of a few common health economic count data models.

1. INTRODUCTION

In this paper we address the problem of testing the validity of parametric conditional distribution specifications for cross-section data. Our approach is based on the well-known fact that two distribution functions are the same if and only if their characteristic functions are the same. Therefore, similar to the integrated conditional moment (ICM) tests proposed by Bierens (1982, 1990) and Bierens and Ploberger (1997) for regression models, we propose consistent tests for the correctness of parametric conditional distribution function specifications based on the integrated squared difference of the empirical characteristic function of the data and the empirical characteristic function corresponding to the estimated conditional distribution function involved.

There exists a substantial body of literature on consistent testing of the functional form of conditional expectation models. See, e.g., Bierens (1982, 1990),

A previous version of this paper was presented by Herman Bierens at the Symposium on Nonparametric Testing in Econometrics, 2008, Indiana University, and the European Econometric Society Meeting 2009, Barcelona, and by Li Wang at the Health Econometrics Workshop 2008, University of Milan-Bicocca. The helpful comments of Juan Carlos Escanciano, Matteo Manera, Michael Jansson, and a referee are gratefully acknowledged. Support for research within the Center for the Study of Auctions, Procurements, and Competition Policy (CAPCP) at Penn State has been provided by a gift from the Human Capital Foundation (http://www.hcfoundation.ru/en/). Address correspondence to Herman J. Bierens, Department of Economics and CAPCP, Pennsylvania State University, 510 Kern Graduate Building, University Park, PA 16802, USA; e-mail: hbierens@psu.edu.

Bierens and Ploberger (1997), Stute (1997), Stinchcombe and White (1998), Boning and Sowell (1999), Fan and Li (2000), and Escanciano (2006) for ICM and related tests and Härdle and Mammen (1993), Gozalo (1993), Horowitz and Härdle (1994), Hong and White (1995), Li and Wang (1998), Zheng (1996), and Lavergne and Vuong (2000), among others, for tests based on comparisons of parametric functional forms with corresponding nonparametric or semi-nonparametric estimates. However, the literature on *consistent* specification testing of conditional distribution models is very limited;[1] to the best of our knowledge it consists only of two papers, Andrews (1997) and Zheng (2000).

Andrews (1997) generalized the Kolmogorov test for testing unconditional distribution to a conditional Kolmogorov (CK) test for testing general conditional distributions. This test is in the same spirit as the ICM test we will propose in the next section, in that Andrews (1997) compared the empirical distribution function of a pair (Y, X) with the corresponding empirical distribution function implied by the model, whereas our ICM test is based on the comparison of the corresponding empirical characteristic functions. In particular, the CK test statistic takes the form

$$\max_{1\le i\le n}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\mathbf{1}(Y_j \le Y_i) - F(Y_i|X_j,\widehat{\theta})\right)\mathbf{1}(X_j \le X_i)\right|, \qquad \textbf{(1)}$$

where $F(y|X_j,\widehat{\theta})$ is the estimated conditional distribution model and $\mathbf{1}(\cdot)$ is the indicator function. Because the asymptotic null distribution is case-dependent, the critical values have to be derived via a bootstrap method. This test is consistent and has nontrivial power against $\sqrt{n}$ local alternatives. However, a practical problem with the CK test is that if the dimension of X_j is large the inequality $X_j < X_i$ for $i \neq j$ may never happen, even in large samples. This appears to be the case in our empirical application in Section 4. This problem does not happen with the proposed ICM test.

Zheng (2000) proposed a test for the validity of conditional densities by comparing a parametric conditional density with a corresponding nonparametric kernel estimator via an approximation of the Kullback–Leibler (Kullback and Leibler, 1951) information criterion. Thus, this test is only applicable to absolutely continuous conditional distribution models. Zheng's test has nontrivial local power but only against local alternatives that approach the null at a slower rate than $\sqrt{n}$. Moreover, this rate decreases with the number of covariates, and so Zheng's test suffers from the curse of dimensionality.

This paper is organized as follows. In Section 2 we introduce our ICM test for conditional distributions, and in Section 3 we propose a simulated ICM (SICM) test to avoid the computation of conditional characteristic functions by numerical integration. In Section 4 we apply the SICM test to a conditional Poisson model and negative binomial logit models for health economic count data. In Section 5 we make some concluding remarks. Most of the proofs are given in the Appendix at the end of this paper.

Throughout the paper we will use the following notation. Convergence in distribution will be denoted by $\xrightarrow{d}$ and convergence in probability by $\xrightarrow{p}$ or $p\lim_{n\to\infty}$. The symbol $\sim$ stands for "is distributed as" and $\Rightarrow$ indicates weak convergence of random functions. See, e.g., Billingsley (1968) for the latter notion. As said before, the indicator function will be denoted by $\mathbf{1}(\cdot)$. Finally, a bar over a complex variable, vector, or function denotes the complex conjugate; i.e., if $z = a + \mathbf{i}.b$ then $\overline{z} = a - \mathbf{i}.b$, where here and in what follows $\mathbf{i} = \sqrt{-1}$.

2. THE ICM TEST FOR CONDITIONAL DISTRIBUTIONS

We will develop our test for cross-section models only. Because the parametric model takes the form of a conditional distribution function specification $F(y|X;\theta)$, we will assume that the parameter vector θ involved is estimated by maximum likelihood (ML). Of course, if the model is misspecified then ML becomes quasi maximum likelihood (QML).

2.1. Quasi Maximum Likelihood Conditions

Throughout we will assume that the standard regularity conditions for the convergence in probability and asymptotic normality of the QML estimator of θ hold. See White (1982, 1994).

Assumption 1. We observe a random sample $(Y_1, X_1), \ldots, (Y_n, X_n)$ from $(Y, X) \in \mathbb{R}^m \times \mathbb{R}^k$. The conditional distribution function of Y given X is assumed to belong to a given parametric family $F(y|X;\theta)$, $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^p$ is a compact and convex parameter space. The support of $F(y|X;\theta)$ does not depend on θ. The log-likelihood involved takes the form $\ln L_n(\theta) = \sum_{j=1}^n \ell(Y_j, X_j;\theta)$ where $\ell(Y, X;\theta)$ is almost surely (a.s.) twice continuously differentiable in θ. The QML estimator $\widehat{\theta} = \arg\max_{\theta\in\Theta} \ln L_n(\theta)$ converges in probability to $\theta_0 = \arg\max_{\theta\in\Theta} \mathrm{E}[\ell(Y, X;\theta)]$, which is a unique interior point of Θ. Moreover, using the notation[2]

$$\Delta\ell(Y, X;\theta) = \partial\ell(Y, X;\theta)/\partial\theta', \qquad \Delta^2\ell(Y, X;\theta) = \frac{\partial^2\ell(Y, X;\theta)}{\partial\theta\partial\theta'},$$

we have that $\mathrm{E}[\Delta\ell(Y, X;\theta_0)] = 0$ and the matrix $A = \mathrm{E}\left[-\Delta^2\ell(Y, X;\theta_0)\right]$ is positive definite. Furthermore,

$$\sqrt{n}(\widehat{\theta} - \theta_0) = A^{-1}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n \Delta\ell\left(Y_j, X_j;\theta_0\right)\right) + o_p(1),$$

so that by the central limit theorem, $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N_p\left[0, A^{-1}BA^{-1}\right]$, where $B = \mathrm{Var}(\Delta\ell(Y, X;\theta_0))$.

Note that nothing is said about whether the parametric specification $F(y|X;\theta_0)$ is correct or not because all the expectations involved are unconditional. In particular, the condition $\mathrm{E}[\Delta\ell(Y,X;\theta_0)]=0$ follows from the first-order condition of a maximum of $\mathrm{E}[\ell(Y,X;\theta)]$ in $\theta=\theta_0$ together with the implicit standard condition that by the dominated convergence theorem, $\mathrm{E}\left[\partial\ell(Y,X;\theta)/\partial\theta'\right]=\partial\mathrm{E}[\ell(Y,X;\theta)]/\partial\theta'$. Moreover, the assumption that $\widehat{\theta}$ is the QML estimator is not essential as long as $\sqrt{n}(\widehat{\theta}-\theta_0)=A^{-1}\left(n^{-1/2}\sum_{j=1}^{n}V_j\right)+o_p(1)$ for some nonsingular matrix A and independent random vectors V_j with zero expectation and finite variance matrix.

2.2. Model Verification via Characteristic Functions

The null hypothesis to be tested is that the conditional distribution specification $F(y|X;\theta)$ is correct; i.e.,

$$H_0:\Pr[Y\le y|X]=F(y|X;\theta_0)\quad\text{a.s. for all } y\in\mathbb{R}^m,$$

where θ_0 is the probability limit of the QML estimator $\widehat{\theta}$, and the alternative hypothesis is that H_0 is incorrect:

$$H_1:\sup_{y\in\mathbb{R}^m}|\Pr[Y\le y|X]-F(y|X;\theta)|>0\quad\text{a.s. for all } \theta\in\Theta.$$

The proposed ICM test is based on the comparison of the actual conditional characteristic function $\mathrm{E}\left[\exp(\mathbf{i}.\tau'Y)|X\right]$ with the conditional characteristic function $\int\exp(\mathbf{i}.\tau'y)dF(y|X,\theta_0)$ implied by the model. As is well known, H_0 is true if and only if

$$\Pr\left(\mathrm{E}\left[\exp(\mathbf{i}.\tau'Y)|X\right]=\int\exp(\mathbf{i}.\tau'y)dF(y|X,\theta_0)\right)=1$$

for all $\tau\in\mathbb{R}^m$. In its turn this is true if and only if

$$\mathrm{E}\left[\exp(\mathbf{i}.\tau'Y)\exp(\mathbf{i}.\xi'X)\right]=\mathrm{E}\left[\int\exp(\mathbf{i}.\tau'y)dF(y|X,\theta_0)\exp(\mathbf{i}.\xi'X)\right]$$

for all $\tau\in\mathbb{R}^m$ and $\xi\in\mathbb{R}^k$. Thus under H_1,

$$\mathrm{E}\left[\exp(\mathbf{i}.\tau'Y)\exp(\mathbf{i}.\xi'X)\right]\ne\mathrm{E}\left[\int\exp(\mathbf{i}.\tau'y)dF(y|X,\theta_0)\exp(\mathbf{i}.\xi'X)\right]$$

for some points $(\tau,\xi)\in\mathbb{R}^m\times\mathbb{R}^k$.

Without loss of generality we may assume that Y and X are bounded random vectors, because if not one may replace Y and X by bounded one-to-one transformations $\Phi_1(Y)$ and $\Phi_2(X)$, respectively. These bounded transformation only lead to a few minor changes in the notation. See the remark following Theorem 1 in Section 2.3.

As is well known, characteristic functions of bounded random vectors are completely determined by their shape in an arbitrary open neighborhood of the zero vector. Therefore, denoting

$$\overline{T}(\theta) = \int_{\Upsilon\times\Xi} |\varsigma(\tau,\xi;\theta)|^2 d\mu(\tau,\xi), \tag{2}$$

where

$$\varsigma(\tau,\xi|\theta) = \mathrm{E}\left[\left(\exp\left(\mathbf{i}.\tau'Y\right) - \int \exp\left(\mathbf{i}.\tau'y\right) dF(y|X,\theta)\right)\exp\left(\mathbf{i}.\xi'X\right)\right],$$

$$\Upsilon = \times_{j=1}^{m}\left[-\overline{\tau}_j,\overline{\tau}_j\right], \qquad \overline{\tau}_j > 0, \tag{3}$$

$$\Xi = \times_{j=1}^{k}\left[-\overline{\xi}_j,\overline{\xi}_j\right], \qquad \overline{\xi}_j > 0, \tag{4}$$

and μ is the uniform distribution function on $\Upsilon\times\Xi$, i.e.,

$$d\mu(\tau,\xi) = \frac{d\tau d\xi}{2^{k+m}\prod_{j=1}^{m}\overline{\tau}_j\prod_{j=1}^{k}\overline{\xi}_j},$$

it follows from the continuity of characteristic functions that under H_1, $\overline{T}(\theta) > 0$ for all $\theta \in \Theta$, whereas of course under H_0, $\overline{T}(\theta_0) = 0$. This suggests that similar to Bierens and Ploberger (1997) the null hypothesis can be tested consistently by an ICM test of the form

$$\widehat{T}_n = \int_{\Upsilon\times\Xi} |Z_n(\tau,\xi)|^2 d\mu(\tau,\xi), \tag{5}$$

where

$$Z_n(\tau,\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(\mathbf{i}.\tau'Y_j\right) - \int \exp\left(\mathbf{i}.\tau'y\right) dF(y|X_j,\hat{\theta})\right)\exp\left(\mathbf{i}.\xi'X_j\right). \tag{6}$$

Note that the choice of the uniform distribution $\mu(\tau,\xi)$ as weight function in (5) is a matter of convenience rather than necessity. Our asymptotic results carry over if we replace $d\mu(\tau,\xi)$ by $w(\tau,\xi)d\tau d\xi$, where $w(\tau,\xi)$ is a continuous density with support $\Upsilon\times\Xi$. However, Boning and Sowell (1999) have shown for the ICM test of Bierens and Ploberger (1997) that the uniform weight function is optimal (in some sense), and their argument will likely carry over to our case, although we have not verified that. Also, to some extent it is possible to choose the integration domain $\Upsilon\times\Xi$ optimally. The latter issue will be addressed in Section 2.7.

2.3. Asymptotic Properties

To derive the asymptotic properties of the ICM statistic (5) we need to separate the sample variation in $Z_n(\tau,\xi)$ from the estimation error of the QML estimator $\hat{\theta}$. For this we need the following conditions.

