

Integrated Conditional Moment Tests for Parametric Conditional Distributions*

Herman J. Bierens^a and Li Wang^b

^aDepartment of Economics and CAPCP[†]

Pennsylvania State University

University Park, PA 16802

^bDepartment of Public Health Sciences

PennState College of Medicine

Hershey, PA 17033

April 12, 2009

Abstract

In this paper we propose consistent Integrated Conditional Moment (ICM) 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 (SICM) test is proposed. As an empirical application we test the validity of a few common health economic count data models

*A previous version of this paper was presented by Herman Bierens at the "Symposium on Nonparametric Testing in Econometrics", 2008, Indiana University, and by Li Wang at the "Health Econometrics Workshop", 2008, University of Milan-Bicocca, Italy. The helpful comments of Juan Carlos Escanciano and Matteo Manera 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/>).

Key words: Specification tests, Integrated conditional moments, Empirical characteristic functions, Bootstrap
JEL classifications: C12, C15

1 Introduction

Quoting Hausman (1978), “Specification tests form one of the most important areas for research in econometrics”, because the correct specification of a model constitutes a fundamental assumption for its estimation and inference. Hausman’s specification test is based on the difference between an efficient estimator under the null and a non-efficient estimator. White (1981) utilized a specification-robust estimator of nonlinear regression models to test conditional mean specifications.

Newey (1985) proposed the conditional moment (CM) test and suggested that both the Hausman (1978) and the White (1981) methods can be viewed as special cases of a CM test. The idea behind the CM test is that a correct model specification implies that certain conditional moments are zero, which can be converted to unconditional moment restrictions by multiplying these conditional moments by instrumental variables. The sample counterparts of these unconditional moments are the basis for a CM test.

These tests are not consistent because they only employ a finite number of moment restrictions implied by the model. Bierens (1982) and Holly (1983) observed this inconsistency for the Hausman test. In general, for tests based on a finite number of moment restrictions one can always construct a data-generating process for which the null hypothesis is false but the moment restrictions involved hold.

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. Because the asymptotic properties of our ICM test are

closely related to those of the ICM tests for conditional expectation models, we will review the latter and related literature first.

The first consistent model specification test, for nonlinear regression models, was proposed by Bierens (1982). This test is based on the fact that an integrable Borel measurable function is uniquely identified by its Fourier transform. Therefore, with U the error term of a nonlinear regression model with vector $X \in \mathbb{R}^k$ of explanatory variables, $\Pr[E[(U|X) = 0] = 1]$ if and only if $E[U \exp(i.\xi'X)] = 0$ for all $\xi \in \mathbb{R}^k$. Moreover, if X is bounded then it suffices to check the latter equality for all ξ in an arbitrary hypercube Ξ around the origin of \mathbb{R}^k . If X is not bounded then without loss of generality one may replace X in the Fourier transform involved with a bounded one-to-one transformation $\Phi(X)$. Thus, $\Pr[E[(U|X) = 0] = 1]$ if and only if $\int_{\Xi} |E[U.w(\xi'\Phi(X))]|^2 d\mu(\xi)$, where $w(x) = \exp(i.x)$ and μ is the uniform probability measure on Ξ . This suggests that the functional form of the regression model involved can be tested consistently on the basis of the integrated conditional moment (ICM) statistic of the form $\hat{T}_n = \int_{\Xi} |\hat{Z}_n(\xi)|^2 d\mu(\xi)$, where $\hat{Z}_n(\xi) = (1/\sqrt{n}) \sum_{j=1}^n [\hat{U}_j w(\xi'\Phi(X_j))]$ with the \hat{U}_j 's the regression residuals. Bierens (1982) showed that if the model is correctly specified then $\hat{T}_n \rightarrow T$ in distribution,¹ whereas $\hat{T}_n \rightarrow \infty$ in probability if the model is misspecified. The ICM test was generalized to time series models by Bierens (1984) and De Jong (1996). Bierens (1990) showed for the case $w(x) = \exp(x)$ that under the null hypothesis, \hat{Z}_n converges weakly to a zero mean Gaussian process Z , so that $T = \int_{\Xi} |Z(\xi)|^2 d\mu(\xi)$. Bierens and Ploberger (1997) proved for more general real-valued weight functions $w(x)$ that the limiting null distribution takes the form $T = \sum_{m=1}^{\infty} \lambda_m \varepsilon_m^2$, where the ε_m 's are i.i.d. $N(0, 1)$ and the λ_m 's are the eigenvalues of the covariance function $E[Z(t_1)Z(t_2)]$. They also showed that the ICM test has non-trivial \sqrt{n} local power and is admissible. Very recently the local power properties of the ICM test has been analyzed further by Escanciano (2009).

Stinchcombe and White (1998) have shown that the ICM test is consistent for a wide range of non-polynomial analytical weight functions w , for which they cast the name “totally revealing”, and Boning and Sowell (1999) showed that the uniform probability measure μ yields the best ICM test according to the weighted average power criterion considered by Andrews and Ploberger

¹Bierens (1982) was only able to derive an expression for $E[T]$, but not for T itself. Therefore, he proposed to derive upper bounds of the critical values via Chebyshev's inequality for first moments: $\Pr[T > t] \leq E[T]/t$ for $t > 0$.

(1994).

A related specification test for regression models has been proposed by Stute (1997), based on the fact that $\Pr[E(U|X) = 0] = 1$ is equivalent to $E[U\mathbf{1}(X \leq x)] \equiv 0$ for all conformable vectors x , where $\mathbf{1}(\cdot)$ denotes the indicator function. Recently, Escanciano (2006) proposed a combined Bierens-Stute type ICM test.

Another strand of literature on model specification testing is based on a comparison of parametric functional forms with corresponding nonparametric or semi-nonparametric estimates. See, for example, 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. However, these tests have only non-trivial power against local alternatives that approach the null at a slower rate than $1/\sqrt{n}$.

Although at first sight the ICM test and the nonparametric kernel regression based tests for nonlinear regression models seem fundamentally different, they are related to each other in an interesting way. Fan and Li (2000) have shown that the ICM test can be viewed as a special case of the kernel-based tests but with a fixed bandwidth.

The literature on specification testing of conditional distribution models is rather limited. White's (1982) test is based on comparison of two different expressions of the Fisher information matrix, which should be equal if the conditional distribution is correctly specified. However, this test is not consistent. Andrews (1988) extended the Pearson's Chi-square test to a test for parametric conditional distributions. This test is based on partitioning the dependent and explanatory variables in cells, and then comparing the frequencies involved with the frequencies implied by the model. However, it is unknown what the best way is to choose these cells. See Justel et. al. (1997). 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) compares 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

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

where $F(y|X_j, \hat{\theta})$ is the estimated conditional distribution model. Since the asymptotic null distribution is case-dependent, the critical values have to be derived via a bootstrap method. This test is consistent, and has non-trivial 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 for quite a large sample size n . This appears to be the case in our empirical application in Section 4. This problem does not happen with our 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 (1951) information criterion. Thus, this test is only applicable to absolutely continuous conditional distribution models. Zheng's test has non-trivial local power, but only against local alternatives that approach the null at a slower rate than $1/\sqrt{n}$. Moreover, this rate decreases with the number of covariates, so that Zheng's test suffers from the curse of dimensionality.

Bai (2003) proposed a test for the validity of absolutely continuous conditional distribution models based on the well-known fact that plugging in an absolutely continuous distributed random variable in its own conditional distribution function yields a uniformly $[0, 1]$ distributed random variable. Although Bai's test aims to test conditional time series distribution models, it applies to cross-section models as well. However, Bai's (2003) test is not consistent. See the Appendix for a counter example.

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 in order 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 notations. Convergence in distribution will be denoted by \xrightarrow{d} and convergence in probability by \xrightarrow{p} or $p\lim_{n \rightarrow \infty}$. The wiggle \sim stands for "is distributed as". The double-arrow \Rightarrow indicates weak convergence of random functions. See for example Billingsley (1968) for the latter notion. Finally, a bar over a complex variable, vector or function denotes the complex conjugate, i.e., if $z = a + i.b$ then $\bar{z} = a - i.b$.

2 The ICM test for conditional distributions

We will develop our test for cross-section models only. Since 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. Of course, if the model is misspecified then maximum likelihood 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 holds. See White (1982, 1994).

