Integrated Conditional Moment Tests for Parametric Conditional Distributions

Herman J. Bierens* and Li Wang
Department of Economics and CAPCP†
Pennsylvania State University
University Park, PA 16802

April 11, 2008

Abstract

This paper extends the Integrated Conditional Moment (ICM) test for the functional form of nonlinear regression models to tests for parametric conditional distributions. This test is formed on the basis of the integrated squared difference between the empirical characteristic function of the actual data and the characteristic function implied by the model. This test is consistent, and has nontrivial power against $\sqrt{n}$ -local alternatives.

To avoid numerical evaluation of the conditional characteristic function of the model distribution, a simulated integrated conditional moment (SICM) test is proposed, where each theoretical conditional characteristic function is replaced by a simulated counterpart, based on a single random drawing from the corresponding conditional distribution. All the properties of the exact ICM test carry over to the SICM test.

*A previous version of this paper was presented by Herman Bierens at the "Symposium on Nonparametric Testing in Econometrics", March 21, 2008, Indiana University. The helpful comments of Juan Carlos Escanciano 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

Simulation results show that this method works well. As an empirical application we test the validity of the conditional Poisson distribution for count data, using a health economics related data set. Our test firmly rejects this model.

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 in the estimation and inference of an econometric model.

The specification tests proposed in the literature address various issues, such as conditional independence and conditional endogeneity. The current paper addresses the problem of testing the validity of a parametric conditional distribution specification. The approach in this paper is based on the well-known fact that two distribution functions are the same if and only if their characteristic functions are the same. This paper generalizes the Integrated Conditional Moment (ICM) test proposed by Bierens (1982) and Bierens and Ploberger (1997) to an ICM test of the correctness of parametric specifications of conditional distributions.

This test is consistent, i.e., it has asymptotic power one against all alternatives. In addition, it has non-trivial power against $\sqrt{n}$ local alternatives, which is faster than for Zheng's (2000) test, which is based on a comparison of a parametric density specification with a corresponding nonparametric kernel density estimator. The critical values can be obtained by a parametric bootstrap method.

This paper is organized as follows. The relevant literature is reviewed in Section 2. In Section 3 we introduce our ICM test for conditional distributions, and in Section 4 we propose a Simulated ICM (SICM) test. In Section 5 we conduct a simulation study similar to the setup of the Monte Carlo analysis in Zheng (2000), showing that the ICM test has good finite-sample properties. In Section 6 we apply the SICM test to a conditional Poisson model, using health economics data. In Section 7 we make concluding remarks. Section 8 is a technical appendix containing the proofs.

Throughout we denote the indicator function by $I(.)$, weak convergence is denoted by $\Rightarrow$, and convergence in distribution and probability by $\stackrel{d}{\to}$ and $\stackrel{p}{\to}$, respectively. Moreover, the norm $||z||$ of a complex vector $z \in \mathbb{C}^p$ is defined as

$$||z|| = \sqrt{z'\overline{z}} = \sqrt{(\mathrm{Re}(z))'(\mathrm{Re}(z)) + (\mathrm{Im}(z))'(\mathrm{Im}(z))},$$

where here and in the sequel the bar above a complex variable, vector or function denotes the complex conjugate.[1] In the case $z = a + i.b \in \mathbb{C}$ the norm $||z||$ becomes the absolute value $|z| = \sqrt{a^2 + b^2}$. Finally, the matrix norm $||A||$ is the maximum absolute value of the elements of A.

2 Literature Review

The Hausman (1978) 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. White (1982) compared two different expressions of the Fisher information matrix, which should be equal if the conditional distribution is correctly specified.

Newey (1985) proposed the conditional moment (CM) test and suggested that both the Hausman and the White 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.

Specific alternatives are also studied in the literature on specification testing, such as selecting a particular competing model. See Vuong (1989) and Fan and Li (1996) for examples.

[1] Recall that if $z = a + i.b$ with a and b real-valued then $\overline{z} = a - i.b$.

The first consistent specification test for nonlinear regression models was proposed by Bierens (1982). This test is also a conditional moment tests, but based on uncountable many moments conditions of the form $E[U.w(\xi'X)] = 0$, where U is the regression model error, X is a bounded vector of explanatory variables, and $w(.)$ is a weight function. The key result that delivers the consistency is Theorem 1 in Bierens and Ploberger (1997), which is a generalization of earlier versions in Bierens (1982, 1990):

Lemma 1. *Let U be a random variable satisfying $E[|U|] < \infty$ and let X* be a *bounded random vector in $\mathbb{R}^k$ such that $P[E(U|X) = 0] < 1$. Let $w(u)$ be a function on $\mathbb{R}$ which is infinitely many times differentiable and satisfies $d^m w(u)/(du)^m|_{u=0} \neq 0$ for all but a finite number of nonnegative integers m. Then the set $S = \{\xi \in \mathbb{R}^k : E[U.w(\xi'X)] = 0\}$ has Lebesgue measure zero and is nowhere dense.*

Examples of such functions $w(u)$ are $w(u) = \exp(i.u) = \cos(u) + i.\sin(u)$, where $i = \sqrt{-1}$, (used in Bierens 1982, 1984), $w(u) = \exp(u)$ (used in Bierens 1990) and $w(u) = \cos(u) + \sin(u)$. Stinchcombe and White (1998) have shown that Lemma 1 holds for a wide range of non-polynomial analytical weight functions $w(.)$, for which they cast the name "totally revealing".