Assumption 2. The conditional characteristic function of $F(y|X,\theta)$,

$$\varphi(\tau|X;\theta) = \int \exp\left(\mathbf{i}.\tau' y\right) dF(y|X,\theta), \tag{7}$$

is a.s. continuously differentiable in $\theta \in \Theta_0$, where Θ_0 is a closed and convex subset of Θ containing θ_0 in its interior, with column vector of partial derivatives $\Delta\varphi(\tau|X;\theta) = \partial\varphi(\tau|X;\theta)/\partial\theta'$ satisfying $\mathrm{E}[\sup_{(\tau,\theta)\in\Upsilon\times\Theta_0} \|\Delta\varphi(\tau|X;\theta)\|] < \infty$.

Then we can state the following result.

LEMMA 1. *Under Assumptions 1 and 2,*

$$\sup_{(\tau,\xi)\in\Upsilon\times\Xi} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\varphi(\tau|X_j;\widehat{\theta}) - \varphi(\tau|X_j;\theta_0)\right) \exp\left(\mathbf{i}.\xi' X_j\right) - b(\tau,\xi)' A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Delta\ell(Y_j, X_j;\theta_0) \right| = o_p(1), \tag{8}$$

where $b(\tau,\xi) = \mathrm{E}[\Delta\varphi(\tau|X;\theta_0)\exp(\mathbf{i}.\xi' X)]$. *Consequently, denoting*

$$\widetilde{Z}_n(\tau,\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \phi(\tau,\xi|Y_j,X_j), \tag{9}$$

where

$$\phi(\tau,\xi|Y,X) = \left(\exp\left(\mathbf{i}.\tau' Y\right) - \varphi(\tau|X;\theta_0)\right) \exp\left(\mathbf{i}.\xi' X\right) + b(\tau,\xi)' A^{-1} \Delta\ell(Y,X;\theta_0) \tag{10}$$

it follows that

$$\sup_{(\tau,\xi)\in\Upsilon\times\Xi} |\widetilde{Z}_n(\tau,\xi) - Z_n(\tau,\xi)| = o_p(1), \tag{11}$$

regardless of whether the null hypothesis is true or not.

Proof. See the Appendix. ■

We are now able to state the first main result.

THEOREM 1. *Let Y and X be bounded random vectors. Under Assumptions 1 and 2 and H_0,*

$$Z_n \Rightarrow Z \quad \text{on } \Upsilon \times \Xi, \tag{12}$$

where Z is a zero-mean complex-valued Gaussian process with covariance function

$$\Gamma((\tau_1,\xi_1),(\tau_2,\xi_2)) = \mathrm{E}\left[\phi(\tau_1,\xi_1|Y,X)\overline{\phi(\tau_2,\xi_2|Y,X)}\right]. \tag{13}$$

Then by the continuous mapping theorem,

$$\widehat{T}_n \xrightarrow{d} T = \int_{\Upsilon\times\Xi} |Z(\tau,\xi)|^2 d\mu(\tau,\xi),$$

whereas under H_1,

$$\widehat{T}_n/n \xrightarrow{p} \overline{T}(\theta_0) > 0, \tag{14}$$

where $\overline{T}(\theta_0)$ *is defined in (2).*

Proof. See the Appendix. ■

Remark. Recall that it is not essential that Y and X are bounded random vectors because we may without loss of generality replace Y and X by bounded one-to-one mappings $\Phi_1(Y)$ and $\Phi_2(X)$, respectively, and redefine $\varphi(\tau|X;\theta)$, $b(\tau,\xi)$, and $\phi(\tau,\xi|Y,X)$ as

$$\varphi(\tau|X;\theta) = \int \exp\left(\mathbf{i}.\tau'\Phi_1(y)\right) dF(y|X,\theta), \tag{15}$$

$$b(\tau,\xi) = \mathrm{E}\left[\Delta\varphi(\tau|X;\theta_0)\exp\left(\mathbf{i}.\xi'\Phi_2(X)\right)\right], \tag{16}$$

$$\begin{aligned}\phi(\tau,\xi|Y,X) &= \left(\exp\left(\mathbf{i}.\tau'\Phi_1(Y)\right) - \varphi(\tau|X;\theta_0)\right)\exp\left(\mathbf{i}.\xi'\Phi_2(X)\right),\\ &\quad - b(\tau,\xi)'A^{-1}\sum_{j=1}^{n}\Delta\ell(Y|X;\theta_0),\end{aligned} \tag{17}$$

respectively. However, there are two practical problems involved. The first one is that it may be difficult to compute the conditional characteristic function $\int \exp\left(\mathbf{i}.\tau'\Phi_1(y)\right) dF(y|X,\theta)$, even if the original conditional characteristic function $\int \exp\left(\mathbf{i}.\tau'y\right) dF(y|X,\theta)$ has a closed form. We will solve that problem in Section 3. The second problem is how to choose Φ_1 and Φ_2 such that enough variation in $\Phi_1(Y)$ and $\Phi_2(X)$ is preserved. How to solve this problem will be addressed in Section 2.4.

2.4. Standardization

Consider the case where Y is the average dollar amount that a household spends on food per month and suppose that we have chosen $\Phi_1(y) = \arctan(y)$. Assuming that all the households in the sample spend at least 100 dollars per month on food we then have $\pi/2 - 0.01 \leq \arctan(Y) < \pi/2$ a.s. Clearly, in this case our ICM test will have no finite-sample power. Therefore, it is important to standardize Y before taking any bounded transformation. The same applies to the components of X.

In particular, let $V_{i,j}$ be component i of $V = (Y', X')'$ and let $S_{n,i,j} = \sigma_{n,i}^{-1}\left(V_{i,j} - \mu_{n,i}\right)$, where $\mu_{n,i}$ and $\sigma_{n,i} > 0$ are location and scale parameters, and choose the arctan function as the bounded transformation. For example, choose for $\mu_{n,i}$ the sample mean and for $\sigma_{n,i}$ the sample standard error of the $V_{i,j}$'s. Alternatively, choose $\mu_{n,i}$ and $\sigma_{n,i}$ such that most of the values of $S_{n,i,j}$ fall in the interval $[-1,1]$, because in this interval the arctan function still has substantial variation. For example, let $\mu_{n,i} = 0.5\left(Q_{n,i}(0.95) + Q_{n,i}(0.05)\right)$ and $\sigma_{n,i} = 0.5\left(Q_{n,i}(0.95) - Q_{n,i}(0.05)\right)$, where $Q_{n,i}(\alpha)$ is the $\alpha \times 100\%$ sample quantile of the $V_{i,j}$'s. Then $1/n\sum_{j=1}^{n}\mathbf{1}(|S_{n,i,j}| \leq 1) \approx 0.9$.

The question now arises whether this standardization affects the asymptotic properties of our ICM test. The answer is no, provided that the following conditions hold.

Assumption 3. For $i = 1,2,\ldots,m+k$, there exist constants μ_i and $\sigma_i > 0$ such that $\sqrt{n}(\mu_{n,i} - \mu_i) = O_p(1)$ and $\sqrt{n}(\sigma_{n,i} - \sigma_i) = O_p(1)$.

These conditions hold for sample means and sample standard errors provided that the variables $V_{i,j}$ have finite fourth moments, and they hold under mild conditions for quantiles also.

Denote

$$\left(\widehat{\Phi}_1(Y_j)', \widehat{\Phi}_2(X_j)'\right) = \left(\arctan\left(S_{n,1,j}\right), \ldots, \arctan\left(S_{n,m+k,j}\right)\right),$$

$$\left(\Phi_1(Y_j)', \Phi_2(X_j)'\right) = \left(\arctan\left(S_{1,j}\right), \ldots, \arctan\left(S_{m+k,j}\right)\right),$$

where $S_{i,j} = \sigma_i^{-1}\left(V_{i,j} - \mu_i\right)$. Redefine $Z_n(\tau,\xi)$ as

$$Z_n(\tau,\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(\mathbf{i}.\tau'\Phi_1(Y_j)\right) - \int \exp\left(\mathbf{i}.\tau'\Phi_1(y)\right) dF(y|X_j,\hat{\theta})\right) \times \exp\left(\mathbf{i}.\xi'\Phi_2(X_j)\right)$$

and let

$$\widehat{Z}_n(\tau,\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(\mathbf{i}.\tau'\widehat{\Phi}_1(Y_j)\right) - \int \exp\left(\mathbf{i}.\tau'\widehat{\Phi}_1(y)\right) dF(y|X_j,\hat{\theta})\right) \times \exp\left(\mathbf{i}.\xi'\widehat{\Phi}_2(X_j)\right).$$

Then the following results hold.

LEMMA 2. *Under the null hypothesis and Assumptions 1–3,*

$$\sup_{(\tau,\xi)\in\Upsilon\times\Xi}\left|\widehat{Z}_n(\tau,\xi) - Z_n(\tau,\xi)\right| \xrightarrow{p} 0,$$

whereas under the alternative hypothesis,

$$\sup_{(\tau,\xi)\in\Upsilon\times\Xi}\left|\widehat{Z}_n(\tau,\xi) - Z_n(\tau,\xi)\right| / \sqrt{n} \xrightarrow{p} 0.$$

Proof. See the Appendix. ■

Although choice of the bounded transformations Φ_1 and Φ_2 does not affect the consistency of our ICM test, provided that Y_j and X_j are appropriately standardized, it will affect the small-sample power in addition to the null distribution.[3] However, the extent to which the choice of Φ_1 and Φ_2 affects the small-sample power depends on the actual data-generating process, and therefore it is difficult, if not impossible, to devise a general rule for choosing Φ_1 and Φ_2.

2.5. Null Distribution

To analyze the limiting null distribution of $\hat{T}_n$ along the lines in Bierens and Ploberger (1997) we need a generalized version of Mercer's theorem for complex-valued symmetric positive semidefinite functions.

A complex-valued positive semidefinite function relative to a probability measure μ defined on the Borel sets in a euclidean space $\mathbb{R}^q$ is a Borel measurable function $\Gamma(\beta_1,\beta_2) : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{C}$, such that for any complex-valued Borel measurable function $\psi(\beta)$, $\int\int \psi(\beta_1)\Gamma(\beta_1,\beta_2)\overline{\psi(\beta_2)}\mu(d\beta_1)\mu(d\beta_2) \geq 0$. The covariance function (13) is such a function, with $\beta = (\tau,\xi)$, $q = m+k$, and μ the uniform probability measure on $\mathbf{B} = \Upsilon \times \Xi$. Moreover, the covariance function (13) is symmetric, in the sense that $\Gamma(\beta_1,\beta_2) = \overline{\Gamma(\beta_2,\beta_1)}$, and is continuous on $\mathbf{B}\times\mathbf{B}$.

LEMMA 3. *Let μ be a probability measure with compact support $\mathbf{B} \subset \mathbb{R}^q$ and let $\Gamma(\beta_1,\beta_2) : \mathbf{B}\times\mathbf{B} \to \mathbb{C}$ be a symmetric and continuous positive semidefinite function relative to μ. Consider the eigenvalue equation*

$$\lambda.\psi(\beta_1) = \int \Gamma(\beta_1,\beta_2)\overline{\psi(\beta_2)}\mu(d\beta_2),$$

where λ is an eigenvalue with corresponding eigenfunction $\psi(\cdot)$. This eigenvalue equation has countable-infinite many solutions,

$$\lambda_j.\psi_j(\beta_1) = \int \Gamma(\beta_1,\beta_2)\overline{\psi_j(\beta_2)}\mu(d\beta_2),$$

$j = 1,2,3,\ldots$. The eigenvalues λ_j are real valued[4] *and nonnegative and satisfy $\sum_{j=1}^{\infty}\lambda_j < \infty$. The eigenfunctions $\psi_j(\cdot)$ are complex valued and continuous on $\mathbf{B}$ and can be chosen orthonormal; i.e.,*

$$\int \psi_{j_1}(\beta)\overline{\psi_{j_2}(\beta)}\mu(d\beta) = \mathbf{1}(j_1 = j_2).$$

The function Γ has the series representation[5]

$$\Gamma(\beta_1,\beta_2) = \sum_{j=1}^{\infty}\lambda_j\psi_j(\beta_1)\overline{\psi_j(\beta_2)}. \tag{18}$$

Moreover, every complex-valued continuous function Z on $\mathbf{B}$ *can be written as* $Z(\beta) = \sum_{j=1}^{\infty} g_j \psi_j(\beta)$, *where* $g_j = \int Z(\beta)\overline{\psi_j(\beta)}\mu(d\beta)$ *satisfying* $\sum_{j=1}^{\infty} |g_j|^2 < \infty$.

Proof. See Hadinejad-Mahram, Dahlhaus, and Blomker (2002) and Krein (1998). ■

LEMMA 4. *Suppose that the function Z in Lemma 3 is a zero-mean complex-valued continuous Gaussian process on* $\mathbf{B}$ *with covariance function*

$$\Gamma(\beta_1, \beta_2) = \mathrm{E}\left[Z(\beta_1)\overline{Z(\beta_2)}\right].$$

Then the Fourier coefficients $g_j = \int Z(\beta)\overline{\psi_j(\beta)}\mu(d\beta)$ *satisfy*

$$\begin{pmatrix} \mathrm{Re}(g_j) \\ \mathrm{Im}(g_j) \end{pmatrix} = \sqrt{\lambda_j} e_j,$$

where the e_j's are independently $N_2[0, I_2]$ *distributed and the λ_j's are the corresponding eigenvalues of* Γ. *Consequently,*

$$\int |Z(\beta)|^2 \mu(d\beta) = \sum_{j=1}^{\infty} \lambda_j e_j' e_j.$$

Proof. See the Appendix. ■

It follows now straightforwardly from Lemma 4 that the following result holds.