Assumption 1. *We observe a random sample $(Y_1, X_1), \dots, (Y_n, X_n)$ 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 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 a.s. twice continuously differentiable in θ . The QML estimator $\hat{\theta} = \arg \max_{\theta \in \Theta} \ln L_n(\theta)$ converges in probability to $\theta_0 = \arg \max_{\theta \in \Theta} E[\ell(Y|X; \theta)]$, which is a unique interior point of Θ . Moreover, using the notation²*

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

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

$$\sqrt{n}(\hat{\theta} - \theta_0) = A^{-1} \left(\frac{1}{\sqrt{n}} \sum_{j=1}^n \Delta \ell(Y_j|X_j; \theta_0) \right) + o_p(1) \quad (2)$$

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

²We adopt the convention that the partial derivative to a row vector produces a column vector of partial derivatives.

Note that nothing is said about whether the parametric specification $F(y|X; \theta_0)$ is correct or not. Moreover, the assumption that $\hat{\theta}$ is the QML estimator is not essential as long as $\sqrt{n}(\hat{\theta} - \theta_0) = A^{-1} \left(n^{-1/2} \sum_{j=1}^n V_j \right) + o_p(1)$ 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.,

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

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

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

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

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

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

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

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

$$E[\exp(i\tau'Y) \exp(i\xi'X)] \neq E \left[\int \exp(i\tau'y)dF(y|X, \theta_0) \exp(i\xi'X) \right] \quad (3)$$

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. Once we have completed

the introduction of our ICM test and its asymptotic properties we will show that these bounded transformation only lead to a few minor changes in the notation. See the remark following Theorem 1 below.

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

$$\begin{aligned} \varsigma(\tau, \xi | \theta) &= E \left[\left(\exp(i\tau'Y) - \int \exp(i\tau'y) dF(y|X, \theta) \right) \exp(i\xi'X) \right] \\ \Upsilon &= \times_{j=1}^m [-\bar{\tau}_j, \bar{\tau}_j], \quad \bar{\tau}_j > 0 \\ \Xi &= \times_{j=1}^k [-\bar{\xi}_j, \bar{\xi}_j], \quad \bar{\xi}_j > 0 \end{aligned} \tag{4}$$

$$\tag{5}$$

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

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

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

$$\hat{T}_n = \int_{\Upsilon \times \Xi} |Z_n(\tau, \xi)|^2 d\mu(\tau, \xi), \tag{6}$$

where

$$Z_n(\tau, \xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\exp(i\tau'Y_j) - \int \exp(i\tau'y) dF(y|X_j, \hat{\theta}) \right) \exp(i\xi'X_j). \tag{7}$$

2.3 Asymptotic properties

To derive the asymptotic properties of the ICM statistic (6) 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 $F(y|X, \theta)$,*

$$\varphi(\tau|X; \theta) = \int \exp(i\tau'y) dF(y|X, \theta), \tag{8}$$

is a.s. continuously differentiable in θ in an open neighborhood Θ_0 of θ_0 , with column vector of partial derivatives $\Delta\varphi(\tau|X; \theta) = \partial\varphi(\tau|X; \theta) / \partial\theta'$ satisfying $E[\sup_{\theta \in \Theta_0} \|\Delta\varphi(\tau|X; \theta)\|] < \infty$.

Then

Lemma 1. *Under Assumptions 1-2,*

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(\tau|X_j; \hat{\theta}) \exp(i.\xi' X_j) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(\tau|X_j; \theta_0) \exp(i.\xi' X_j) \\ &\quad + b(\tau, \xi)' A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^n \Delta\ell(Y_j|X_j; \theta_0) + o_p(1) \end{aligned}$$

pointwise in (τ, ξ) , where $b(\tau, \xi) = E[\Delta\varphi(\tau|X; \theta_0) \exp(i.\xi' X)]$.

Proof: Appendix.

Consequently, denoting

$$\begin{aligned} \phi(\tau, \xi|Y, X) &= (\exp(i.\tau' Y) - \varphi(\tau|X; \theta_0)) \exp(i.\xi' X) \\ &\quad + b(\tau, \xi)' A^{-1} \Delta\ell(Y|X; \theta_0) \end{aligned} \tag{9}$$

$$\tilde{Z}_n(\tau, \xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \phi(\tau, \xi|Y_j, X_j), \tag{10}$$

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

it follows from Lemma 1 that

Lemma 2. *Under Assumptions 1-2, $\hat{T}_n = \tilde{T}_n + o_p(1)$, regardless whether H_0 is true or not.*

Proof: Appendix.

We are now able to state the first main result:

Theorem 1. *Let Y and X be bounded random vectors. Then under Assumptions 1-2 and H_0 ,*

$$\tilde{Z}_n \Rightarrow Z \text{ on } \Upsilon \times \Xi, \tag{11}$$

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

$$\Gamma((\tau_1, \xi_1), (\tau_2, \xi_2)) = E \left[\phi(\tau_1, \xi_1 | Y, X) \overline{\phi(\tau_2, \xi_2 | Y, X)} \right], \quad (12)$$

hence by Lemma 2 and the continuous mapping theorem,

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

Moreover, under H_1 ,

$$\widetilde{Z}_n(\tau, \xi) / \sqrt{n} \xrightarrow{p} E[(\exp(i\tau'Y) - \varphi(\tau|X; \theta_0)) \exp(i\xi'X)] = \rho(\tau, \xi)$$

say, where the set $\{(\tau, \xi) \in \Upsilon \times \Xi : \rho(\tau, \xi) > 0\}$ has positive Lebesgue measure, hence by Lemma 2 and bounded convergence,

$$\widehat{T}_n/n \xrightarrow{p} \int_{\Upsilon \times \Xi} |\eta(\tau, \xi)|^2 d\mu(\tau, \xi) > 0. \quad (13)$$

Proof: Appendix.

Remark: Recall that Y and X are bounded random vectors is not essential, 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

$$\varphi(\tau|X; \theta) = \int \exp(i\tau'\Phi_1(y)) dF(y|X, \theta), \quad (14)$$

$$b(\tau, \xi) = E[\Delta\varphi(\tau|X; \theta_0) \exp(i\xi'\Phi_2(X))] \quad (15)$$

$$\begin{aligned} \phi(\tau, \xi|Y, X) &= (\exp(i\tau'\Phi_1(Y)) - \varphi(\tau|X; \theta_0)) \exp(i\xi'\Phi_2(X)) \\ &\quad - b(\tau, \xi)' A^{-1} \sum_{j=1}^n \Delta\ell(Y|X; \theta_0). \end{aligned} \quad (16)$$

respectively.

However, there are two practical problems involved. The first one is that it may be difficult to compute the conditional characteristic function $\int \exp(i\tau'\Phi_1(y)) dF(y|X, \theta)$, even if the original conditional characteristic function $\int \exp(i\tau'y) dF(y|X, \theta)$ 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 the next subsection.

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

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

Assumption 3. For $\ell = 1, 2, \dots, m + k$, there exist constants μ_ℓ and $\sigma_\ell > 0$ such that $\sqrt{n}(\mu_{n,\ell} - \mu_\ell) = O_p(1)$ and $\sqrt{n}(\sigma_{n,\ell} - \sigma_\ell) = O_p(1)$.

These conditions hold for sample means and sample standard errors provided that the variables $V_{\ell,j}$ have finite fourth moments, and under mild conditions for quantiles as well.

Denote

$$\begin{aligned} (\widehat{\Phi}_1(Y_j)', \widehat{\Phi}_2(X_j)') &= (\arctan(S_{n,1,j}), \dots, \arctan(S_{n,m+k,j})), \\ (\Phi_1(Y_j)', \Phi_2(X_j)') &= (\arctan(S_{1,j}), \dots, \arctan(S_{m+k,j})), \end{aligned}$$

where $S_{\ell,j} = \sigma_\ell^{-1}(V_{\ell,j} - \mu_\ell)$. Redefine $Z_n(\tau, \xi)$ as

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

and let

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

Then the following results hold.

Lemma 3. *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: Appendix.