Under H_0, $E[U.w(\xi'X)] = 0$ for all ξ, and Lemma 1 states that under H_1, the same expression is nonzero except for at most a non-dense set with Lebesgue measure zero. Therefore, given a random sample $(Y_1, X_1), ..., (Y_n, X_n)$ from (Y, X), Bierens (1982) and Bierens and Ploberger (1997) propose to test the validity of the nonlinear regression model via the ICM test statistic

$$\widehat{T} = \int_{\Xi} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \widehat{U}_j w(\xi' X_j) \right|^2 d\mu(\xi),$$

where the $\widehat{U}_j$'s are nonlinear least squares residuals, Ξ is a compact subset of $\mathbb{R}^k$ and $\mu(.)$ is an absolutely continuous probability measure with support Ξ. Bierens and Ploberger (1997) showed that the ICM test has nontrivial $\sqrt{n}$ local power and is admissible. Boning and Sowell (1999) showed that this ICM test is the best ICM test according to the weighted average power criterion considered by Andrews and Ploberger (1994).

A somewhat related specification test for regression models has been proposed by Stute (1997), based on an empirical process of the form

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widehat{U}_j I(X_j \leq x).$$

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 Hardle and Mammen (1993), Zheng (1996) and Hong and White (1995). However, these tests have only nontrivial power against local alternatives that aproach 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) discovered that the ICM test can be viewed as a special case of the kernel-based tests but with a fixed bandwidth.

The literature on consistent specification testing of conditional distributions is rather limited. Zheng (2000) proposed a test for the validity of conditional densities by comparing a parametric conditional density with a corresponding nonparametric kernel estimator via the Kullback-Leibler (1951) information criterion. Andrews (1988) extended the Pearson's Chi-square test to a test for parametric conditional distributions. This test is based on partitioning the the dependent and explanatory variables in cells, and then comparing the frequencies involved with the frequencies implied by the model. However, 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 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.[2] In particular, the CK test statistic takes the form

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

[2]The characteristic function approach is used by Pinkse (1998) in a nonparametric test of serial independence, and by Su and White (2006) in a non-parametric test of distributional independence without assuming parametric distributions.

where $F(y|X_j, \widehat{\theta})$ is the estimated conditional distribution model. Since the asymptotic null distribution is case-dependent, a bootstrap method is used to derive critical values. This test is consistent, and has nontrivial power against $\sqrt{n}$ local alternatives. However, a practical problem with the CK test is that if the dimension of X_j is large the inequality $X_j < X_i$ for $i \neq j$ may never happen, even for quite a large sample size n. This appears to be the case in our empirical application in Section 6, where $X_j \in \mathbb{R}^{16}$ and $n = 4406$. Then effectively the CK test statistic becomes $\max_{1 \leq j \leq n} \left|1 - F(Y_j|X_j, \widehat{\theta})\right| / \sqrt{n}$. This problem does not happen with our ICM test.

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 conditional distribution function yields an 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. In particular, with $F(y|X_j, \widehat{\theta})$ the estimated conditional distribution of Y_j given X_j, Bai's test is based on an empirical process of the form

$$\widehat{V}(u) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(I\left(F(Y_j|X_j, \widehat{\theta}) \leq u\right) - u\right), \ u \in [0,1].$$

However, Bai's test is not consistent, as we will show by the following counterexample. Let X be uniformly $[0,1]$ distributed and let Y be a nonnegative random variable. Suppose that the true conditional distribution of Y given X is exponential: $F_1(y|X) = 1 - \exp(-y/X)$. Let the incorrect null model be $F_0(y|X) = G(y/X^2)$, where $G(z)$ is an absolutely continuous distribution function on $(0,\infty)$ with inverse G^{-1}. Then for $u \in [0,1]$,

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

hence

$$\Pr\left[F_0(Y|X) \leq u\right] = 1 - \int_0^1 \exp\left(-x.G^{-1}(u)\right) dx = 1 - \frac{1 - \exp\left(-G^{-1}(u)\right)}{G^{-1}(u)}.$$

Now let $G(z) = 1 - (1 - \exp(-z))/z$, which is a distribution function on $(0,\infty)$ with density $g(z) = z^{-2}\exp(-z).(\exp(z) - 1 - z)) > 0$. Then it is trivial that $\Pr[F_0(Y|X) \leq u] = u$, which demonstrates that Bai's test is not consistent.