THEOREM 2. *Denote* $g_j = \int_{\Upsilon\times\Xi} Z(\tau,\xi)\overline{\psi}_j(\tau,\xi)d\mu(\tau,\xi)$, *where the ψ_j's are the eigenfunctions of the covariance function (13). Under H_0 and the conditions of Theorem 1,*

$$T = \int_{\Upsilon\times\Xi} |Z(\tau,\xi)|^2 d\mu(\tau,\xi) = \sum_{j=1}^{\infty} |g_j|^2 \sim \sum_{j=1}^{\infty} \lambda_j \chi^2_{2,j},$$

where the $\chi^2_{2,j}$'s are independently χ^2_2 distributed and the λ_j's are the corresponding eigenvalues of the covariance function (13).

Because the covariance function (13) depends on the joint distribution of Y and X, so do the eigenvalues λ_j, and hence the distribution of T is case dependent. However, the critical values of the test can be computed by a parametric bootstrap method, as will be shown in Section 2.8.

2.6. Local Power

Let $Q(y|X)$ be a conditional distribution function that is not identically equal to $F(y|X,\theta_0)$; i.e.,

$$\Pr\left[\sup_{y\in\mathbb{R}^m} |Q(y|X) - F(y|X,\theta_0)| = 0\right] < 1. \qquad \textbf{(19)}$$

Consider the $\sqrt{n}$-local alternative

$$H_1^L : F_n(y|X,\theta_0) = \left(1 - n^{-1/2}\right) F(y|X,\theta_0) + n^{-1/2} Q(y|X).$$

It follows straightforwardly from (10) that under H_1^L,

$$\mathrm{E}[Z(\tau,\xi)] = \varphi_Q(\tau,\xi) - \varphi_F(\tau,\xi), \tag{20}$$

where

$$\varphi_Q(\tau,\xi) = \mathrm{E}\left[\left(\int \exp(\mathbf{i}.\tau' y) dQ(y|X)\right) \exp(\mathbf{i}.\xi' X)\right], \tag{21}$$

$$\varphi_F(\tau,\xi) = \mathrm{E}\left[\left(\int \exp(\mathbf{i}.\tau' y) dF(y|X,\theta_0)\right) \exp(\mathbf{i}.\xi' X)\right].$$

Recall from Lemma 3 that we can write $\mathrm{E}[Z(\tau,\xi)] = \sum_{j=1}^{\infty} \eta_j \psi_j(\tau,\xi)$, where

$$\eta_j = \int_{\Upsilon\times\Xi} \mathrm{E}[Z(\tau,\xi)] \overline{\psi}_j(\tau,\xi) d\mu(\tau,\xi).$$

It follows now from Lemma 4 that, under H_1^L,

$$\begin{aligned}
T_{alt} &= \int_{\Upsilon\times\Xi} |Z(\tau,\xi)|^2 d\mu(\tau,\xi) = \sum_{j=1}^{\infty} |\eta_j + g_j|^2 \\
&= \sum_{j=1}^{\infty} \left(\mathrm{Re}(\eta_j) + \mathrm{Re}(g_j)\right)^2 + \sum_{j=1}^{\infty} \left(\mathrm{Im}(\eta_j) + \mathrm{Im}(g_j)\right)^2 \\
&\sim \sum_{j=1}^{\infty} \left(\mathrm{Re}(\eta_j) + \sqrt{\lambda_j} e_{1,j}\right)^2 + \sum_{j=1}^{\infty} \left(\mathrm{Im}(\eta_j) + \sqrt{\lambda_j} e_{2,j}\right)^2,
\end{aligned}$$

where the $e_{1,j}$'s and $e_{2,j}$'s are independently $N(0,1)$ distributed.

Because condition (19) implies that $\varphi_Q(\tau,\xi)$ and $\varphi_F(\tau,\xi)$ are not identical on $\Upsilon \times \Xi$, at least one η_j is nonzero. Therefore, it follows similar to Corollary 1 in Bierens and Ploberger (1997) that

$$\Pr[T_{alt} > K] > \Pr[T > K]$$

for all $K > 0$, and so the ICM test has nontrivial power against $\sqrt{n}$-local alternatives.

2.7. Maximizing the ICM Test over the Integration Domain

The choice of the hypercubes Υ and Ξ defined by (3) and (4), respectively, does not affect the consistency of the ICM tests but may affect the small-sample power. Therefore, we may improve the small-sample power by maximizing the ICM statistic $\widehat{T}_n$ to Υ and Ξ, under the restrictions $\underline{\Upsilon} \subset \Upsilon \subset \overline{\Upsilon}$ and $\underline{\Xi} \subset \Xi \subset \overline{\Xi}$, where $\underline{\Upsilon}$ and $\overline{\Upsilon}$ are given hypercubes in $\mathbb{R}^m$ of the type (3) and $\underline{\Xi}$ and $\overline{\Xi}$ are given hypercubes in $\mathbb{R}^k$ of the type (4), provided it can be shown that under the null hypothesis,

$$\sup_{\underline{\Upsilon} \subset \Upsilon \subset \overline{\Upsilon},\ \underline{\Xi} \subset \Xi \subset \overline{\overline{\Xi}}} \int_{\Upsilon \times \Xi} |Z_n(\tau,\xi)|^2 d\mu(\tau,\xi)$$
$$\xrightarrow{d} \sup_{\underline{\Upsilon} \subset \Upsilon \subset \overline{\Upsilon},\ \underline{\Xi} \subset \Xi \subset \overline{\overline{\Xi}}} \int_{\Upsilon \times \Xi} |Z(\tau,\xi)|^2 d\mu(\tau,\xi). \tag{22}$$

Indeed, (22) is true, as will be shown for the following special case.

THEOREM 3. *Let* $\Upsilon(c) = [-c,c]^m$ *and* $\Xi(c) = [-c,c]^k$, *where* $c \in [\underline{c},\overline{c}]$, *with* $0 < \underline{c} < \overline{c} < \infty$ *given constants, and let*

$$\widehat{T}_n(c) = \frac{1}{(2c)^{m+k}} \int_{\Upsilon(c)\times\Xi(c)} |Z_n(\tau,\xi)|^2 d\tau d\xi,$$
$$T(c) = \frac{1}{(2c)^{m+k}} \int_{\Upsilon(c)\times\Xi(c)} |Z(\tau,\xi)|^2 d\tau d\xi. \tag{23}$$

Then, under Assumptions 1 and 2 and H_0, $\sup_{\underline{c}\le c\le\overline{c}} \widehat{T}_n(c) \xrightarrow{d} \sup_{\underline{c}\le c\le\overline{c}} T(c)$.

Proof. See the Appendix. ∎

Although it is too much of a computational burden to compute this supremum exactly, let alone the supremum in (22), this result is a motivation to conduct the ICM test for various values of c, and use the maximum of $\widehat{T}_n(c)$ for these values as the actual ICM test, as is done by Bierens and Carvalho (2007) in testing logit model specifications in nonlinear regression form.

2.8. Parametric Bootstrap

Since the seminal work by Efron (1979), bootstrap has become a popular method for deriving null distributions of tests, especially if the null distribution cannot be derived analytically or is case dependent. Bickel and Freedman (1981) developed the asymptotic theory for general bootstrap cases. Conditions under which the bootstrap method fails can be found in Athreya (1987). For more discussion on the bootstrap, see Chernick (1999).

In this section we set forth mild additional conditions for the asymptotic validity of the following parametric bootstrap approach, which is an adaptation to the ICM case of the bootstrap method proposed by Li and Tkacz (1996). Given the null distribution model $F(y|X;\theta)$ and the QML estimator $\widehat{\theta}$, generate M bootstrap samples $\left\{(\widetilde{Y}_{b,1},X_1),\ldots,(\widetilde{Y}_{b,n},X_n)\right\}$, $b = 1,\ldots,M$, where $\widetilde{Y}_{b,j}$ is a random drawing from $F(y|X_j;\widehat{\theta})$ in bootstrap sample b. The vectors

X_j of covariates are the same as in the actual sample. Let $\widetilde{\theta}_b$ be the ML estimator on the basis of this bootstrap sample, i.e., $\widetilde{\theta}_b = \arg\max_{\theta\in\Theta} \ln L_{b,n}(\theta)$, where $\ln L_{b,n}(\theta) = \sum_{j=1}^{n} \ell(\widetilde{Y}_{b,j}, X_j;\theta)$.

Without loss of generality we may assume that $(Y_j', X_j')'$ and $(\widetilde{Y}_{b,j}', X_j')'$ are bounded random vectors. Then the bootstrap ICM test statistic (in the exact ICM case) is

$$\widehat{T}_{b,n} = \int_{\Upsilon\times\Xi} \left|Z_{b,n}(\tau,\xi)\right|^2 d\mu(\tau,\xi),$$

where

$$\widehat{Z}_{b,n}(\tau,\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(\mathbf{i}.\tau'\widetilde{Y}_{b,j}\right) - \varphi(\tau|X_j;\widetilde{\theta}_b)\right)\exp\left(\mathbf{i}.\xi' X_j\right)$$

with $\varphi(\tau|X_j;\widetilde{\theta}_b) = \int \exp(\mathbf{i}.\tau' y)dF(y|X_j;\widetilde{\theta}_b)$. We will set forth conditions such that $\widehat{Z}_{b,n} \Rightarrow Z_b$ as $n\to\infty$, where Z_b is a zero-mean complex-valued Gaussian process on $\Upsilon\times\Xi$ with the same covariance function as the limiting process Z in Theorem 1.

The first step is to set forth conditions such that $(\widetilde{\theta}_b - \widehat{\theta}) \xrightarrow{p} 0$. As is well known, the standard proof of the consistency of QML estimators is based on the uniform law of large numbers for the log-likelihood divided by the sample size. Rather than listing the primitive conditions involved, which are standard, we simply assume that the uniform convergence results involved hold.

Assumption 4. Let $G(x)$ be the distribution function of X_j. The function $\kappa(\theta_1,\theta_2) = \int\int \ell(y,x;\theta_1)dF(y|x;\theta_2)dG(x)$ is continuous on $\Theta\times\Theta$. Moreover,

$$p\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^{n}\int \ell(y,X_j;\theta)dF(y|X_j;\theta) - \kappa(\theta,\theta)\right| = 0$$

and

$$p\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|n^{-1}\ln L_{b,n}(\theta) - \frac{1}{n}\sum_{j=1}^{n}\int \ell(y,X_j;\theta)dF(y|X_j;\widehat{\theta})\right| = 0.$$

Furthermore, $\theta_0 = \arg\max_{\theta\in\Theta}\kappa(\theta,\theta_0)$.

Then it follows from Assumption 1 that $\sup_{\theta\in\Theta}|n^{-1}\ln L_{b,n}(\theta) - \kappa(\theta,\theta_0)| \xrightarrow{p} 0$, which in turn implies[6] that $\widetilde{\theta}_b = \arg\max_{\theta\in\Theta}\ln L_{b,n}(\theta)/n \xrightarrow{p} \theta_0$.

The next step is to show that $\sqrt{n}(\widetilde{\theta}_b - \widehat{\theta})$ has the same limiting distribution as $\sqrt{n}(\widehat{\theta} - \theta_0)$. For this we need the following standard regularity conditions on the vector $\Delta\ell(y,X;\theta)$ and the matrix $\Delta^2\ell\left(y,X_j;\theta\right)$ defined in Assumption 1.

Assumption 5. The elements and components, respectively, of $\int \Delta\ell(y, X_j;\theta)\, dF(y|X_j;\theta)$ and $\int \Delta^2\ell\left(y, X_j;\theta\right) dF(y|X_j;\theta)$ are a.s. continuous on an arbitrary open neighborhood Θ_0 of θ_0, and

$$\int \Delta\ell(y, X;\theta)\, dF(y|X;\theta) = \left(\partial/\partial\theta'\right) \int \ell(y, X;\theta)\, dF(y|X;\theta) = 0 \tag{24}$$

on Θ_0. Moreover, for an arbitrarily small $\delta > 0$,

$$\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left\|\int \Delta\ell\left(y, X_j;\theta\right) dF(y|X_j;\theta)\right\|^{2+\delta}\right] < \infty, \tag{25}$$

$$\mathrm{E}\left[\sup_{\theta\in\Theta_0}\left\|\int \Delta^2\ell\left(y, X_j;\theta\right) dF(y|X_j;\theta)\right\|\right] < \infty. \tag{26}$$

The matrix norm $||\cdot||$ in (26) is the maximum absolute value of the elements of the matrix involved.

Note that Assumption 1 and part (24) of Assumption 5 imply that

$$\lim_{n\to\infty} \Pr\left[\int \Delta\ell\left(y, X_j;\widehat{\theta}\right) dF(y|X_j;\widehat{\theta}) = 0\right] = 1.$$

Next, let $\mathcal{D}_n = \sigma\left(\left\{(Y_j, X_j)\right\}_{j=1}^n\right)$ be the σ-algebra generated by the sample. Then, conditional on $\mathcal{D}_n$,

$$U_{j,n} = \Delta\ell\left(\widetilde{Y}_{b,j}, X_j;\widehat{\theta}\right) - \int \Delta\ell\left(y, X_j;\widehat{\theta}\right) dF(y|X_j;\widehat{\theta})$$

is a double array of independent random vectors, for which Liapounov's central limit theorem applies. See, e.g., Chung (1974, p. 200). This is the reason for the δ in (25). In particular, choose an arbitrary nonzero vector $\xi \in \mathbb{R}^p$ and denote $z_{j,n} = \xi' U_{j,n}$ and $\sigma_n^2 = \frac{1}{n}\sum_{j=1}^n \xi'\mathrm{E}[U_{j,n}U_{j,n}'|\mathcal{D}_n]\xi$. Then it follows from Liapounov's central limit theorem and Assumptions 1 and 5 that $(1/\sqrt{n})\sum_{j=1}^n z_{j,n} \xrightarrow{d} N[0,\sigma]$, where $\sigma^2 = p\lim_{n\to\infty}\sigma_n^2 = \xi' B\xi$. This result can also be proved using a martingale difference central limit theorem. See McLeish (1974) for the latter. Thus,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \Delta\ell\left(\widetilde{Y}_{b,j}, X_j;\widehat{\theta}\right) \xrightarrow{d} N_p[0, B], \tag{27}$$

where B is defined in Assumption 1. It is now a standard exercise to verify that

$$\sqrt{n}(\widetilde{\theta}_b - \widehat{\theta}) \xrightarrow{d} N_p\left[0, A^{-1}BA^{-1}\right],$$

where A is the same as in Assumption 1.