2.5 The null distribution

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

A complex-valued positive semi-definite function relative to a probability measure μ defined on the Borel sets in a Euclidean space \mathbb{R}^q is a Borel measurable function $\Gamma(\beta_1, \beta_2) : \mathbb{R}^q \times \mathbb{R}^q \rightarrow \mathbb{C}$, 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 (12) 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 (12) 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 4. *Let μ be a probability measure with compact support $\mathbf{B} \subset \mathbb{R}^q$, and let $\Gamma(\beta_1, \beta_2) : \mathbf{B} \times \mathbf{B} \rightarrow \mathbb{C}$ be a symmetric and continuous positive semi-definite function relative to μ . Consider the eigenvalue equation*

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

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

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

$j = 1, 2, 3, \dots$. The eigenvalues λ_j are real-valued³ 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 representation⁴

$$\Gamma(\beta_1, \beta_2) = \sum_{j=1}^{\infty} \lambda_j \psi_j(\beta_1) \overline{\psi_j(\beta_2)}. \quad (17)$$

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 et al. (2002) and Krein (1998).

Lemma 5. Suppose that the function Z in Lemma 4 is a zero-mean complex-valued continuous Gaussian process on \mathbf{B} with covariance function

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

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

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

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,

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

³Due to the symmetry of Γ .

⁴The result (17) is the actual Mercer's theorem.

Proof: Appendix.

It follows now straightforwardly from Lemmas 4 and 5 that the following result holds.

Theorem 2. Denote $g_j = \int_{\Upsilon \times \Xi} Z(\tau, \xi) \bar{\psi}_j(\tau, \xi) d\mu(\tau, \xi)$, where the ψ_j 's are the eigenfunctions of the covariance function (12). Under H_0 and the conditions of Theorem 1,

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

where the $\chi_{2,j}^2$'s are independently χ_2^2 distributed and the λ_j 's are the corresponding eigenvalues of the covariance function (12).

Because the covariance function (12) depends on the joint distribution of Y and X , so do the eigenvalues λ_j , 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.,

$$\Pr \left[\sup_{y \in \mathbb{R}^m} |Q(y|X) - F(y|X, \theta_0)| = 0 \right] < 1. \quad (18)$$

Consider the \sqrt{n} -local alternative

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

It follows straightforwardly from (9) that under H_1^L ,

$$E [Z(\tau, \xi)] = \varphi_Q(\tau, \xi) - \varphi_F(\tau, \xi). \quad (19)$$

where

$$\begin{aligned} \varphi_Q(\tau, \xi) &= E \left[\left(\int \exp(i \cdot \tau' y) dQ(y|X) \right) \exp(i \xi' X) \right], \\ \varphi_F(\tau, \xi) &= E \left[\left(\int \exp(i \cdot \tau' y) dF(y|X, \theta_0) \right) \exp(i \xi' X) \right]. \end{aligned} \quad (20)$$

Recall from Lemma 4 that we can write $E[Z(\tau, \xi)] = \sum_{j=1}^{\infty} \eta_j \psi_j(\tau, \xi)$, where

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

It follow now from Lemmas 4 and 5 that under H_1^L ,

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

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

Since condition (18) 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 from Corollary 1 in Bierens and Ploberger (1997)⁵ that

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

for all $K > 0$. This implies that the ICM test has non-trivial power against \sqrt{n} -local alternatives.

Note that the argument of Escanciano (2009) regarding the local power of the ICM test for regression models applies to our case as well. Translated to our case, Escanciano's point boils down to the following. Let $K(\alpha)$ be the $\alpha \times 100\%$ critical value of T , i.e., $\Pr [T > K(\alpha)] = \alpha$. Then for arbitrary $\varepsilon > 0$ there exists a $\delta > 0$ such that $\alpha < \Pr [T_{alt} > K(\alpha)] < \alpha + \varepsilon$ if $\sum_{j=1}^{\infty} |\eta_j|^2 \in (0, \delta)$. Moreover, given $\delta > 0$, it is possible to choose $Q(y|X)$ so close to $F(y|X, \theta_0)$ that $\sum_{j=1}^{\infty} |\eta_j|^2 \in (0, \delta)$.⁶ Thus, the local power may be near-trivial if the local alternative is too close to the null model.

⁵Let $T_{alt}^* = \sum_{j=1}^{\infty} \lambda_j e_{1,j}^2 + \sum_{j=1}^{\infty} (\operatorname{Im}(\eta_j) + \sqrt{\lambda_j} e_{2,j})^2$. It follows from the proof of Corollary 1 in Bierens and Ploberger (1997) that for any $K > 0$, $\Pr [T_{alt} > K] \geq \Pr [T_{alt}^* > K]$, where the inequality is strict if for some j , $\operatorname{Re}(\eta_j) \neq 0$. By the same argument, $\Pr [T_{alt}^* > K] \geq \Pr [T > K]$, where the inequality is strict if for some j , $\operatorname{Im}(\eta_j) \neq 0$.

⁶For example, let $Q_\ell(y|X) = F(y|X, \theta_0)^{1+1/\ell}$, $\ell > 0$, with corresponding characteristic function $\varphi_{Q_\ell}(\tau, \xi)$ of the type (20) and Fourier coefficients $\eta_{\ell,j}$. Then there exists an $\ell(\delta)$ such that for all finite $\ell > \ell(\delta)$, $\int_{\Upsilon \times \Xi} |\varphi_{Q_\ell}(\tau, \xi) - \varphi_F(\tau, \xi)|^2 d\mu(\tau, \xi) = \sum_{j=1}^{\infty} |\eta_{\ell,j}|^2 \in (0, \delta)$.

2.7 Maximizing the ICM test over the integration domain

The choice of the hypercubes Υ and Ξ defined by (4) and (5), 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 (4) and $\underline{\Xi}$ and $\overline{\Xi}$ are given hypercubes in \mathbb{R}^k of the type (5), provided that it can be shown that under the null hypothesis,

$$\begin{aligned} & \sup_{\underline{\Upsilon} \subset \Upsilon \subset \overline{\Upsilon}, \underline{\Xi} \subset \Xi \subset \overline{\Xi}} \int_{\Upsilon \times \Xi} |Z_n(\tau, \xi)|^2 d\mu(\tau, \xi) \\ & \xrightarrow{d} \sup_{\underline{\Upsilon} \subset \Upsilon \subset \overline{\Upsilon}, \underline{\Xi} \subset \Xi \subset \overline{\Xi}} \int_{\Upsilon \times \Xi} |Z(\tau, \xi)|^2 d\mu(\tau, \xi). \end{aligned} \quad (21)$$

Indeed, (21) 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*

$$\begin{aligned} \widehat{T}_n(c) &= \frac{1}{(2c)^{m+k}} \int_{\Upsilon(c) \times \Xi(c)} |Z_n(\tau, \xi)|^2 d\tau d\xi, \\ T(c) &= \frac{1}{(2c)^{m+k}} \int_{\Upsilon(c) \times \Xi(c)} |Z(\tau, \xi)|^2 d\tau d\xi \end{aligned} \quad (22)$$

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

Proof: Appendix

Although it is too much of a computational burden to compute this supremum exactly, let alone the supremum in (21), this result motivates 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 distrib-

ution cannot be derived analytically or is case dependent. Bickle and Freedman (1987) developed the asymptotic theory for general bootstrap cases. Conditions under which the bootstrap method fails can be found in Athreya (1987). For more discussions 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 $\hat{\theta}$, generate M bootstrap samples $\{(\tilde{Y}_{b,1}, X_1), \dots, (\tilde{Y}_{b,n}, X_n)\}$, $b = 1, \dots, M$, where $\tilde{Y}_{b,j}$ is a random drawing from $F(y|X_j; \hat{\theta})$ in bootstrap sample b . The vectors X_j of covariates are the same as in the actual sample. Let $\tilde{\theta}_b$ be the ML estimator on the basis of this bootstrap sample, i.e., $\tilde{\theta}_b = \arg \max_{\theta \in \Theta} \ln L_{b,n}(\theta)$, where $\ln L_{b,n}(\theta) = \sum_{j=1}^n \ell(\tilde{Y}_{b,j}|X_j; \theta)$.