It is now easy to verify, similar to Lemma 1, that under Assumptions 1–2 and 4–5,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varphi(\tau|X_j;\widetilde{\theta}_b)\exp\left(\mathbf{i}.\xi'X_j\right)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varphi\left(\tau|X_j;\widehat{\theta}\right)\exp\left(\mathbf{i}.\xi'X_j\right)$$
$$+b(\tau,\xi)'A^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Delta\ell\left(\widetilde{Y}_{b,j},X_j;\widehat{\theta}\right)+o_p(1)$$

uniformly in $(\tau,\xi)\in\Upsilon\times\Xi$, where $b(\tau,\xi)$ is the same as in Lemma 1. Consequently, denoting

$$\phi_{b,j}(\tau,\xi|\widehat{\theta})=\left(\exp\left(\mathbf{i}.\tau'\widetilde{Y}_{b,j}\right)-\varphi\left(\tau|X_j;\widehat{\theta}\right)\right)\exp\left(\mathbf{i}.\xi'X_j\right)$$
$$+b(\tau,\xi)'A^{-1}\Delta\ell\left(\widetilde{Y}_{b,j},X_j;\widehat{\theta}\right), \tag{28}$$

$$\widetilde{Z}_{b,n}(\tau,\xi)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_{b,j}(\tau,\xi|\widehat{\theta}), \tag{29}$$

we have

$$\sup_{(\tau,\xi)\in\Upsilon\times\Xi}\left|\widehat{Z}_{b,n}(\tau,\xi)-\widetilde{Z}_{b,n}(\tau,\xi)\right|=o_p(1).$$

Therefore, it suffices to prove that $\widetilde{Z}_{b,n}\Rightarrow Z_*$ on $\Upsilon\times\Xi$, where Z_* is a zero-mean complex-valued Gaussian process with the same covariance function as Z in Theorem 1.

Denote

$$\widetilde{Z}_{1,b,n}(\tau,\xi)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(\mathbf{i}.\tau'\widetilde{Y}_{b,j}\right)-\varphi\left(\tau|X_j;\widehat{\theta}\right)\right)\exp\left(\mathbf{i}.\xi'X\right),$$

$$\widetilde{Z}_{2,b,n}(\tau,\xi)=b(\tau,\xi)'A^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Delta\ell\left(\widetilde{Y}_{b,j},X_j;\widehat{\theta}\right).$$

Because $b(\tau,\xi)$ is uniformly continuous on $\Upsilon\times\Xi$, the tightness of $\widetilde{Z}_{2,b,n}$ follows from (27). The tightness of $\widetilde{Z}_{1,b,n}$ follows from the proof of Theorem 1, simply by replacing the expectations involved by the corresponding conditional expectations $\mathrm{E}[\cdot|\mathcal{D}_n]$. Moreover, similar to (27) it can be shown that the finite distributions of $\widetilde{Z}_{1,b,n}(\tau,\xi)$ converge to a multivariate normal distribution. Thus, $\widetilde{Z}_{b,n}\Rightarrow Z_*$. Furthermore, it is easy to verify that the covariance function of this limiting process takes the form

$$p\lim_{n\to\infty}\mathrm{E}\left[\widetilde{Z}_{b,n}(\tau_1,\xi_1)\overline{\widetilde{Z}_{b,n}(\tau_2,\xi_2)}\middle|\mathcal{D}_n\right]$$
$$=p\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\left[\phi_{b,j}(\tau_1,\xi_1|\widehat{\theta})\overline{\phi_{b,j}(\tau_2,\xi_2|\widehat{\theta})}\middle|\mathcal{D}_n\right]$$
$$=\Gamma\left((\tau_1,\xi_1),(\tau_2,\xi_2)\right),$$

where the latter is the same as in Theorem 1. This proves the following result.

THEOREM 4. *Let Y and X be bounded random vectors.*[7] *Then under Assumptions 1–2 and 4–5, $\widetilde{Z}_{b,n} \Rightarrow Z_b$ on $\Upsilon \times \Xi$, where the Z_b's have the same distribution as the process Z in Theorem 1 and are independent. Hence by the continuous mapping theorem,*

$$\left(\widehat{T}_{1,n}, \ldots, \widehat{T}_{M,n}\right)' \xrightarrow{d} (T_1, \ldots, T_M)', \tag{30}$$

where the T_b's are independent random drawings from the distribution of $T = \int_{\Upsilon \times \Xi} |Z(\tau, \xi)|^2 d\mu(\tau, \xi)$.

Note that (30) carries over if the components of these random vectors are sorted in decreasing order. Then it follows that for $\alpha \in (0, 1)$, $\widehat{T}_{[\alpha M],n} \xrightarrow{d} T_{[\alpha M]}$, where $[\alpha M]$ is the largest integer $\leq \alpha M$. The statistic $\widehat{T}_{[\alpha M],n}$ is the $\alpha \times 100\%$ bootstrap critical value, and $T_{[\alpha M]}$ is approximately the actual asymptotic $\alpha \times 100\%$ critical value of T.

It is not hard to verify that this bootstrap approach remains valid after standardizing and transforming the variables involved. The same applies to the simulated ICM test proposed in the next section.

3. THE SIMULATED ICM TEST

3.1. How to Avoid Numerical Integration

The theoretical conditional characteristic function poses a computational challenge in two ways. First, some conditional distributions have no closed-form expression for their characteristic functions, especially if Y has to be transformed first by a bounded one-to-one transformation. But even for distributions with closed-form characteristic functions the integration over τ has to be carried out numerically, which is time consuming, especially if Y is multidimensional. Moreover, the need for numerical integration will slow down the bootstrap too much.

To cope with this problem, a simulated integrated conditional moment (SICM) test will be proposed, in which the process $Z_n(\tau, \xi)$ in the exact ICM test statistic is replaced by either

$$\widehat{Z}_n^{(s)}(\tau, \xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\exp(\mathbf{i}.\tau' Y_j) - \exp(\mathbf{i}.\tau' \tilde{Y}_j)\right) \exp(\mathbf{i}.\xi' X_j)$$

if Y_j and X_j are bounded random vectors, or

$$\widehat{Z}_n^{(s)}(\tau, \xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\exp(\mathbf{i}.\tau' \Phi_1(Y_j)) - \exp(\mathbf{i}.\tau' \Phi_1(\tilde{Y}_j)\right) \exp(\mathbf{i}.\xi' \Phi_2(X_j))$$

if not, where $\tilde{Y}_j$ is a random drawing from the estimated conditional distribution $F(y|X_j;\hat{\theta})$, and in the latter case $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ are bounded one-to-one mappings. The SICM test statistic is then

$$\widehat{T}_n^{(s)} = \int_{\Upsilon\times\Xi} |\widehat{Z}_n^{(s)}(\tau,\xi)|^2 d\mu(\tau,\xi).$$

THEOREM 5. *Let the conditions of Theorem 1 hold. Write* $\widehat{Z}_n^{(s)}(\tau,\xi)$ *as* $\widehat{Z}_n^{(s)}(\tau,\xi) = Z_n(\tau,\xi) - \widetilde{Z}_n^{(s)}(\tau,\xi)$, *where* $Z_n(\tau,\xi)$ *is the process (6) and*

$$\widetilde{Z}_n^{(s)}(\tau,\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp(i\tau'\tilde{Y}_j) - \int \exp(i\tau' y)dF(y|X_j,\hat{\theta})\right)\exp(i\xi X_j).$$

Under H_0, $\widehat{T}_n^{(s)} \xrightarrow{d} T_s = \int_{\Upsilon\times\Xi} |Z(\tau,\xi) - Z_s(\tau,\xi)|^2 d\mu(\tau,\xi)$, *where Z is the same as in Theorem 1 and* Z_s *is a complex-valued zero-mean Gaussian process with covariance function*

$$\begin{aligned}\Gamma_s&((\tau_1,\xi_1),(\tau_2,\xi_2))\\ &= \mathrm{E}\left[(\varphi(\tau_1-\tau_2|X;\theta_0) - \varphi(\tau_1|X;\theta_0)\varphi(-\tau_2|X;\theta_0))\exp(\mathbf{i}(\xi_1-\xi_2)'X)\right].\end{aligned}$$

Moreover, Z and Z_s *are independent. Under* H_1,

$$\widehat{T}_n^{(s)}/n \xrightarrow{p} \int_{\Upsilon\times\Xi} |\eta(\tau,\xi)|^2 d\mu(\tau,\xi) > 0,$$

which is the same as (14).

Proof. See the Appendix. ■

Note that under H_0, $\overline{Z}_s(\tau,\xi) = Z(\tau,\xi) - Z_s(\tau,\xi)$ has covariance function

$$\overline{\Gamma}_s((\tau_1,\xi_1),(\tau_2,\xi_2)) = \Gamma((\tau_1,\xi_1),(\tau_2,\xi_2)) + \Gamma_s((\tau_1,\xi_1),(\tau_2,\xi_2)),$$

where Γ is defined by (13). Clearly, all the previous results for the exact ICM test carry over to the SICM test, simply by replacing Z with $\overline{Z}_s$ and Γ with $\overline{\Gamma}_s$.

The main advantage of the SICM test is that the validity of quite complicated conditional distribution models $F(y|X;\theta)$ can be tested, as long as it is feasible to generate random drawings $\widetilde{Y}$ from them. Another advantage is that $\widehat{T}_n^{(s)}$ has a closed form. In particular, if the Y_j's and X_j's are bounded then, with $Y_{i,j}$, $\widetilde{Y}_{i,j}$, and $X_{i,j}$ components i of Y_j, $\widetilde{Y}_j$, and X_j, respectively, the SICM version of the test statistic (23) takes the form

$$\begin{aligned}
\widehat{T}_n^{(s)}(c) &= \frac{1}{(2c)^{k+m}} \int_{[-c,c]^m} \int_{[-c,c]^k} |\widehat{Z}_n^{(s)}(\tau,\xi)|^2 d\tau d\xi \\
&= \frac{2}{n} \sum_{j_1=1}^{n-1} \sum_{j_2=j_1+1}^{n} \left(\prod_{i=1}^{m} \frac{\sin\left(c(Y_{i,j_1} - Y_{i,j_2})\right)}{c(Y_{i,j_1} - Y_{i,j_2})} + \prod_{i=1}^{m} \frac{\sin\left(c(\widetilde{Y}_{i,j_1} - \widetilde{Y}_{i,j_2})\right)}{c(\widetilde{Y}_{i,j_1} - \widetilde{Y}_{i,j_2})} \right. \\
&\qquad \left. - \prod_{i=1}^{m} \frac{\sin\left(c(Y_{i,j_1} - \widetilde{Y}_{i,j_2})\right)}{c(Y_{i,j_1} - \widetilde{Y}_{i,j_2})} - \prod_{i=1}^{m} \frac{\sin\left(c(\widetilde{Y}_{i,j_1} - Y_{i,j_2})\right)}{c(\widetilde{Y}_{i,j_1} - Y_{i,j_2})} \right) \\
&\quad \times \left(\prod_{i=1}^{k} \frac{\sin\left(c(X_{i,j_1} - X_{i,j_2})\right)}{c(X_{i,j_1} - X_{i,j_2})} \right) \\
&\quad + \frac{2}{n} \sum_{j=1}^{n} \left(1 - \prod_{i=1}^{m} \frac{\sin\left(c(Y_{i,j} - \widetilde{Y}_{i,j})\right)}{c(Y_{i,j} - \widetilde{Y}_{i,j})} \right) \qquad \textbf{(31)}
\end{aligned}$$

as is not hard to verify.

Note that $\lim_{c\downarrow 0} \widehat{T}_n^{(s)}(c) = \lim_{c\to\infty} \widehat{T}_n^{(s)}(c) = 0$, and so choosing too small or too large a c will destroy the small-sample power of the test. This is another reason for computing $\widehat{T}_n^{(s)}(c)$ for a range of values of c, e.g., $c_1 < c_2 < \cdots < c_r$, and using the MAXSICM statistic

$$\max_{i=1,2,\ldots,r} \widehat{T}_n^{(s)}(c_i) \qquad \textbf{(32)}$$

as the actual test statistic.

3.2. Small-Sample Performance

Zheng (2000) presents small-sample simulation results for data-generating processes of the type $Y = 1 + X + U$, where $X \sim N(0,1)$ and

$H_Z^{(0)}$: $U|X \sim N(0,1)$,

$H_Z^{(1)}$: $U|X \sim$ Standard Logistic,

$H_Z^{(2)}$: $U|X \sim t_5$,

$H_Z^{(3)}$: $U|X \sim N(0, X^2)$.

In each of these cases the null hypothesis is that the data-generating process corresponds to the classical linear regression model

$$H_0 : Y = \alpha + \beta X + U \quad \text{with } U|X \sim N(0,\sigma^2).$$

To compare the small-sample performance of our MAXSICM test with Zheng's tests for these cases we have generated 1,000 replications of samples of size 200 from the preceding data-generating processes and used bootstrap samples of size 500 to compute the bootstrap p-values of our test. In the first instance we have chosen the range $c = 5, 10, 15, 20, 25$ for the constants c_i in (32). Moreover, to see how sensitive the size and power of the SICM test are for the choice of c, we present the results for the five separate SICM(c) tests also. Furthermore, we also include the corresponding results for the Andrews (1997) CK test (1), which is bootstrapped in the same way as the SICM tests. In the case of the SICM tests both Y and X are standardized by taking them in deviation of their sample means, dividing them by their sample standard errors, and then using the arctan transformation to make them bounded. The simulated $\widetilde{Y}$'s are transformed similarly, using the sample mean and sample standard error of the actual Y variables.

The results are presented in Table 1, where Max is the MAXSICM test (32), CK is the Andrews (1997) CK test (1), and the Z(·)'s are the corresponding results of the Zheng (2000) tests[8] for bandwidth constants $0.5, 1.0, 2.0$. The numerical entries in Table 1 are the rejection percentages at the 1, 5, and 10% significance levels, respectively.

Note that Zheng's results are based on the right-sided critical values of the standard normal asymptotic null distribution. Apparently, the small-sample null distribution of Zheng's test is not yet normal. Surprisingly, the small-sample power of Zheng's test against the alternative $U|X \sim t_5$ is much higher than of the MAXSICM and CK tests. We do not have an explanation of this. The power of all the tests against the logistic alternative is low, which is not surprising because the shapes of the logistic and normal densities differ mainly in the tails, and so we need much larger samples to distinguish them. All tests do a good job against the heteroskedastic alternative.

Observe that the small-sample power of the SICM(c) test is maximal for $c = 5$. At first sight it is surprising that the SICM(5) test is even more powerful than the MAXSICM test because the latter statistic is always larger than the individual SICM statistics. However, the bootstrap critical values of the MAXSICM test are also always larger than those of the individual SICM tests, which in

TABLE 1. Monte Carlo results for the linear regression model: Part 1

	$H_Z^{(0)}$			$H_Z^{(1)}$			$H_Z^{(2)}$			$H_Z^{(3)}$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$c = 5$	1.5	4.8	10.5	2.1	8.2	15.1	6.6	17.5	26.6	100	100	100
$c = 10$	1.3	5.4	9.9	1.6	7.4	13.1	3.7	12.0	21.5	100	100	100
$c = 15$	1.0	5.6	10.6	1.5	5.8	11.6	3.0	9.3	18.1	99.6	100	100
$c = 20$	1.2	5.6	9.7	1.4	5.4	11.1	2.2	7.8	14.2	99.1	99.9	100
$c = 25$	1.0	5.3	11.0	1.5	6.1	11.0	1.7	7.5	13.4	96.4	99.0	99.8
Max	1.6	5.2	9.7	1.8	6.9	13.7	5.4	16.0	22.7	100	100	100
CK	1.2	5.7	10.6	3.7	11.2	20.0	6.4	19.8	28.5	99.8	100	100
Z(0.5)	1.3	2.2	4.6	12.4	15.1	17.2	31.6	40.2	46.1	95.1	95.3	95.3
Z(1.0)	1.7	2.5	4.0	12.1	14.2	16.7	43.7	55.0	59.5	96.3	96.6	96.7
Z(2.0)	2.8	3.0	4.7	18.7	22.8	25.6	55.5	66.6	74.0	89.8	93.1	94.0

small samples may result in a lower power than the best individual SICM test. Moreover, recall from Theorem 5 that the variability of the SICM test statistic comes from two sources, namely, the asymptotically independent empirical processes $Z_n(\tau,\xi)$ and $\widetilde{Z}_n^{(s)}(\tau,\xi)$. The latter adds additional variation to the SICM test statistic, resulting in a reduction of power in small samples compared with the corresponding exact ICM test. These effects will of course vanish in large samples once the consistency of the SICM test kicks in. This may be the reason why the Andrews CK test has somewhat better power against the logistic and t_5 alternatives than the MAXSICM test.

Because the small-sample power of the SICM(c) test in Table 1 is maximal for $c = 5$, we have redone the simulations for $c = 1,2,3,4,5$. The results are presented in Table 2.

Note that the MAXSICM test in Table 2 has slightly better power than the version in Table 1. More remarkably however, observe that the power of the SICM(c) test against $H_Z^{(2)}$ and $H_Z^{(3)}$ increases dramatically with c up to $c = 4$.

Next we consider a model for which the Zheng (2000) test is not applicable, namely, the conditional Poisson model. Another reason for focusing on the conditional Poisson model is that we will use this model in our empirical application. Thus, the null hypothesis is that

$$H_0 : Y|X \sim \text{Poisson}\,(\exp(\alpha+\beta X))$$

with actual data-generating processes of the type Poisson and negative binomial (NB) logit:

$$H_P^{(0)} \;:\; Y|X \sim \text{Poisson}\,(\exp(X)),$$
$$H_P^{(1)} \;:\; Y|X \sim \text{NB}(1, p(X)),$$
$$H_P^{(2)} \;:\; Y|X \sim \text{NB}(5, p(X)),$$
$$H_P^{(3)} \;:\; Y|X \sim \text{NB}(10, p(X)),$$

where $p(x) = (1+\exp(-x))^{-1}$ is the logit function and the single covariate X is standard normally distributed. The simulation setup in Table 3 is the same as

TABLE 2. Monte Carlo results for the linear regression model: Part 2

	$H_Z^{(0)}$			$H_Z^{(1)}$			$H_Z^{(2)}$			$H_Z^{(3)}$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$c=1$	1.5	4.9	9.7	0.5	4.9	10.3	1.7	6.9	11.3	7.0	24.0	42.2
$c=2$	1.0	4.8	10.1	0.8	6.6	12.2	2.9	9.2	16.5	64.1	88.7	96.5
$c=3$	1.0	4.3	8.9	1.8	8.6	15.7	5.6	14.1	24.6	97.7	99.6	99.9
$c=4$	1.1	3.2	8.0	2.5	9.0	16.5	6.3	18.5	28.1	99.9	100	100
$c=5$	1.1	4.1	9.9	2.9	9.7	15.6	6.4	18.5	28.0	100	100	100
Max	1.0	4.1	8.6	2.4	8.8	15.6	5.7	17.4	27.7	100	100	100

TABLE 3. Monte Carlo results for the Poisson model: Part 1

	$H_P^{(0)}$			$H_P^{(1)}$			$H_P^{(2)}$			$H_P^{(3)}$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$c = 5$	1.7	6.0	10.7	75.3	89.8	94.8	69.2	88.9	94.3	47.0	76.9	88.4
$c = 10$	1.0	5.2	10.0	41.2	66.5	78.4	53.6	77.1	87.0	54.3	79.2	89.4
$c = 15$	1.1	4.8	10.8	32.8	57.8	70.9	34.6	60.2	72.9	39.7	67.1	80.6
$c = 20$	0.7	4.4	9.5	32.1	57.4	67.8	23.4	47.2	49.8	28.9	56.0	69.8
$c = 25$	0.8	3.7	9.5	21.7	44.1	60.5	16.3	37.1	49.8	21.1	45.8	61.5
Max	1.0	5.2	9.6	57.8	77.7	85.4	36.5	55.9	67.2	29.1	55.6	68.7
CK	0.5	5.6	11.4	98.3	99.3	99.6	53.6	74.3	83.7	21.1	47.5	61.8

before, and in the first instance we have again chosen the range $5, 10, 15, 20, 25$ for c.

The small-sample power of the SICM(c) test against the alternatives $H_P^{(1)}$ and $H_P^{(2)}$ is maximal for $c = 5$, but now the small-sample power against $H_P^{(3)}$ is maximal for $c = 10$. Moreover, the Andrews CK test has much better small-sample power against the alternatives $H_P^{(1)}$ and $H_P^{(2)}$ than the MAXSICM test, but the MAXSICM test performs better against $H_P^{(3)}$. The reason for the latter may be that the SICM test has better local power properties than the CK test because the decrease in power along the alternatives $H_P^{(1)}$, $H_P^{(2)}$, and $H_P^{(3)}$ is due to the fact that the NB$(m, p(X))$ distribution approaches the conditional Poisson distribution as $m \to \infty$. However, this conjecture has not been analytically verified. Furthermore, the small-sample size properties of the SICM and CK tests are similar.

The simulation results for $c = 1, 2, 3, 4, 5$ are presented in Table 4.

Again we see a dramatic increase of the power of the SICM(c) in c up to $c = 4$. Moreover, note that the power of the MAXSICM test involved is substantially higher than in the case of Table 3.

In conclusion, the results in Tables 1–4 show that the choice of the integration domain of the SICM test is important for the small-sample power of the test.

TABLE 4. Monte Carlo results for the Poisson model: Part 2

	$H_P^{(0)}$			$H_P^{(1)}$			$H_P^{(2)}$			$H_P^{(3)}$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$c = 1$	1.4	5.0	10.0	29.0	59.0	73.2	15.9	35.7	51.3	7.3	21.4	36.8
$c = 2$	1.2	5.2	9.7	56.0	80.4	90.5	34.7	59.2	74.2	14.7	37.8	53.6
$c = 3$	0.8	5.1	9.5	71.3	90.2	96.1	52.3	78.1	89.8	24.1	53.3	71.1
$c = 4$	0.4	4.3	9.6	77.5	93.0	96.6	65.3	86.8	93.4	36.8	67.5	81.2
$c = 5$	0.9	4.4	9.6	73.5	90.1	94.9	70.1	88.3	94.2	48.2	74.6	86.3
Max	0.6	4.4	8.8	72.3	90.4	95.0	67.5	88.1	94.0	44.4	74.3	86.0

4. APPLICATION TO HEALTH ECONOMIC COUNT DATA MODELS

Count data are often encountered in health economics, such as the number of physician visits and the number of days of hospital stays. A popular model for count data is the Poisson distribution. See Cameron and Trivedi (1986). In this section we apply the MAXSICM method to test whether a conditional Poisson model is correctly specified. To the best of our knowledge, the only other consistent specification test available for the conditional Poisson distribution is the Andrews (1997) CK test (1). Lee (1986) has proposed several tests for the validity of the conditional Poisson distribution, but none of his tests are consistent. Cameron and Trivedi (1986) have tested the validity of the Poisson distribution by testing the null hypothesis that the mean and the variance are equal, but this feature is not exclusively a property of the Poisson distribution, and therefore such a test is not consistent.

The data source is the 1987–1988 National Medical Expenditure Survey used by Deb and Trivedi (1997). There are 4,406 observations of individuals over age of 66. The variable Y of interest is the number of physician visits by elderly, which is explained by a vector of various variables of health conditions and demographic characteristics, as listed in Table 5.

However, because it is conceivable that the effects of the covariates X_3 through X_{16} are different for people with excellent health ($X_1 = 1$) and poor health ($X_2 = 1$), we have augmented the list of covariates with $X_1 \times X_j$ and $X_2 \times X_j$ for $j = 3, 4, \ldots, 16$, and so the actual number of covariates is 44. A preliminary

TABLE 5. Model variables

	Name	Meaning
Y	ofp	number of visits to physicians in an office setting
X_1	exclhlth	self-perceived health condition: excellent
X_2	poorhlth	self-perceived health condition: poor
X_3	numchron	number of chronic diseases and conditions
X_4	adldiff	a measure of disability status
X_5	noreast	region: northeast
X_6	midwest	region: midwest
X_7	west	region: west
X_8	age	age in years (divided by 10)
X_9	black	$= 1$ if black
X_{10}	male	$= 1$ if male
X_{11}	married	$= 1$ if married
X_{12}	school	years of schooling
X_{13}	faminc	family income (in \$10,000)
X_{14}	employed	employment status
X_{15}	privins	private insurance status
X_{16}	medicaid	public insurance status

data analysis reveals that the extended Poisson model has lower values for the Hannan–Quinn (Hannan and Quinn, 1979) and Schwarz (1978) information criteria than the model with the original 16 covariates, and so the additional 28 covariates contribute to the fit of the model.

The null hypothesis to be tested is that conditional on these 44 explanatory variables, the number Y of physician visits by the elderly follows a Poisson distribution with conditional expectation $\mu(X) = \exp((1, X')\theta_0)$.

We will use the MAXSICM test (32) to test the Poisson hypothesis, with $c = 5, 10, 15, 20, 25$ and bootstrap sample size 500. It suffices to include only the original 16 covariates in Table 5 as conditioning variables in the test, as conditioning on these 16 covariates is equivalent to conditioning on the augmented list of 44 covariates.[9] The dependent variable Y and the 16 X variables have been standardized and transformed in the same way as in the simulation study in the previous section. To generate random drawings from the Poisson distribution we have used the fact that for independent random drawings e_j from the standard exponential distribution, the smallest integer Y for which $\sum_{j=1}^{Y} e_j > \mu$ has a Poisson(μ) distribution.

The value of the MAXSICM test involved is 193.197, with bootstrap p-value virtually equal to zero (< 0.000005). Thus, the Poisson model is strongly rejected.

As a comparison we have also conducted the Cameron–Trivedi (Cameron and Trivedi, 1990), test based on the regression

$$((Y_j - \widehat{\mu}_j)^2 - Y_j)/\widehat{\mu}_j = \alpha.\widehat{\mu}_j + \varepsilon_j, \tag{33}$$

where $\widehat{\mu}_j = \exp((1, X_j')\widehat{\theta})$ with $\widehat{\theta}$ the ML estimator of θ_0. Under the null hypothesis that the conditional expectation and the conditional variance of Y_j are equal the parameter α should be zero. Therefore, the test statistic involved is the t-value $\widehat{t}$ of the ordinary least squares (OLS) estimate $\widehat{\alpha}$ of α. The results are $\widehat{\alpha} = 0.874068$, $\widehat{t} = 12.7497$. Thus, the Cameron–Trivedi test also strongly rejects the validity of the Poisson model.

As a further comparison we have also tried to conduct the Andrews (1997) CK test (1). However, for the 16 covariates in Table 1 the inequality $X_j < X_i$ for $i \neq j$ never happened, and so the CK test statistic collapses to $\max_{1 \le j \le n} \left|1 - F(Y_j | X_j, \widehat{\theta})\right| / \sqrt{n} < 1/\sqrt{n} \approx 0.015$.