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

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

where $\hat{Z}_{b,n}(\tau, \xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\exp(i \cdot \tau' \tilde{Y}_{b,j}) - \varphi(\tau|X_j; \tilde{\theta}_b) \right) \exp(i \cdot \xi' X_j)$ with $\varphi(\tau|X_j; \tilde{\theta}_b) = \int \exp(i\tau'y) dF(y|X_j; \tilde{\theta}_b)$. We will set forth conditions such that $\hat{Z}_{b,n} \Rightarrow Z_b$ as $n \rightarrow \infty$, 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 $(\tilde{\theta}_b - \hat{\theta}) \xrightarrow{p} 0$. 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,*

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

and

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

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 its turn implies⁷ that $\tilde{\theta}_b = \arg \max_{\theta \in \Theta} \ln L_{b,n}(\theta) / n \xrightarrow{p} \theta_0$.

The next step is to show that $\sqrt{n}(\tilde{\theta}_b - \hat{\theta})$ has the same limiting distribution as $\sqrt{n}(\hat{\theta} - \theta_0)$. For this we need the following standard regularity conditions on the vector $\Delta \ell(y|X; \theta)$ and the matrix $\Delta^2 \ell(y|X_j; \theta)$ 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 $\int \Delta^2 \ell(y|X_j; \theta) dF(y|X_j; \theta)$ are a.s. continuous on an arbitrary open neighborhood Θ_0 of θ_0 , and*

$$\int \Delta \ell(y|X; \theta) dF(y|X; \theta) = (\partial/\partial \theta') \int \ell(y|X; \theta) dF(y|X; \theta) = 0 \quad (23)$$

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

$$E \left[\sup_{\theta \in \Theta_0} \left\| \int \Delta \ell(y|X_j; \theta) dF(y|X_j; \theta) \right\|^{2+\delta} \right] < \infty \quad (24)$$

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

The matrix norm $\|\cdot\|$ in (25) is the maximum absolute value of the elements of the matrix involved.

Note that Assumption 1 and part (23) of Assumption 5 imply that

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

⁷See for example Theorem 6.11 in Bierens (2004), which originates from Jennrich (1969).

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

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

is a double array of independent random vectors, for which Liapounov's central limit theorem applies. See for example Chung (1974, p. 200). This is the reason for the δ in (24). In particular, choose an arbitrary nonzero vector $\xi \in \mathbb{R}^p$, and denote $z_{j,n} = \xi'U_{j,n}$ and $\sigma_n^2 = \frac{1}{n} \sum_{j=1}^n \xi' E[U_{j,n}'U_{j,n}|\mathcal{D}_n]\xi$. 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 $\sigma^2 = p \lim_{n \rightarrow \infty} \sigma_n^2 = \xi' B \xi$. This result can also be proved using a martingale difference central limit theorem. See McLeish (1974) for the latter. Thus,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \Delta\ell\left(\tilde{Y}_{b,j}|X_j;\hat{\theta}\right) \xrightarrow{d} N_p[0, B], \quad (26)$$

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

$$\sqrt{n}(\tilde{\theta}_b - \hat{\theta}) \xrightarrow{d} N_p[0, A^{-1}BA^{-1}]$$

where A is the same as in Assumption 1. Therefore, it is easy to verify that similar to Lemma 1 the following result holds.

Lemma 6. *Under Assumptions 1-2 and 4-5,*

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(\tau|X_j;\tilde{\theta}_b) \exp(i.\xi'X_j) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi\left(\tau|X_j;\hat{\theta}\right) \exp(i.\xi'X_j) \\ &\quad + b(\tau, \xi)' A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^n \Delta\ell\left(\tilde{Y}_{b,j}|X_j;\hat{\theta}\right) + o_p(1) \end{aligned}$$

pointwise in (τ, ξ) , where $b(\tau, \xi)$ is the same as in Lemma 1.

Consequently, denoting

$$\begin{aligned} \phi_{b,j}(\tau, \xi|\hat{\theta}) &= \left(\exp\left(i.\tau'\tilde{Y}_{b,j}\right) - \varphi\left(\tau|X_j;\hat{\theta}\right) \right) \exp(i.\xi'X_j) \\ &\quad + b(\tau, \xi)' A^{-1} \Delta\ell\left(\tilde{Y}_{b,j}|X_j;\hat{\theta}\right) \end{aligned} \quad (27)$$

$$\begin{aligned}
\tilde{Z}_{b,n}(\tau, \xi) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \phi_{b,j}(\tau, \xi | \hat{\theta}), \\
\tilde{T}_{b,n} &= \int_{\Upsilon \times \Xi} |\tilde{Z}_{b,n}(\tau, \xi)|^2 d\mu(\tau, \xi)
\end{aligned} \tag{28}$$

it follows from Lemma 6 that

Lemma 7. *Under Assumptions 1-2 and 4-5, $\hat{T}_{b,n} = \tilde{T}_{b,n} + o_p(1)$.*

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

Let

$$\begin{aligned}
\tilde{Z}_{1,b,n}(\tau, \xi) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\exp \left(i \cdot \tau' \tilde{Y}_{b,j} \right) - \varphi \left(\tau | X_j; \hat{\theta} \right) \right) \exp \left(i \cdot \xi' X \right) \\
\tilde{Z}_{2,b,n}(\tau, \xi) &= b(\tau, \xi)' A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^n \Delta \ell \left(\tilde{Y}_{b,j} | X_j; \hat{\theta} \right)
\end{aligned}$$

Since $b(\tau, \xi)$ is uniformly continuous on $\Upsilon \times \Xi$, the tightness of $\tilde{Z}_{2,b,n}$ follows from (26). The tightness of $\tilde{Z}_{1,b,n}$ follows from the proof of Theorem 1, simply by replacing the expectations involved by the corresponding conditional expectations $E[\cdot | \mathcal{D}_n]$. Moreover, similar to (26) it can be shown that the finite distributions of $\tilde{Z}_{1,b,n}(\tau, \xi)$ converge to a multivariate normal distribution. Thus, $\tilde{Z}_{b,n} \Rightarrow Z_*$. Furthermore, it is easy to verify that the covariance function of this limiting process takes the form

$$\begin{aligned}
& p \lim_{n \rightarrow \infty} E \left[\tilde{Z}_{b,n}(\tau_1, \xi_1) \overline{\tilde{Z}_{b,n}(\tau_2, \xi_2)} \middle| \mathcal{D}_n \right] \\
&= p \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n E \left[\phi_{b,j}(\tau_1, \xi_1 | \hat{\theta}) \overline{\phi_{b,j}(\tau_2, \xi_2 | \hat{\theta})} \middle| \mathcal{D}_n \right] \\
&= \Gamma \left((\tau_1, \xi_1), (\tau_2, \xi_2) \right)
\end{aligned}$$

where the latter is the same as in Theorem 1. Thus, we have the following result.

Theorem 4. *Let Y and X be bounded random vectors.⁸ Then under Assumptions 1-2 and 4-5, $\tilde{Z}_{b,n} \Rightarrow Z_b$ 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,*

$$\left(\hat{T}_{1,n}, \dots, \hat{T}_{M,n}\right)' \xrightarrow{d} (T_1, \dots, T_M)' \quad (29)$$

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 (29) carries over if the components of these random vectors are sorted in decreasing order. Then it follows that for $\alpha \in (0, 1)$, $\hat{T}_{[\alpha M],n} \xrightarrow{d} T_{[\alpha M]}$, where $[\alpha M]$ is the largest integer $\leq \alpha M$. The statistic $\hat{T}_{[\alpha M],n}$ 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 multi-dimensional. 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

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

⁸Note that then by Assumption 1 the bootstrap $\tilde{Y}_{b,j}$'s are bounded too.

if Y_j and X_j are bounded random vectors, or

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

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 (7) 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

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

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 (13).

Proof: 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 (12). 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 \tilde{Y} from it. 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 $Y_{\ell,j}$, $\tilde{Y}_{\ell,j}$ and $X_{\ell,j}$ components ℓ of Y_j , \tilde{Y}_j and X_j , respectively, the SICM version of the test statistic (22) takes the form

$$\begin{aligned}
\widehat{T}_n^{(s)}(c) &= \frac{1}{(2c)^{k+m}} \int_{[-c,c]^m} \int_{[-c,c]^k} |\widehat{Z}_n^{(s)}(\tau, \xi)|^2 d\tau d\xi \\
&= \frac{2}{n} \sum_{j_1=1}^{n-1} \sum_{j_2=j_1+1}^n \left(\prod_{\ell=1}^m \frac{\sin(c(Y_{\ell,j_1} - Y_{\ell,j_2}))}{c(Y_{\ell,j_1} - Y_{\ell,j_2})} + \prod_{\ell=1}^m \frac{\sin(c(\tilde{Y}_{\ell,j_1} - \tilde{Y}_{\ell,j_2}))}{c(\tilde{Y}_{\ell,j_1} - \tilde{Y}_{\ell,j_2})} \right. \\
&\quad \left. - \prod_{\ell=1}^m \frac{\sin(c(Y_{\ell,j_1} - \tilde{Y}_{\ell,j_2}))}{c(Y_{\ell,j_1} - \tilde{Y}_{\ell,j_2})} - \prod_{\ell=1}^m \frac{\sin(c(\tilde{Y}_{\ell,j_1} - Y_{\ell,j_2}))}{c(\tilde{Y}_{\ell,j_1} - Y_{\ell,j_2})} \right) \\
&\quad \times \left(\prod_{\ell=1}^k \frac{\sin(c(X_{\ell,j_1} - X_{\ell,j_2}))}{c(X_{\ell,j_1} - X_{\ell,j_2})} \right) \\
&\quad + \frac{2}{n} \sum_{j=1}^n \left(1 - \prod_{\ell=1}^m \frac{\sin(c(Y_{\ell,j} - \tilde{Y}_{\ell,j}))}{c(Y_{\ell,j} - \tilde{Y}_{\ell,j})} \right) \tag{30}
\end{aligned}$$

as is not hard to verify.

Note that $\lim_{c \downarrow 0} \widehat{T}_n^{(s)}(c) = \lim_{c \rightarrow \infty} \widehat{T}_n^{(s)}(c) = 0$, so that 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 , for example $c_1 < c_2 < \dots < c_r$, and using the MAXSICM statistic

$$\max_{i=1,2,\dots,r} \widehat{T}_n^{(s)}(c_i) \tag{31}$$

as the actual test statistic.

3.2 Small sample performance

In first instance we have compared the small sample size and power of the bootstrapped SICM test $\widehat{T}_n^{(s)}(c)$ [c.f. (30)] for $c = 4, 5, 6$, and bootstrap sample size $M = 500$, with the results in Zheng (2000), using the same

data generating processes, null hypotheses, sample sizes ($n = 200, 300$) and number of replications (1000) as Zheng. The performance of the SICM test is similar to Zheng's test, except that due to the bootstrap the empirical size of our tests is closer to the nominal size than in Zheng (2000), who uses asymptotic critical values. The detailed results are not reported here, but are available from the authors upon request.

Instead, we will present the results of a limited Monte Carlo analysis where we focus on a model for which Zheng's (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.

The null hypothesis is that the dependent variable Y is generated by the following data generating process

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

with actual data generating processes of the type Poisson and Negative Binomial (NB) Logit:

$$\begin{aligned} H_1^{(0)} & : Y|X \sim \text{Poisson}(\exp(X)) \\ H_1^{(1)} & : Y|X \sim \text{NB}(1, p(X)) \\ H_1^{(2)} & : Y|X \sim \text{NB}(5, p(X)) \\ H_1^{(3)} & : Y|X \sim \text{NB}(10, p(X)) \end{aligned}$$

where $p(x) = (1 + \exp(-x))^{-1}$ is the Logit function, and the single covariate X is standard normally distributed. The sample size is $n = 200$, and the bootstrap sample size is $M = 500$. We have only used 200 replications, as this is sufficient to shed light on the small sample size and power properties of the SICM test. The SICM test involved is the MAXSICM test (31) with $c_i = 5 \cdot i$ for $i = 1, 2, 3, 4, 5$. Both Y and X are first 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 \tilde{Y} are transformed similarly, using the sample mean and sample standard error of the actual Y variable. The rejection percentages at the 1, 5 and 10% significance levels are reported in Table 1.

Table 1: MAXSICM test

Monte Carlo results			
	1%	5%	10%
$H_1^{(0)}$	0	5	10
$H_1^{(1)}$	52	71	83
$H_1^{(2)}$	33	56	68
$H_1^{(3)}$	30	52	66

Clearly, the empirical size is very good, and the small sample power against the negative binomial alternatives is what can be expected for such a small sample size. Note that the decrease in power is due to the fact that the conditional Negative Binomial Logit (NBL) distribution $NB(m, p(X))$ approaches the conditional Poisson distribution for $m \rightarrow \infty$.

4 Application to health economic count data models

Count data are often encountered in health economics, like 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). This section applies the MAXSICM method to test whether a conditional Poisson model is correctly specified. To the best of our knowledge, there is no other consistent specification test available for the conditional Poisson distribution. 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 4406 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 2. However, since 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 $j = 3, 4, \dots, 16$, so that the actual number of covariates is 44. A preliminary data analysis reveals that the extended Poisson model has lower values for the Hannan-Quinn (1979) and Schwarz (1978) information criteria than the model with the original sixteen covariates, so that the additional 28 covariates contribute to the fit of the model.

Table 2: Model variables

	Variable name	Meaning
Y	ofp	# of visit 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	# of chronicle 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

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 (31) to test the Poisson hypothesis, with $c_i = 5 \cdot i$ for $i = 1, 2, 3, 4, 5$, and bootstrap sample size 500. It suffices to include only the original sixteen covariates in Table 2 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

⁹Because the corresponding σ -algebras are the same.

the sixteen 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 $\text{Poisson}(\mu)$ 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 (1990) test, based on the regression

$$((Y_j - \hat{\mu}_j)^2 - Y_j)/\hat{\mu}_j = \alpha \cdot \hat{\mu}_j + \varepsilon_j, \quad (32)$$

where $\hat{\mu}_j = \exp((1, X_j')\hat{\theta})$ with $\hat{\theta}$ 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 \hat{t} of the OLS estimate $\hat{\alpha}$ of α . The results are $\hat{\alpha} = 0.874068$, $\hat{t} = 12.7497$. Thus, the Cameron-Trivedi test also strongly rejects the validity of the Poisson model.

As a further comparison we have also tried to conduct 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, so that the CK test statistic collapses to $\max_{1 \leq j \leq n} |1 - F(Y_j|X_j, \hat{\theta})|/\sqrt{n} < 1/\sqrt{n} \approx 0.015$.

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

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

where V represent unobserved heterogeneity which is independent of X , and if V is $\text{Gamma}(m, \beta)$ distributed, then the conditional distribution of Y given X alone is of the Negative Binomial Logit (NBL) type:

$$Y|X \sim \text{NB}(m, p(-\ln \beta - (1, X')\theta_0)), \quad (33)$$

where again $p(x)$ is the Logit function. Note that due to the presence of a constant term we may without loss of generality assume that $E[V] = 1$, which corresponds to $\beta = 1/m$ in the $\text{NBL}(m)$ case, so that $E[Y|X] =$

$\exp((1, X')\theta_0)$. Moreover, it is easy to verify that then the Poisson ML estimator $\hat{\theta}$ is still consistent, and that in the NBL(m) case the OLS estimator $\hat{\alpha}$ of the parameter α in the Cameron-Trivedi model (32) converges in probability to $1/m$. Since $1/\hat{\alpha} \approx 1.144$ is somewhat close to $m = 1$ we will therefore now try a 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, so that 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.

The model estimation and test computations have been conducted by a modified version of the free econometric software package *EasyReg International* developed by the first author. See Bierens (2008). The modified EasyReg modules involved are available from Bierens upon requests.¹⁰

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, 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, so that 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

¹⁰These modified modules only work within the EasyReg environment. Therefore, EasyReg has to be installed first.

with a large number of covariates, contrary to Andrews' (1997) Conditional Kolmogorov test, and does not suffer from the curse of dimensionality as is the case with Zheng's (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 (like the 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 a negative binomial Logit (NBL) model might be more appropriate for this data. However the NBL(m) models for $m = 1, 2$ are strongly rejected as well 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.