It is well known that if conditional on a gamma(m, β) distributed random variable V, where $m \geq 1$ is integer valued, Y is Poisson$(\mu.V)$ distributed, then the unconditional distribution of Y is negative binomial (m, p), with $p = (1 + \beta\mu)^{-1}$. More generally, if

$$Y | X, V \sim \text{Poisson}\left(V \exp((1, X')\theta_0)\right),$$

where V represents unobserved heterogeneity that is independent of X, and if V is gamma(m, β) distributed, then the conditional distribution of Y given X alone is of the negative binomial logit (NBL) type:

$$Y | X \sim \text{NB}\left(m, p\left(-\ln\beta - (1, X')\theta_0\right)\right), \tag{34}$$

where again $p(x)$ is the logit function. Note that because of the presence of a constant term we may without loss of generality assume that $\mathrm{E}[V] = 1$, which corresponds to $\beta = 1/m$ in the NBL(m) case, and so $\mathrm{E}[Y|X] = \exp((1, X')\theta_0)$. Moreover, it is easy to verify that then the Poisson ML estimator $\widehat{\theta}$ is still consistent and that in the NBL(m) case the OLS estimator $\widehat{\alpha}$ of the parameter α in the Cameron–Trivedi model (33) converges in probability to $1/m$. Because $1/\widehat{\alpha} \approx 1.144$ is somewhat close to $m = 1$ we will therefore now try an NBL(1) model.

The MAXSICM test statistic involved, with the same c values as before, is now 10.796, which is much lower than in the Poisson case. However, the bootstrap p-value is still virtually zero, and so also the NBL(1) model is strongly rejected. The same applies to the NBL(2) model: the MAXSICM test statistic is 15.990 with again virtually zero bootstrap p-value. Moreover, the values of the Hannan–Quinn (1979) and Schwarz (1978) information criteria for the NBL(2) model are much higher than for the NBL(1) case, which is reflected by the higher value of the MAXSICM statistic in the former case. Because apparently these three models are misspecified, the MAXSICM test statistic divided by the sample size may be interpreted as a measure of the deviation of the characteristic function corresponding to the model with the characteristic function corresponding to the data-generating process, and thus it may be interpreted as a measure of fit or the lack thereof.

The model estimation and test computations have been conducted via the free econometric software package EasyReg International developed by the first author. See Bierens (2010).

5. CONCLUDING REMARKS

This paper extends the ICM specification test for the functional form of regression models to specification tests for parametric conditional distributions, on the basis of the integrated squared difference between the empirical conditional characteristic function and the theoretical characteristic function. This test is consistent and has $\sqrt{n}$-local power, and the conditional distributions tested can be of any type: continuous, discrete, or mixed. The null distribution is an infinite weighted sum of independent χ^2_2 random variables, with case-dependent weights, and so the critical values have to be derived via a parametric bootstrap method. To avoid numerical integration for computing the theoretical characteristic function, the simulated integrated conditional moment (SICM) test is proposed, in which the conditional characteristic function implied by the estimated model is simulated using only a single random drawing from this distribution for each data point. This test is much easier and faster to compute than the exact ICM test, whereas it has similar asymptotic properties as the latter test. The SICM test works well for models with a large number of covariates, contrary to the Andrews (1997) conditional Kolmogorov test, and does not suffer from the curse of dimensionality as is the case with the Zheng (2000) test.

The SICM test has been applied to test a conditional Poisson model for count data, using health economics data. The conditional Poisson model is a popular model in health economic research for modeling counts (such as number of doctor's office visits by elderly as in the paper). The SICM test firmly rejects this Poisson model specification, and so does the Cameron–Trivedi test. The result of the latter test suggests that an NBL model with parameter $m = 1$ might be more appropriate for these data. However the NBL(m) models for $m = 1, 2$ are strongly rejected also by the SICM test, although with much lower values of the test statistics than for the Poisson model. At least we can conclude from our results that the Poisson model is inferior to the NBL(1) model, and to a lesser extent the same applies to the NBL(2) model.

These empirical applications merely serve as illustrations of the power of the SICM test. Searching for the right model for the data involved is beyond the scope of this paper.

NOTES

1. Of course, there are many more tests for the validity of conditional distributions, but none of these tests are consistent.
2. We adopt the convention that the partial derivative to a row vector produces a column vector of partial derivatives.
3. Via the covariance function (13).
4. Due to the symmetry of Γ.
5. The result (18) is the actual Mercer's theorem.
6. See, e.g., Theorem 6.11 in Bierens (2004), which originates from Jennrich (1969).
7. Note that then by Assumption 1 the bootstrap $\widetilde{Y}_{b,j}$'s are bounded too.
8. Taken from his Table 1.
9. Because the corresponding σ-algebras are the same.

REFERENCES

Andrews, D.W. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097–1128.

Athreya, K.B. (1987) Bootstrap of the mean in the infinite variance case. *Annals of Statistics* 15, 724–731.

Bickel, P. & D. Freedman (1981) Some asymptotic theory for the bootstrap. *Annals of Statistics* 9, 1196–1217.

Bierens, H.J. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.

Bierens, H.J. (2004) *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge University Press.

Bierens, H.J. (2010) EasyReg International. Free econometrics software downloadable from http://econ.la.psu.edu/~hbierens/EASYREG.HTM.

Bierens, H.J. & J.R. Carvalho (2007) Semi-nonparametric competing risks analysis of recidivism. *Journal of Applied Econometrics* 22, 971–993.

Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Boning, B. & F. Sowell (1999) Optimality of the integrated conditional moment test. *Econometric Theory* 15, 710–719.

Cameron, A.C. & P.K. Trivedi (1986) Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1, 29–53.

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

Chernick, M. (1999) *Bootstrap Methods: A Practitioner's Guide*. Wiley.

Chung, K.L. (1974) *A Course in Probability Theory*. Academic Press.

Deb, P. & P.K. Trivedi (1997) Demand for medical care by the elderly. *Journal of Applied Econometrics* 12, 313–336.

Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7, 1–26.

Escanciano, J.C. (2006) A consistent diagnostic test for regression models using projections. *Econometric Theory* 22, 1030–1051.

Fan, Y.Q. & Q. Li (2000) Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory* 16, 1016–1041.

Gozalo, P.L. (1993) A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory* 9, 451– 477.

Hadinejad-Mahram, H., D. Dahlhaus, & D. Blomker (2002) Karhunen-Loeve Expansion of Vector Random Processes. Technical Report IKT-NT 1019, Communications Technological Laboratory, Swiss Federal Institute of Technology, Zurich.

Hannan, E.J. & B.G. Quinn (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B* 41, 190–195.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.

Horowitz, J.T. & W. Härdle (1994) Testing a parametric model against a semiparametric alternative. *Econometric Theory* 10, 821–848.

Jennrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 40, 633–643.

Krein, M.G. (1998) Compact linear operators on functional spaces with two norms. *Integral Equations and Operator Theory* 30, 140–162.

Kullback, L. & R.A. Leibler (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22, 79–86.

Lavergne, P. & Q.H. Vuong (2000) A nonparametric significance test. *Econometric Theory* 16, 576–601.

Lee, L.F. (1986) Specification test for Poisson regression models. *International Economic Review* 27, 689–706.

Li, F. & G. Tkacz (1996) A consistent bootstrap test for conditional density functions with time-series data. *Journal of Econometrics* 133, 863–886.

Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145–165.

McLeish, D.L. (1974) Dependent central limit theorems and invariance principles. *Annals of Probability* 2, 620–628.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.

Stinchcombe, M. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* 25, 613–641.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.

White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press.

Zheng, J.X. (1996) A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics* 75, 263–289.

Zheng, J.X. (2000) A consistent test of conditional parametric distributions. *Econometric Theory* 16, 667–691.

APPENDIX

Proof of Lemma 1. By the mean value theorem and Assumption 1,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\operatorname{Re}\left(\varphi(\tau|X_j;\widehat{\theta})-\varphi\left(\tau|X_j;\theta_0\right)\right)\exp(\mathbf{i}.\xi' X_j)\right]$$
$$=\sqrt{n}(\widehat{\theta}-\theta_0)'\left(\frac{1}{n}\sum_{j=1}^{n}\operatorname{Re}\left[\Delta\varphi\left(\tau|X_j;\theta_0+\widetilde{\lambda}(\tau,\xi)(\widehat{\theta}-\theta_0)\right)\exp(\mathbf{i}.\xi' X_j)\right]\right),$$

where $\widetilde{\lambda}(\tau,\xi)\in[0,1]$, whereas by the Jennrich (1969) uniform strong law of large numbers and the conditions in Assumption 2,

$$\frac{1}{n}\sum_{j=1}^{n}\operatorname{Re}\left[\Delta\varphi\left(\tau|X_j;\theta\right)\exp(\mathbf{i}.\xi' X_j)\right]\to \mathrm{E}\left(\operatorname{Re}\left[\Delta\varphi\left(\tau|X;\theta\right)\exp(\mathbf{i}.\xi' X)\right]\right)$$

a.s., uniformly in $(\tau,\xi,\theta)\in\Upsilon\times\Xi\times\Theta_0$. This implies that

$$\frac{1}{n}\sum_{j=1}^{n}\operatorname{Re}\left[\Delta\varphi\left(\tau|X_j;\theta_0+\widetilde{\lambda}(\tau,\xi)(\widehat{\theta}-\theta_0)\right)\exp(\mathbf{i}.\xi' X_j)\right]\xrightarrow{p}\operatorname{Re}[b(\tau,\xi)]$$

uniformly on $\Upsilon\times\Xi$ because $\widehat{\theta}\xrightarrow{p}\theta_0$. The same result (with possibly a different $\widetilde{\lambda}(\tau,\xi)$) applies to the $\operatorname{Im}[\cdot]$ case. Consequently,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\varphi(\tau|X_j;\widehat{\theta})-\varphi\left(\tau|X_j;\theta_0\right)\right)\exp(\mathbf{i}.\xi' X_j)$$
$$=\sqrt{n}(\widehat{\theta}-\theta_0)'b(\tau,\xi)+o_p(1) \qquad \textbf{(A.1)}$$

uniformly on $\Upsilon\times\Xi$. Equation (8) of Lemma 1 follows now straightforwardly from Assumptions 1 and 2 and (A.1), and (11) follows trivially from (8). ■

Proof of Theorem 1. Let $Z_n(\beta)$ be an empirical process on a compact subset $\mathbf{B}$ of a euclidean space. Then $Z_n\Rightarrow Z$ if Z_n is tight, and the finite distributions of Z_n converge. The latter means that for arbitrary $\beta_1,\beta_2,\ldots,\beta_M$ in $\mathbf{B}$,

$$(Z_n(\beta_1),Z_n(\beta_2),\ldots,Z_n(\beta_M))\xrightarrow{d}(Z(\beta_1),Z(\beta_2),\ldots,Z(\beta_M)).$$

In the case of the empirical process (9) this condition follows straightforwardly from the central limit theorem. The tightness concept is a generalization of the stochastic boundedness concept for sequences of random variables: Let $Z_n\in\mathcal{M}$ for all $n\geq 1$, where $\mathcal{M}$ is a metric space of functions on $\mathbf{B}$. For each $\varepsilon\in(0,1)$ there exists a compact set $K\subset\mathcal{M}$ such that $\inf_{n\geq 1}\Pr[Z_n\in K]>1-\varepsilon$.

According to Billingsley (1968, Thm. 8.2), two conditions deliver the tightness of Z_n:

(a) For each $\eta>0$ and each $\beta\in\mathbf{B}$ there exists a $\delta>0$ such that

$$\sup_{n\geq 1}\Pr[|Z_n(\beta)|>\delta]\leq\eta.$$

(b) For each $\eta > 0$ and $\delta > 0$ there exists an $\varepsilon > 0$ such that

$$\sup_{n\geq 1} \Pr\left[\sup_{\|\beta_1-\beta_2\|<\varepsilon} |Z_n(\beta_1) - Z_n(\beta_2)| \geq \delta \right] \leq \eta.$$

Condition (a) is a pointwise stochastic boundedness condition, which holds if for each $\beta \in \mathbf{B}$, $Z_n(\beta)$ converges in distribution. Condition (b) is also known as the stochastic equicontinuity condition, which is the difficult part of the tightness proof.

As a result of Lemma 1 it suffices to prove $\widetilde{Z}_n(\tau,\xi) \Rightarrow Z(\tau,\xi)$, where $\widetilde{Z}_n(\tau,\xi)$ is defined by (9). Thus, in our case $\beta = (\tau,\xi)$, $\mathbf{B} = \Upsilon \times \Xi$, and $Z_n(\beta) = \widetilde{Z}_n(\tau,\xi)$.

To prove the tightness of $\widetilde{Z}_n$, note that $\widetilde{Z}_n(\tau,\xi) = \widetilde{Z}_{1,n}(\tau,\xi) - \widetilde{Z}_{2,n}(\tau,\xi)$, where

$$\widetilde{Z}_{1,n}(\tau,\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\exp(\mathbf{i}.\tau' Y_j) - \mathrm{E}[\exp(\mathbf{i}.\tau' Y_j)|X_j] \right) \exp(\mathbf{i}.\xi' X_j),$$

$$\widetilde{Z}_{2,n}(\tau,\xi) = b(\tau,\xi)' A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Delta\ell\,(Y|X;\theta_0).$$

Because under H_0, $A^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \Delta\ell\,(Y,X;\theta_0) \xrightarrow{d} N_p\left[0, A^{-1}\right]$, and $b(\tau,\xi)$ is continuous, it follows straightforwardly that $\widetilde{Z}_{2,n}(\tau,\xi)$ is tight. Therefore, $\widetilde{Z}_n(\tau,\xi)$ is tight if $\widetilde{Z}_{1,n}(\tau,\xi)$ is tight.