References

- Andrews, D.W. (1988) Chi-square diagnostic tests for econometric models: Theory. *Econometrica* 56, 1419-1453.
- Andrews, D.W. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097-1128.
- Andrews, D.W. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383-1414.
- Athreya, K.B. (1987) Bootstrap of the mean in the infinite variance case. *The Annals of Statistics* 15, 724-731.
- Bai, J. (2003) Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85, 531-549.
- Bickel, P. & D. Freedman (1981) Asymptotic theory for the bootstrap. *The Annals of Statistics* 9, 1196-1927.
- Bierens, H.J. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105-134.
- Bierens, H.J. (1984) Model specification testing of time series regressions. *Journal of Econometrics* 26, 323-353.
- 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. (2008) *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*. John 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*. John Wiley.
- Chung, K.L. (1976) *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.
- De Jong, R. (1996) The Bierens test under data dependence. *Journal of Econometrics* 72, 1-32.
- Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7, 1-26.
- Escanciano, J.C. (2006) A Consistent diagnostic test for regression models using projections. *Econometric Theory* 22, 1030-1051.
- Escanciano, J.C. (2009) On the lack of power of omnibus specification tests. *Econometric Theory* 25, 162-194.
- 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 No. IKT-NT 1019, Communications Technological Laboratory, Swiss Federal Institute of Technology, Zurich.

- Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926-1947.
- Hausman, J. (1978) Specification tests in econometrics. *Econometrica* 46, 1251-1271.
- 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.
- Holly, A. (1982) A remark on Hausman's specification test. *Econometrica* 50, 749-760.
- 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.
- Justel, A., D. Pena & R. Zamar (1997) A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics and Probability Letters* 35, 251-259.
- Krein, M.G (1998) Compact linear operators on functional spaces with two norms. *Integral Equations and Operator Theory* 30, 140-162.
- 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.
- Kullback, L. & R.A. Leibler (1951) On information and sufficiency. *Annals of Mathematical Statistics* 22, 79-86.
- McLeish, D.L. (1974) Dependent central limit theorems and invariance principles. *Annals of Probability* 2, 620-628.
- Newey, W.K. (1985) Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53, 1047-1070.
- 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. (1981) Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* 76, 419-433.

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

Inconsistency of Bai's (2003) test

Consider the following counter-example. Let X be standard exponentially distributed and suppose that the true conditional distribution of Y given X is exponential: $F_1(y|X) = 1 - \exp(-y/X)$. Let the null model be $F_0(y|X) = G(y/X^2)$, where

$$G(z) = 1 - \frac{1}{1+z}, \quad z \geq 0.$$

Note that the inverse of G is $G^{-1}(u) = (1-u)^{-1} - 1$, $u \in (0, 1)$. Then for $u \in (0, 1)$,

$$\begin{aligned} \Pr [F_0(Y|X) \leq u|X] &= \Pr [G(Y/X^2) \leq u|X] = \Pr [Y \leq X^2 G^{-1}(u) |X] \\ &= 1 - \exp(-X \cdot G^{-1}(u)), \end{aligned}$$

hence

$$\begin{aligned} \Pr [F_0(Y|X) \leq u] &= 1 - \int_0^\infty \exp(-x \cdot G^{-1}(u)) \exp(-x) dx \\ &= 1 - \frac{1}{1 + G^{-1}(u)} = u. \end{aligned}$$

This demonstrates that Bai's test is not consistent. The reason is that Bai tests the null hypothesis that $F_0(Y|X)$ has a uniform $[0, 1]$ distribution,

whereas the correct null hypothesis should have been that *conditional on* X , $F_0(Y|X)$ has a uniform $[0, 1]$ distribution.

Proof of Lemma 1

By the mean value theorem and Assumption 1

$$\begin{aligned}
& \operatorname{Re} \left[\sqrt{n} \varphi(\tau|X; \hat{\theta}) \right] \\
&= \operatorname{Re} \left[\sqrt{n} \varphi(\tau|X; \theta_0) \right] + \left(\operatorname{Re} \left[\Delta \varphi \left(\tau|X; \theta_0 + \tilde{\lambda}_1(\tau|X) (\hat{\theta} - \theta_0) \right) \right] \right)' \\
&\quad \times \sqrt{n} (\hat{\theta} - \theta_0) \\
&= \operatorname{Re} \left[\sqrt{n} \varphi(\tau|X; \theta_0) \right] + (\operatorname{Re} [\Delta \varphi(\tau|X; \theta_0)])' \sqrt{n} (\hat{\theta} - \theta_0) \\
&\quad + \operatorname{Re} \left[\hat{R}_n(\tau|X) \right]' \sqrt{n} (\hat{\theta} - \theta_0)
\end{aligned}$$

and

$$\begin{aligned}
& \operatorname{Im} \left[\sqrt{n} \varphi(\tau, \hat{\theta}|X) \right] \\
&= \operatorname{Im} \left[\sqrt{n} \varphi(\tau|X; \theta_0) \right] + \left(\operatorname{Im} \left[\Delta \varphi \left(\tau|X; \theta_0 + \tilde{\lambda}_2(\tau|X) (\hat{\theta} - \theta_0) \right) \right] \right)' \\
&\quad \times \sqrt{n} (\hat{\theta} - \theta_0) \\
&= \operatorname{Im} \left[\sqrt{n} \varphi(\tau|X; \theta_0) \right] + (\operatorname{Im} [\Delta \varphi(\tau|X; \theta_0)])' \sqrt{n} (\hat{\theta} - \theta_0) \\
&\quad + \operatorname{Im} \left[\hat{R}_n(\tau|X) \right]' \sqrt{n} (\hat{\theta} - \theta_0)
\end{aligned}$$

where $\tilde{\lambda}_2(\tau|X) \in [0, 1]$ and $\tilde{\lambda}_1(\tau|X) \in [0, 1]$ a.s. for all $\tau \in \mathbb{R}^m$, and

$$\begin{aligned}
\hat{R}_n(\tau|X) &= \operatorname{Re} \left[\Delta \varphi \left(\tau|X; \theta_0 + \tilde{\lambda}_1(\tau|X) (\hat{\theta} - \theta_0) \right) - \Delta \varphi(\tau|X; \theta_0) \right] \\
&\quad + i \operatorname{Im} \left[\Delta \varphi \left(\tau|X; \theta_0 + \tilde{\lambda}_2(\tau|X) (\hat{\theta} - \theta_0) \right) - \Delta \varphi(\tau|X; \theta_0) \right]
\end{aligned}$$

Note that

$$\left\| \hat{R}_n(\tau|X) \right\| \leq \sup_{\|\theta - \theta_0\| \leq \|\hat{\theta} - \theta_0\|} \left\| \Delta \varphi(\tau|X; \theta) - \Delta \varphi(\tau|X; \theta_0) \right\|$$

so that by Assumptions 1-2, for sufficiently small $\delta > 0$,¹¹

$$\frac{1}{n} \sum_{j=1}^n \left\| \hat{R}_n(\tau|X_j) \right\|$$

¹¹I.e., $\{\theta \in \mathbb{R}^p : \|\theta - \theta_0\| \leq \delta\} \subset \Theta_0$.

$$\begin{aligned}
&\leq \mathbf{1} \left(\left\| \widehat{\theta} - \theta_0 \right\| \leq \delta \right) \left(\frac{1}{n} \sum_{j=1}^n \sup_{\|\theta - \theta_0\| \leq \delta} \|\Delta\varphi(\tau|X_j; \theta) - \Delta\varphi(\tau|X_j; \theta_0)\| \right) \\
&+ 2.1 \left(\left\| \widehat{\theta} - \theta_0 \right\| > \delta \right) \frac{1}{n} \sum_{j=1}^n \sup_{\theta \in \Theta_0} \|\Delta\varphi(\tau|X_j; \theta_0)\| \\
&\leq \frac{1}{n} \sum_{j=1}^n \sup_{\|\theta - \theta_0\| \leq \delta} \|\Delta\varphi(\tau|X_j; \theta) - \Delta\varphi(\tau|X_j; \theta_0)\| + o_p(1).
\end{aligned}$$

By Chebyshev's inequality for first moments,

$$\Pr \left[\frac{1}{n} \sum_{j=1}^n \left\| \widehat{R}_n(\tau|X_j) \right\| > \varepsilon \right] \leq \varepsilon^{-1} E \left[\sup_{\|\theta - \theta_0\| \leq \delta} \|\Delta\varphi(\tau|X; \theta) - \Delta\varphi(\tau|X; \theta_0)\| \right]$$

for arbitrary $\varepsilon > 0$, which by the continuity of $\Delta\varphi(\tau|X; \theta)$ in $\theta \in \Theta_0$ can be made arbitrarily small by decreasing δ towards zero. Hence pointwise in τ ,

$$p \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \left\| \widehat{R}_n(\tau|X_j) \right\| = 0.$$

Consequently, it follows now from Assumptions 1-2 that pointwise in τ and ξ ,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(\tau|X_j; \widehat{\theta}) \exp(i.\xi'X_j) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(\tau|X_j; \theta_0) \exp(i.\xi'X_j) \\
&+ \left(\frac{1}{n} \sum_{j=1}^n \Delta\varphi(\tau|X_j; \theta_0) \exp(i.\xi'X_j) \right)' \sqrt{n}(\widehat{\theta} - \theta_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^n \varphi(\tau|X_j; \theta_0) \exp(i.\xi'X_j) \\
&+ b(\tau, \xi)' A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^n \Delta\ell(Y_j|X_j; \theta_0) + o_p(1)
\end{aligned}$$

where $b(\tau, \xi) = E[\Delta\varphi(\tau|X; \theta_0)(i.\xi'X)]$.

Proof of Lemma 2

Let $U_n(\tau, \xi) = Z_n(\tau, \xi) - \tilde{Z}_n(\tau, \xi)$ and observe from Lemma 1 that $p \lim_{n \rightarrow \infty} U_n(\tau, \xi) = 0$, pointwise in (τ, ξ) , hence by bounded convergence

$$\int_{\Upsilon \times \Xi} |U_n(\tau, \xi)|^2 d\mu(\tau, \xi) = o_p(1) \quad (34)$$

Next, observe that

$$E \left[\int_{\Upsilon \times \Xi} |\tilde{Z}_n(\tau, \xi)|^2 d\mu(\tau, \xi) \right] = \int_{\Upsilon \times \Xi} E [|\phi(\tau, \xi|Y, X)|^2] d\mu(\tau, \xi) = O(1),$$

hence

$$\int_{\Upsilon \times \Xi} |\tilde{Z}_n(\tau, \xi)|^2 d\mu(\tau, \xi) = O_p(1). \quad (35)$$

Finally, it is easy to verify that

$$\left| |Z_n(\tau, \xi)|^2 - |\tilde{Z}_n(\tau, \xi)|^2 \right| \leq |U_n(\tau, \xi)|^2 + 2|U_n(\tau, \xi)| \cdot |\tilde{Z}_n(\tau, \xi)|$$

and therefore

$$\begin{aligned} |\hat{T}_n - \tilde{T}_n| &\leq \int_{\Upsilon \times \Xi} |U_n(\tau, \xi)|^2 d\mu(\tau, \xi) \\ &\quad + 2 \int_{\Upsilon \times \Xi} |U_n(\tau, \xi)| \cdot |\tilde{Z}_n(\tau, \xi)| d\mu(\tau, \xi) \\ &\leq \int_{\Upsilon \times \Xi} |U_n(\tau, \xi)|^2 d\mu(\tau, \xi) \\ &\quad + 2 \sqrt{\int_{\Upsilon \times \Xi} |\tilde{Z}_n(\tau, \xi)|^2 d\mu(\tau, \xi)} \\ &\quad \times \sqrt{\int_{\Upsilon \times \Xi} |U_n(\tau, \xi)|^2 d\mu(\tau, \xi)} \\ &= o_p(1) \end{aligned}$$

where the latter follows from (34) and (35).

Proof of Theorem 1

Let $Z_n(\beta)$ be an empirical process on a compact subset \mathbf{B} of an 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 $\beta_1, \beta_2, \dots, \beta_M$ in \mathbf{B} ,

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

In the case of the empirical process (10) 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 compact set $K \subset \mathcal{M}$ such that $\inf_{n \geq 1} \Pr[Z_n \in K] > 1 - \varepsilon$.

According to Billingsley (1968, Theorem 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.

Due to Lemma 2 it suffices to prove $\tilde{Z}_n(\tau, \xi) \Rightarrow Z(\tau, \xi)$, where $\tilde{Z}_n(\tau, \xi)$ defined by (10). Thus, in our case $\beta = (\tau, \xi)$, $\mathbf{B} = \Upsilon \times \Xi$, and $Z_n(\beta) = \tilde{Z}_n(\tau, \xi)$.

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

$$\begin{aligned} \tilde{Z}_{1,n}(\tau, \xi) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\exp(i \cdot \tau' Y_j) - \int \exp(i \cdot \tau' y) dF(y|X_j; \theta_0) \right) \exp(i \cdot \xi' X_j) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n (\exp(i \cdot \tau' Y_j) - E[\exp(i \cdot \tau' Y_j)|X_j]) \exp(i \cdot \xi' X_j) \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

Hence,

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

and thus

$$\begin{aligned}
& E \left[\left(\sup_{\|(\tau, \xi)' - (\rho, \zeta)'\| < \varepsilon} \left| \tilde{Z}_{1,n}(\tau, \xi) - \tilde{Z}_{1,n}(\rho, \zeta) \right| \right)^2 \right] \\
& \leq 16\varepsilon^2 \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{r!s!} (4c)^{r+s} \\
& \quad \times \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{r!s!} (2c)^{r+s} \sum_{b=0}^r \binom{r}{b} \sum_{q=0}^s \binom{s}{q} E (Y_{1,1}^b Y_{2,1}^{r-b} X_{1,1}^q X_{2,1}^{s-q})^2 \right) \\
& = 16\varepsilon^2 \exp(8c) E \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{r!s!} (2c)^{r+s} \|Y\|^{2r} \|X\|^{2s} \right) \\
& = 16\varepsilon^2 \exp(8c) E \left[\exp(2c(\|Y\|^2 + \|X\|^2)) \right]
\end{aligned}$$

where the inequality is due to Schwarz inequality. Thus, 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.

Proof of Lemma 3

Assume that the Y_j 's and X_j 's are scalar random variables, and that X_j is already bounded. Then

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

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

$$\sqrt{n}(\sigma_n - \sigma) = O_p(1), \quad \sqrt{n}(\mu_n - \mu) = O_p(1) \tag{36}$$

Note that

$$\Psi'(x) = \frac{1}{1+x^2}, \quad \Psi''(x) = \frac{-2x}{(1+x^2)^2} \quad (37)$$

hence

$$\begin{aligned} \frac{\partial \Phi(\sigma^{-1}(Y_j - \mu))}{\partial(\sigma, \mu)} &= -\Psi'(\sigma^{-1}(Y_j - \mu)) \begin{pmatrix} \sigma^{-2}Y_j \\ \sigma^{-1} \end{pmatrix} = \Delta_1(Y_j|\sigma, \mu) \\ \frac{\partial^2 \Psi(\sigma^{-1}(Y_j - \mu))}{\partial(\sigma, \mu) \partial(\sigma, \mu)'} &= \Psi''(\sigma^{-1}(Y_j - \mu)) \begin{pmatrix} \sigma^{-2}Y_j \\ \sigma^{-1} \end{pmatrix} (\sigma^{-2}Y_j, \sigma^{-1}) \\ &\quad + \Psi'(\sigma^{-1}(Y_j - \mu)) \begin{pmatrix} 2\sigma^{-3}Y_j & 0 \\ \sigma^{-2} & 0 \end{pmatrix} = \Delta_2(Y_j|\sigma, \mu) \end{aligned}$$

say. Note that due to (37), $\Delta_1(Y_j|\sigma, \mu)$ and $\Delta_2(Y_j|\sigma, \mu)$ are bounded in Y_j . Now by Taylor's theorem,

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

and similarly,

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

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

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

$$\begin{aligned}
& \times dF(y|X_j, \widehat{\theta}) \\
& = \sqrt{n}(\sigma_n - \sigma, \mu_n - \mu) \frac{1}{n} \sum_{j=1}^n \Delta_1(Y_j|\sigma, \mu) \exp(i\tau\Phi(\sigma(Y_j - \mu))) \\
& \quad \times \exp(i\xi X_j) \\
& \quad - \sqrt{n}(\sigma_n - \sigma, \mu_n - \mu) \frac{1}{n} \sum_{j=1}^n \int \Delta_1(y|\sigma, \mu) \exp(i\tau\Phi(\sigma^{-1}(y - \mu))) \\
& \quad \times dF(y|X_j, \widehat{\theta}) \exp(i\xi X_j) + O_p(1/\sqrt{n})
\end{aligned}$$