Because by the central limit theorem, $\widetilde{Z}_{1,n}(\tau,\xi)$ converges in distribution, pointwise in (τ,ξ), to a complex-valued random variable $Z_1(\tau,\xi)$, e.g., condition (a) is satisfied.

To verify condition (b), observe that if the Y_j's and X_j's are bounded we can write

$$\begin{aligned}
\widetilde{Z}_{1,n}(\tau,\xi) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{r=0}^{\infty} \frac{\mathbf{i}^r}{r!} \left((\tau' Y_j)^r - \mathrm{E}[(\tau' Y_j)^r|X_j] \right) \left(\sum_{s=0}^{\infty} \frac{\mathbf{i}^s}{s!} (\xi' X_j)^s \right) \\
&= \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{\mathbf{i}^{r+s}}{r!s!} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left((\tau' Y_j)^r - \mathrm{E}[(\tau' Y_j)^r|X_j] \right) (\xi' X_j)^s.
\end{aligned}$$

To keep the notation tractable, let us focus on the case $m = k = 2$. With $\tau = (\tau_1, \tau_2)'$ and $Y_j = (Y_{1,j}, Y_{2,j})'$ we have

$$(\tau' Y_j)^r = \sum_{b=0}^{r} \binom{r}{b} \tau_1^b \tau_2^{r-b} Y_{1,j}^b Y_{2,j}^{r-b},$$

and, similarly, with $\xi = (\xi_1, \xi_2)'$ and $X_j = (X_{1,j}, X_{2,j})'$ we have

$$(\xi' X_j)^s = \sum_{q=0}^{s} \binom{s}{q} \xi_1^q \xi_2^{s-q} X_{1,j}^q X_{2,j}^{s-q}.$$

Then

$$\begin{aligned}
\widetilde{Z}_{1,n}(\tau,\xi) &= \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{\mathbf{i}^{r+s}}{r!s!} \sum_{b=0}^{r} \binom{r}{b} \tau_1^b \tau_2^{r-b} \sum_{q=0}^{s} \binom{s}{q} \xi_1^q \xi_2^{s-q} \\
&\quad \times \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(Y_{1,j}^b Y_{2,j}^{r-b} - \mathrm{E}[Y_{1,j}^b Y_{2,j}^{r-b}|X_j] \right) X_{1,j}^q X_{2,j}^{s-q}.
\end{aligned}$$

Hence, with $\rho = (\rho_1, \rho_2)'$ and $\zeta = (\zeta_1, \zeta_2)'$ we have

$$\begin{aligned}
&\sup_{\|(\tau,\xi)'-(\rho,\zeta)'\|<\varepsilon} \left|\widetilde{Z}_{1,n}(\tau,\xi) - \widetilde{Z}_{1,n}(\rho,\zeta)\right| \\
&\quad \leq \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \frac{1}{r!s!} \sum_{b=0}^{r}\binom{r}{b}\sum_{q=0}^{s}\binom{s}{q} \sup_{\|(\tau,\xi)'-(\rho,\zeta)'\|<\varepsilon} \left|\tau_1^b\tau_2^{r-b}\xi_1^q\xi_2^{s-q} - \rho_1^b\rho_2^{r-b}\zeta_1^q\zeta_2^{s-q}\right| \\
&\quad \times \left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(Y_{1,j}^b Y_{2,j}^{r-b} - \mathrm{E}[Y_{1,j}^b Y_{2,j}^{r-b}|X_j]\right) X_{1,j}^q X_{2,j}^{s-q}\right|.
\end{aligned}$$

Next, let $\Upsilon \times \Xi = [-c, c]^4$ with $c > 1$. Then it is not too hard to verify that for $\|(\tau,\xi)' - (\rho,\zeta)'\| < \varepsilon < 1$,

$$\begin{aligned}
&\left|\tau_1^b\tau_2^{r-b}\xi_1^q\xi_2^{s-q} - \rho_1^b\rho_2^{r-b}\zeta_1^q\zeta_2^{s-q}\right| \\
&\quad \leq c^{r+s}\left(\left|\tau_1^b - \rho_1^b\right| + \left|\tau_2^{r-b} - \rho_2^{r-b}\right| + \left|\xi_1^q - \zeta_1^q\right| + \left|\xi_2^{s-q} - \zeta_2^{s-q}\right|\right) \\
&\quad \leq \varepsilon.c^{r+s}\left(c^b + c^{r-b} + c^q + c^{s-q}\right) \leq 4\varepsilon\,(2c)^{r+s}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\left(\sup_{\|(\tau,\xi)'-(\rho,\zeta)'\|<\varepsilon} \left|\widetilde{Z}_{1,n}(\tau,\xi) - \widetilde{Z}_{1,n}(\rho,\zeta)\right|\right)^2 \\
&\quad \leq 16\varepsilon^2\left(\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{1}{r!s!}(2c)^{r+s}\sum_{b=0}^{r}\binom{r}{b}\sum_{q=0}^{s}\binom{s}{q}\right. \\
&\qquad\qquad \left.\times\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(Y_{1,j}^b Y_{2,j}^{r-b} - \mathrm{E}[Y_{1,j}^b Y_{2,j}^{r-b}|X_j]\right) X_{1,j}^q X_{2,j}^{s-q}\right|\right)^2 \\
&\quad \leq 16\varepsilon^3\left(\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{1}{r!s!}\sum_{b=0}^{r}\binom{r}{b}\sum_{q=0}^{s}\binom{s}{q}(2c)^{r+s}\right) \\
&\qquad \times\left\{\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{1}{r!s!}(2c)^{r+s}\sum_{b=0}^{r}\binom{r}{b}\sum_{q=0}^{s}\binom{s}{q}\right. \\
&\qquad\quad \left.\times\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(Y_{1,j}^b Y_{2,j}^{r-b} - \mathrm{E}[Y_{1,j}^b Y_{2,j}^{r-b}|X_j]\right) X_{1,j}^q X_{2,j}^{s-q}\right)^2\right\},
\end{aligned}$$

and thus

$$\mathrm{E}\left[\left(\sup_{\|(\tau,\xi)'-(\rho,\zeta)'\|<\varepsilon}\left|\widetilde{Z}_{1,n}(\tau,\xi)-\widetilde{Z}_{1,n}(\rho,\zeta)\right|\right)^2\right]$$

$$\leq 16\varepsilon^2\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{1}{r!s!}(4c)^{r+s}$$

$$\times\left(\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{1}{r!s!}(2c)^{r+s}\sum_{b=0}^{r}\binom{r}{b}\sum_{q=0}^{s}\binom{s}{q}E\left(Y_{1,1}^{b}Y_{2,1}^{r-b}X_{1,1}^{q}X_{2,1}^{s-q}\right)^2\right)$$

$$=16\varepsilon^2\exp(8c)\,\mathrm{E}\left(\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{1}{r!s!}(2c)^{r+s}\|Y\|^{2r}\|X\|^{2s}\right)$$

$$=16\varepsilon^2\exp(8c)\,\mathrm{E}\left[\exp\left(2c\left(\|Y\|^2+\|X\|^2\right)\right)\right],$$

where the inequality is due to Schwarz inequality. Therefore, a sufficient condition for tightness is that the moment generating function of $\|Y\|^2+\|X\|^2$ is everywhere finite, which is of course the case if Y and X are bounded.

This completes the proof of part (12) of Theorem 1. The proof of part (14) is easy and therefore omitted. ■

Proof of Lemma 2. Assume that the Y_j's and X_j's are scalar random variables and that X_j is already bounded. Then

$$\widehat{Z}_n(\tau,\xi)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(\mathbf{i}.\tau\Psi\left(\sigma_n^{-1}(Y_j-\mu_n)\right)\right)\right.$$

$$\left.-\int\exp\left(\mathbf{i}.\tau\Psi\left(\sigma_n^{-1}(y-\mu_n)\right)\right)dF(y|X_j,\widehat{\theta})\right)\exp\left(\mathbf{i}.\xi X_j\right),$$

$$Z_n(\tau,\xi)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(\mathbf{i}.\tau\Psi\left(\sigma^{-1}(Y_j-\mu)\right)\right)\right.$$

$$\left.-\int\exp\left(\mathbf{i}.\tau\Psi\left(\sigma^{-1}(y-\mu)\right)\right)dF(y|X_j,\widehat{\theta})\right)\exp\left(\mathbf{i}.\xi X_j\right),$$

where $\Psi(x)=\arctan(x)$, $(\tau,\xi)\in[-\overline{\tau},\overline{\tau}]\times[-\overline{\xi},\overline{\xi}]$, and $\overline{\tau},\overline{\xi}\in(0,\infty)$, and, by Assumption 3,

$$\sqrt{n}(\sigma_n-\sigma)=O_p(1),\qquad \sqrt{n}(\mu_n-\mu)=O_p(1). \tag{A.2}$$

Note that

$$\Psi'(x)=\frac{1}{1+x^2},\qquad \Psi''(x)=\frac{-2x}{(1+x^2)^2}. \tag{A.3}$$

Hence

$$\frac{\partial \Phi\left(\sigma^{-1}\left(Y_j-\mu\right)\right)}{\partial(\sigma,\mu)} = -\Psi'\left(\sigma^{-1}\left(Y_j-\mu\right)\right)\begin{pmatrix}\sigma^{-2}Y_j\\ \sigma^{-1}\end{pmatrix} = \Delta_1\left(Y_j|\sigma,\mu\right),$$

$$\frac{\partial^2 \Psi\left(\sigma^{-1}\left(Y_j-\mu\right)\right)}{\partial(\sigma,\mu)\partial(\sigma,\mu)'} = \Psi''\left(\sigma^{-1}\left(Y_j-\mu\right)\right)\begin{pmatrix}\sigma^{-2}Y_j\\ \sigma^{-1}\end{pmatrix}\left(\sigma^{-2}Y_j,\sigma^{-1}\right)$$

$$+\Psi'\left(\sigma^{-1}\left(Y_j-\mu\right)\right)\begin{pmatrix}2\sigma^{-3}Y_j & 0\\ \sigma^{-2} & 0\end{pmatrix} = \Delta_2\left(Y_j|\sigma,\mu\right),$$

say. Moreover, note that because of (A.3), $\Delta_1(y|\sigma,\mu)$ and $\Delta_2(y|\sigma,\mu)$ are bounded in y.

Now by Taylor's theorem,

$$\begin{aligned}
&\sqrt{n}\left(\Psi\left(\sigma_n Y_j-\mu_n\right)-\Psi\left(\sigma Y_j-\mu\right)\right)\\
&\quad=\sqrt{n}\left(\sigma_n-\sigma,\mu_n-\mu\right)\Delta_1\left(Y_j|\sigma,\mu\right)\\
&\qquad+\frac{1}{2}\sqrt{n}\left(\sigma_n-\sigma,\mu_n-\mu\right)\Delta_2\left(Y_j|\sigma+\lambda_j\left(\left(\sigma_n-\sigma\right)\right),\mu+\lambda_j\left(\mu_n-\mu\right)\right)\\
&\qquad\times\begin{pmatrix}\sigma_n-\sigma\\ \mu_n-\mu\end{pmatrix}\\
&\quad=\sqrt{n}\left(\sigma_n-\sigma,\mu_n-\mu\right)\Delta_1\left(Y_j|\sigma,\mu\right)+O_p\left(1/\sqrt{n}\right),
\end{aligned}$$

and, similarly,

$$\begin{aligned}
&\sqrt{n}\left(\exp\left(\mathbf{i}.\tau\Phi\left(\sigma_n Y_j-\mu_n\right)\right)-\exp\left(\mathbf{i}.\tau\Phi\left(\sigma Y_j-\mu\right)\right)\right)\\
&\quad=\sqrt{n}\left(\sigma_n-\sigma,\mu_n-\mu\right)\Delta_1\left(Y_j|\sigma,\mu\right)\exp\left(\mathbf{i}.\tau\Phi\left(\sigma Y_j-\mu\right)\right)\\
&\qquad+O_p\left(1/\sqrt{n}\right),
\end{aligned}$$

where the $O_p\left(1/\sqrt{n}\right)$ terms are uniform in $j=1,\ldots,n$ and $\tau\in[-\overline{\tau},\overline{\tau}]$. Then

$$\begin{aligned}
&\widehat{Z}_n(\tau,\xi)-Z_n(\tau,\xi)\\
&\quad=\frac{1}{n}\sum_{j=1}^{n}\sqrt{n}\left(\exp\left(\mathbf{i}.\tau\Phi\left(\sigma_n^{-1}\left(Y_j-\mu_n\right)\right)\right)-\exp\left(\mathbf{i}.\tau\Phi\left(\sigma^{-1}\left(Y_j-\mu\right)\right)\right)\right)\\
&\qquad\times\exp\left(i\xi X_j\right)\\
&\qquad-\sqrt{n}\int\left(\exp\left(\mathbf{i}.\tau\Phi\left(\sigma_n^{-1}\left(y-\mu_n\right)\right)\right)-\exp\left(\mathbf{i}.\tau\Phi\left(\sigma^{-1}\left(y-\mu\right)\right)\right)\right)\\
&\qquad\times dF(y|X_j,\widehat{\theta})
\end{aligned}$$

$$= \sqrt{n}(\sigma_n - \sigma, \mu_n - \mu)\frac{1}{n}\sum_{j=1}^{n} \Delta_1\left(Y_j|\sigma,\mu\right)\exp\left(\mathbf{i}.\tau\Phi\left(\sigma\left(Y_j - \mu\right)\right)\right)$$

$$\times\exp\left(\mathbf{i}.\xi X_j\right)$$

$$-\sqrt{n}(\sigma_n - \sigma, \mu_n - \mu)\frac{1}{n}\sum_{j=1}^{n}\int \Delta_1(y|\sigma,\mu)\exp\left(\mathbf{i}.\tau\Phi\left(\sigma^{-1}(y-\mu)\right)\right)$$

$$\times dF(y|X_j,\widehat{\theta})\exp\left(\mathbf{i}.\xi X_j\right) + O_p\left(1/\sqrt{n}\right)$$

uniformly in $(\tau,\xi) \in [-\overline{\tau},\overline{\tau}] \times [-\overline{\xi},\overline{\xi}]$. Because $\widehat{\theta} \xrightarrow{p} \theta_0$ it follows now that

$$\frac{1}{n}\sum_{j=1}^{n} \Delta_1\left(Y_j|\sigma,\mu\right)\exp\left(\mathbf{i}.\tau\Phi\left(\sigma^{-1}\left(Y_j-\mu\right)\right)\right)\exp\left(\mathbf{i}.\xi X_j\right)$$

$$\xrightarrow{p} \mathrm{E}\left[\int \Delta_1(y|\sigma,\mu)\exp\left(\mathbf{i}.\tau\Phi\left(\sigma^{-1}(y-\mu)\right)\right)dF(y|X,\theta_0)\exp(\mathbf{i}.\xi X)\right],$$

$$\frac{1}{n}\sum_{j=1}^{n}\int \Delta_1(y|\sigma,\mu)\exp\left(\mathbf{i}.\tau\Phi\left(\sigma^{-1}(y-\mu)\right)\right)dF(y|X_j,\widehat{\theta})\exp\left(\mathbf{i}.\xi X_j\right)$$

$$\xrightarrow{p} \mathrm{E}\left[\int \Delta_1(y|\sigma,\mu)\exp\left(\mathbf{i}.\tau\Phi\left(\sigma^{-1}(y-\mu)\right)\right)dF(y|X,\theta_0)\exp(\mathbf{i}.\xi X)\right],$$

uniformly in $(\tau,\xi) \in [-\overline{\tau},\overline{\tau}] \times [-\overline{\xi},\overline{\xi}]$. Hence

$$\sup_{(\tau,\xi)\in[-\overline{\tau},\overline{\tau}]\times[-\overline{\xi},\overline{\xi}]}\left|\widetilde{Z}_n(\tau,\xi) - Z_n(\tau,\xi)\right| = o_p(1).$$