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

Under the null hypothesis, $\widehat{\theta} \xrightarrow{p} \theta_0$ and thus

$$\begin{aligned}
& \frac{1}{n} \sum_{j=1}^n \Delta_1(Y_j|\sigma, \mu) \exp(i\tau\Phi(\sigma^{-1}(Y_j - \mu))) \exp(i\xi X_j) \\
& \quad \xrightarrow{p} E \left[\int \Delta_1(y|\sigma, \mu) \exp(i\tau\Phi(\sigma^{-1}(y - \mu))) dF(y|X, \theta_0) \exp(i\xi X) \right], \\
& \frac{1}{n} \sum_{j=1}^n \int \Delta_1(y|\sigma, \mu) \exp(i\tau\Phi(\sigma^{-1}(y - \mu))) dF(y|X_j, \widehat{\theta}) \exp(i\xi X_j) \\
& \quad \xrightarrow{p} E \left[\int \Delta_1(y|\sigma, \mu) \exp(i\tau\Phi(\sigma^{-1}(y - \mu))) dF(y|X, \theta_0) \exp(i\xi X) \right],
\end{aligned}$$

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

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

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

Proof of Lemma 5

Using the trivial equality $(a + i.b)(c - i.d) = (a.c + b.d) - i.(a.d - b.c)$, 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

$$\int \operatorname{Re}(\psi_{j_1}(\beta)) \operatorname{Re}(\psi_{j_2}(\beta)) \mu(d\beta) + \int \operatorname{Im}(\psi_{j_1}(\beta)) \operatorname{Im}(\psi_{j_2}(\beta)) \mu(d\beta)$$

$$\begin{aligned}
&= \mathbf{1}(j_1 = j_2) \\
&\int \operatorname{Re}(\psi_{j_1}(\beta)) \operatorname{Im}(\psi_{j_2}(\beta)) \mu(d\beta) - \int \operatorname{Im}(\psi_{j_1}(\beta)) \operatorname{Re}(\psi_{j_2}(\beta)) \mu(d\beta) \\
&= 0
\end{aligned}$$

Consequently, denoting

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

it follows that

$$\int Q_{j_1}(\beta) Q_{j_2}(\beta) \mu(d\beta) = \mathbf{1}(j_1 = j_2) \cdot I_2 \quad (38)$$

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

$$\operatorname{Re}(g_j) = \int \operatorname{Re}(Z(\beta)) \operatorname{Re}(\psi_j(\beta)) \mu(d\beta) + \int \operatorname{Im}(Z(\beta)) \operatorname{Im}(\psi_j(\beta)) \mu(d\beta)$$

$$\operatorname{Im}(g_j) = \int \operatorname{Re}(Z(\beta)) \operatorname{Im}(\psi_j(\beta)) \mu(d\beta) - \int \operatorname{Im}(Z(\beta)) \operatorname{Re}(\psi_j(\beta)) \mu(d\beta)$$

Hence, denoting

$$G_{2,j} = \begin{pmatrix} \operatorname{Re}(g_j) \\ \operatorname{Im}(g_j) \end{pmatrix}, \quad Z_2(\beta) = \begin{pmatrix} \operatorname{Re}(Z(\beta)) \\ \operatorname{Im}(Z(\beta)) \end{pmatrix}$$

we can write

$$G_{2,j} = \int Q_j(\beta) Z_2(\beta) \mu(d\beta). \quad (39)$$

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

$$E[Z_2(\beta_1) Z_2(\beta_2)'] = \sum_{k=1}^{\infty} \lambda_k Q_k(\beta_1) Q_k(\beta_2). \quad (40)$$

It follows now from (38), (39) and (40) that

$$\begin{aligned}
E[G_{2,j_1} G_{2,j_2}'] &= \int \int Q_{j_1}(\beta_1) E[Z_2(\beta_1) Z_2(\beta_2)'] Q_{j_2}(\beta_2) \mu(d\beta_1) \mu(d\beta_2) \\
&= \int \int Q_{j_1}(\beta_1) \left(\sum_{k=1}^{\infty} \lambda_k Q_k(\beta_1) Q_k(\beta_2) \right) Q_{j_2}(\beta_2) \mu(d\beta_1) \mu(d\beta_2)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{\infty} \lambda_k \left(\int Q_{j_1}(\beta_1) Q_k(\beta_1) \mu(d\beta_1) \right) \left(\int Q_k(\beta_2) Q_{j_2}(\beta_2) \mu(d\beta_2) \right) \\
&= \sum_{k=1}^{\infty} \lambda_k \mathbf{1}(j_1 = k) \mathbf{1}(j_2 = k) I_2 \\
&= \begin{cases} \lambda_{j_1} I_2 & \text{if } j_1 = j_2, \\ O & \text{if } j_1 \neq j_2. \end{cases} \tag{41}
\end{aligned}$$

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 (41) 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 $N_2[0, I_2]$ distributed.

The last result in Lemma 5 follows from

$$\begin{aligned}
\int |Z(\beta)|^2 \mu(d\beta) &= \int Z(\beta) \overline{Z(\beta)} \mu(d\beta) \\
&= \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} g_{j_1} \int \psi_{j_1}(\beta) \overline{\psi_{j_2}(\beta)} \mu(d\beta) \overline{g_{j_2}} \\
&= \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} g_{j_1} \overline{g_{j_2}} \mathbf{1}(j_1 = j_2) \\
&= \sum_{j=1}^{\infty} g_j \overline{g_j} = \sum_{j=1}^{\infty} |g_j|^2 \\
&= \sum_{j=1}^{\infty} ((\operatorname{Re}(g_j))^2 + (\operatorname{Im}(g_j))^2).
\end{aligned}$$

Proof of Theorem 3

In view of Lemma 2 it suffices to show that

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

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

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

Then

$$\begin{aligned}
& \sup_{|c_2-c_1|<\delta} \left| \tilde{T}_n(c_2) - \tilde{T}_n(c_1) \right| = \frac{1}{(2\underline{c})^{m+k}} \sup_{|c_2-c_1|<\delta} \int_{\Pi(c_1, c_2)} |\tilde{Z}_n(\tau, \xi)|^2 d\tau d\xi \\
& + \sup_{|c_2-c_1|<\delta} \left(\frac{1}{(2c_1)^{m+k}} - \frac{1}{(2c_2)^{m+k}} \right) \int_{\Upsilon(\bar{c}) \times \Xi(\bar{c})} |\tilde{Z}_n(\tau, \xi)|^2 d\tau d\xi \\
& \xrightarrow{d} \frac{1}{(2\underline{c})^{m+k}} \sup_{|c_2-c_1|<\delta} \int_{\Pi(c_1, c_2)} |Z(\tau, \xi)|^2 d\tau d\xi \\
& + \sup_{|c_2-c_1|<\delta} \left(\frac{1}{(2c_1)^{m+k}} - \frac{1}{(2c_2)^{m+k}} \right) \int_{\Upsilon(\bar{c}) \times \Xi(\bar{c})} |Z(\tau, \xi)|^2 d\tau d\xi
\end{aligned}$$

hence for $0 < \delta < 1$,

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

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

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

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

Proof of Theorem 5

It follows similar to Lemma 2 that

$$\hat{T}_n^{(s)} = \int_{\Upsilon \times \Xi} |\tilde{Z}_n(\tau, \xi) - \tilde{Z}_n^{(s)}(\tau, \xi)|^2 d\mu(\tau, \xi) + o_p(1). \quad (42)$$

Denote by \mathcal{D} the σ -algebra generated by $\{(Y_j, X_j)\}_{j=1}^\infty$. Since \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 , $\tilde{Z}_n^{(s)} \Rightarrow Z_s$ conditional on \mathcal{D} , where $Z_s(\tau, \xi)$ is a zero-mean Gaussian process with covariance function

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

where $\varphi(\tau|X; \theta)$ is the conditional characteristic function of $F(y|X_1, \theta)$ [c.f. (8)]. Since Γ_s does not depend on \mathcal{D} it follows now that unconditionally,

$$\tilde{Z}_n^{(s)} \Rightarrow Z_s \text{ under } H_0 \text{ and } H_1, \quad (43)$$

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

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

as is not hard to verify. Hence by (42), (43) and the continuous mapping theorem,

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

The result under H_1 is easy to verify from (43) and Theorem 1.