The proof of the general case is now easy, and so is the proof of the result under the alternative hypothesis. ■

Proof of Lemma 4. Using the trivial equality $(a+\mathbf{i}.b)(c-\mathbf{i}.d) = (a.c+b.d) - \mathbf{i}.(a.d-b.c)$, it is easy to verify that the orthonormality property $\int \psi_{j_1}(\beta)\overline{\psi_{j_2}(\beta)}\mu(d\beta) = \mathbf{1}(j_1 = j_2)$ is equivalent to

$$\int \mathrm{Re}\left(\psi_{j_1}(\beta)\right)\mathrm{Re}\left(\psi_{j_2}(\beta)\right)\mu(d\beta) + \int \mathrm{Im}\left(\psi_{j_1}(\beta)\right)\mathrm{Im}\left(\psi_{j_2}(\beta)\right)\mu(d\beta) = \mathbf{1}(j_1 = j_2),$$

$$\int \mathrm{Re}\left(\psi_{j_1}(\beta)\right)\mathrm{Im}\left(\psi_{j_2}(\beta)\right)\mu(d\beta) - \int \mathrm{Im}\left(\psi_{j_1}(\beta)\right)\mathrm{Re}\left(\psi_{j_2}(\beta)\right)\mu(d\beta) = 0.$$

Consequently, denoting

$$Q_j(\beta) = \begin{pmatrix} \mathrm{Re}\left(\psi_j(\beta)\right) & \mathrm{Im}\left(\psi_j(\beta)\right) \\ \mathrm{Im}\left(\psi_j(\beta)\right) & -\mathrm{Re}\left(\psi_j(\beta)\right) \end{pmatrix}$$

it follows that

$$\int Q_{j_1}(\beta)Q_{j_2}(\beta)\mu(d\beta) = \mathbf{1}(j_1 = j_2).I_2. \quad \textbf{(A.4)}$$

Similarly, $g_j = \int Z(\beta)\overline{\psi_j(\beta)}\mu(d\beta)$ is equivalent to

$$\mathrm{Re}\left(g_j\right) = \int \mathrm{Re}\left(Z(\beta)\right)\mathrm{Re}\left(\psi_j(\beta)\right)\mu(d\beta) + \int \mathrm{Im}\left(Z(\beta)\right)\mathrm{Im}\left(\psi_j(\beta)\right)\mu(d\beta),$$

$$\mathrm{Im}\left(g_j\right) = \int \mathrm{Re}\left(Z(\beta)\right)\mathrm{Im}\left(\psi_j(\beta)\right)\mu(d\beta) - \int \mathrm{Im}\left(Z(\beta)\right)\mathrm{Re}\left(\psi_j(\beta)\right)\mu(d\beta).$$

Hence, denoting

$$G_{2,j} = \begin{pmatrix} \mathrm{Re}\left(g_j\right) \\ \mathrm{Im}\left(g_j\right) \end{pmatrix}, \qquad Z_2(\beta) = \begin{pmatrix} \mathrm{Re}\left(Z(\beta)\right) \\ \mathrm{Im}\left(Z(\beta)\right) \end{pmatrix}$$

we can write

$$G_{2,j} = \int Q_j(\beta) Z_2(\beta)\mu(d\beta). \tag{A.5}$$

Along the same lines it is straightforward to verify that the result of Mercer's theorem, i.e., $\mathrm{E}\left[Z(\beta_1)\overline{Z(\beta_2)}\right] = \sum_{k=1}^{\infty} \lambda_k \psi_k(\beta_1)\overline{\psi_k(\beta_2)}$, now reads

$$\mathrm{E}\left[Z_2(\beta_1)Z_2(\beta_2)'\right] = \sum_{k=1}^{\infty} \lambda_k Q_k(\beta_1) Q_k(\beta_2). \tag{A.6}$$

It follows now from (A.4)–(A.6) that

$$\begin{aligned}
\mathrm{E}\left[G_{2,j_1} G_{2,j_2}'\right] &= \int\int Q_{j_1}(\beta_1)\mathrm{E}\left[Z_2(\beta_1)Z_2(\beta_2)'\right] Q_{j_2}(\beta_2)\mu(d\beta_1)\mu(d\beta_2) \\
&= \int\int Q_{j_1}(\beta_1)\left(\sum_{k=1}^{\infty} \lambda_k Q_k(\beta_1)Q_k(\beta_2)\right) Q_{j_2}(\beta_2)\mu(d\beta_1)\mu(d\beta_2) \\
&= \sum_{k=1}^{\infty} \lambda_k \left(\int Q_{j_1}(\beta_1)Q_k(\beta_1)\mu(d\beta_1)\right)\left(\int Q_k(\beta_2)Q_{j_2}(\beta_2)\mu(d\beta_2)\right) \\
&= \sum_{k=1}^{\infty} \lambda_k \mathbf{1}(j_1 = k)\mathbf{1}(j_2 = k) I_2 \\
&= \begin{cases} \lambda_{j_1} I_2 & \text{if } j_1 = j_2, \\ O & \text{if } j_1 \neq j_2. \end{cases}
\end{aligned} \tag{A.7}$$

Because the random vectors $G_{2,j}$ depend linearly on a common bivariate zero-mean Gaussian process Z_2, they are jointly normally distributed, with zero expectation vectors. Therefore, the result (A.7) implies that the random vectors $G_{2,j}$ are independently $N_2[0, \lambda_j I_2]$ distributed. Consequently, we can write

$$G_{2,j} = \sqrt{\lambda_j} e_j,$$

where the e_j's are independently $N_2\left[0, I_2\right]$ distributed.

Finally, note that

$$\int |Z(\beta)|^2 \mu(d\beta) = \int Z(\beta)\overline{Z(\beta)}\mu(d\beta)$$

$$= \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} g_{j_1} \int \psi_{j_1}(\beta)\overline{\psi}_{j_2}(\beta)\mu(d\beta)\overline{g}_{j_2}$$

$$= \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} g_{j_1}\overline{g}_{j_2}\mathbf{1}(j_1 = j_2)$$

$$= \sum_{j=1}^{\infty} g_j\overline{g}_j = \sum_{j=1}^{\infty} |g_j|^2$$

$$= \sum_{j=1}^{\infty} \left(\left(\mathrm{Re}(g_j)\right)^2 + \left(\mathrm{Im}(g_j)\right)^2\right),$$

which proves the last result in Lemma 4. ■

Proof of Theorem 3. In view of Lemma 1 it suffices to show that

$$\widetilde{T}_n(c) = \frac{1}{(2c)^{m+k}} \int_{\Upsilon(c)\times\Xi(c)} |\widetilde{Z}_n(\tau,\xi)|^2 d\tau d\xi$$

is tight on $[\underline{c}, \overline{c}]$. To show this, let for $\underline{c} \le c_1 < c_2 \le \overline{c}$,

$$\Pi(c_1, c_2) = [-c_2, c_2]^{m+k} \setminus [-c_1, c_1]^{m+k}.$$

Then

$$\sup_{|c_2-c_1|<\delta} \left|\widetilde{T}_n(c_2) - \widetilde{T}_n(c_1)\right|$$

$$= \frac{1}{(2\underline{c})^{m+k}} \sup_{|c_2-c_1|<\delta} \int_{\Pi(c_1,c_2)} |\widetilde{Z}_n(\tau,\xi)|^2 d\tau d\xi$$

$$+ \sup_{|c_2-c_1|<\delta} \left(\frac{1}{(2c_1)^{m+k}} - \frac{1}{(2c_2)^{m+k}}\right) \int_{\Upsilon(\overline{c})\times\Xi(\overline{c})} |\widetilde{Z}_n(\tau,\xi)|^2 d\tau d\xi$$

$$\xrightarrow{d} \frac{1}{(2\underline{c})^{m+k}} \sup_{|c_2-c_1|<\delta} \int_{\Pi(c_1,c_2)} |Z(\tau,\xi)|^2 d\tau d\xi$$

$$+ \sup_{|c_2-c_1|<\delta} \left(\frac{1}{(2c_1)^{m+k}} - \frac{1}{(2c_2)^{m+k}}\right) \int_{\Upsilon(\overline{c})\times\Xi(\overline{c})} |Z(\tau,\xi)|^2 d\tau d\xi.$$

Hence for $0 < \delta < 1$,

$$\sup_{|c_2-c_1|<\delta} \left|\widetilde{T}_n(c_2) - \widetilde{T}_n(c_1)\right| = O_p(\delta)$$

because the Lebesgue measure of $\Pi(c_1, c_2)$ for $|c_2 - c_1| < \delta$ is $(2\delta)^{m+k} < 2^{m+k}\delta$ and

$$\sup_{|c_2-c_1|<\delta}\left(\frac{1}{(2c_1)^{m+k}} - \frac{1}{(2c_2)^{m+k}}\right) \le \sup_{|c_2-c_1|<\delta} \frac{(2c_2)^{m+k}-(2c_1)^{m+k}}{(2\underline{c})^{2m+2k}} = O(\delta).$$

This proves the tightness of $\widetilde{T}_n(c)$. See Theorem 8.2 in Billingsley (1968). ■

Proof of Theorem 5. It follows similar to Lemma 1 that

$$\sup_{(\tau,\xi)\in\Upsilon\times\Xi} |\widetilde{Z}_n(\tau,\xi) - \widetilde{Z}_n^{(s)}(\tau,\xi)| = o_p(1). \quad \textbf{(A.8)}$$

Denote by $\mathcal{D}$ the σ-algebra generated by $\left\{(Y_j, X_j)\right\}_{j=1}^{\infty}$. Because $\tilde{Y}_j$ is generated according to the estimated model $F(y|X_j,\hat{\theta})$, it follows similar to Theorem 1 that under both H_0 and H_1, $\widetilde{Z}_n^{(s)} \Rightarrow Z_s$ conditional on $\mathcal{D}$, where $Z_s(\tau,\xi)$ is a zero-mean Gaussian process with covariance function

$$\begin{aligned}
&\Gamma_s((\tau_1,\xi_1),(\tau_2,\xi_2)) \\
&\quad = p\lim_{n\to\infty} \mathrm{E}\Big[\Big(\varphi\left(\tau_1-\tau_2|X;\widehat{\theta}\right) - \varphi\left(\tau_1|X;\widehat{\theta}\right)\varphi\left(-\tau_2|X;\widehat{\theta}\right)\Big) \\
&\qquad\qquad\qquad \times \exp(\mathbf{i}(\xi_1-\xi_2)'X)\Big|\mathcal{D}\Big] \\
&\quad = \mathrm{E}\left[(\varphi(\tau_1-\tau_2|X;\theta_0) - \varphi(\tau_1|X;\theta_0)\varphi(-\tau_2|X;\theta_0))\exp(\mathbf{i}(\xi_1-\xi_2)'X)\right],
\end{aligned}$$

where $\varphi(\tau|X;\theta)$ is the conditional characteristic function of $F(y|X_1,\theta)$ (cf. (7)). Because Γ_s does not depend on $\mathcal{D}$ it follows now that unconditionally,

$$\widetilde{Z}_n^{(s)} \Rightarrow Z_s \quad \text{under } H_0 \text{ and } H_1. \quad \textbf{(A.9)}$$

The independence of $Z(\tau,\xi)$ and $Z_s(\tau,\xi)$ follows from

$$\mathrm{E}\left[Z(\tau_1,\xi_1)\overline{Z_s(\tau_2,\xi_2)}\right] = 0,$$

as is not hard to verify. Hence by (A.8)–(A.9), and the continuous mapping theorem,

$$\widehat{T}_n^{(s)} \xrightarrow{d} \int_{\Upsilon\times\Xi} |Z(\tau,\xi) - Z_s(\tau,\xi)|^2 d\mu(\tau,\xi).$$

The result under H_1 is easy to verify from (A.9) and Theorem 1. ■