Admittedly, Bai (2003) did not claim consistency, but only that his test has non-trivial power against $\sqrt{n}$ local alternatives. Thus, our example demonstrates that nontrivial local power does not imply consistency against fixed alternatives.

3 The ICM Test for Conditional Distributions

3.1 Quasi-Maximum Likelihood Conditions

Throughout we will assume that

Assumption 1. *We observe a random sample* $(Y_1, X_1), ..., (Y_n, X_n)$ *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 parameter space* Θ *is chosen such that for all* $\theta \in \Theta$ *the support of* $F(y|X;\theta)$ *is the same as the support of the actual conditional distribution of* Y *given* X. *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 (quasi-) maximum likelihood estimator* $\widehat{\theta} = \arg\max_{\theta\in\Theta} \ln L_n(\theta)$ *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 notation*[3]

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

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

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

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

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

Note that nothing is said about whether the parametric specification $F(y|X;\theta_0)$ is correct or not. In either case Assumption 1 implicitly impose standard regularity conditions on $F(y|X;\theta)$ such that sufficient conditions for the consistency and asymptotic normality of quasi-maximum likelihood estimators are satisfied. See White (1982, 1994).

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

H_0 : *There exists a* $\theta \in \Theta$ *such that* $\Pr[Y \leq y|X] = F(y|X;\theta)$ *a.s. for all* $y \in \mathbb{R}^m$,

and the alternative hypothesis is that H_0 is incorrect*:*

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

Then Assumption 1 implies that under H_0, $\Pr[Y \leq y|X] = F(y|X;\theta_0)$ a.s. for all $y \in \mathbb{R}^m$, so that then $A = B$ and $\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{d} N_p\left[0, A^{-1}\right]$, whereas under H_1, $\sup_{y\in\mathbb{R}^m} |\Pr[Y \leq y|X] - F(y|X;\theta_0)| > 0$.

3.2 Model Verification Via Characteristic Functions

The proposed ICM test is based on the comparison of the actual conditional characteristic function $E\left[\exp(i.\tau'Y)|X\right]$ 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

$$\Pr\left(E\left[\exp(i.\tau'Y)|X\right] - \int \exp(i\tau'y)dF(y|X,\theta_0) = 0\right) = 1$$

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

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

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

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

for some points $(\tau, \xi) \in \mathbb{R}^m \times \mathbb{R}^k$. The question now arises where to look for these points in the space $\mathbb{R}^m \times \mathbb{R}^k$. The answer is the following:

Lemma 2. *Let Y be a bounded m-dimensional random vector, let X be a bounded k-dimensional random vector, and let N_o be an arbitrary open neighborhood of the origin of $\mathbb{R}^m \times \mathbb{R}^k$. Under H_1 the set*

$$\begin{aligned}\mathcal{C} &= \left\{(\tau, \xi) \in N_o : \left| E\left[\left(\exp\left(i\tau' Y\right) - \int \exp(i.\tau' Y) dF(y|X, \theta_0)\right)\right.\right.\right. \\ &\qquad \left.\left.\left. \times \exp(i.\xi' X)\right]\right| > 0\right\}\end{aligned}$$

has positive Lebesgue measure. If the random vectors Y and X are not bounded, replace them in the complex exp *functions by bounded one-to-one mappings, $\Phi_1(Y)$ and $\Phi_2(X)$, respectively. Then under H_1 the set*

$$\begin{aligned}\mathcal{C} = &\left\{(\tau, \xi) \in N_o : \left| E\left[\left(\exp\left(i.\tau' \Phi_1(Y)\right) - \int \exp\left(i.\tau' \Phi_1(y)\right) dF(y|X, \theta_0)\right)\right.\right.\right. \\ &\qquad \left.\left.\left. \times \exp\left(i\xi' \Phi_2(X)\right)\right]\right| > 0\right\}\end{aligned}$$

has positive Lebesgue measure.

Proof. Appendix.

It can actually be shown along the lines of the proof of Lemma 1 that the Lebesgue measure of $\mathcal{C}$ is the same as the Lebesgue measure of N_o, but Lemma 2 suffices for our purpose.

For the time being we will assume that Y and X are bounded random vectors. Once we have completed the introduction of the ICM test for conditional distributions and its asymptotic properties we will show that these bounded transformation only lead to a few minor changes in the notation. See Remark 1 following Theorem 1 below.

Denote

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

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

$$\Xi = \times_{j=1}^{k} \left[-\overline{\xi}_j, \overline{\xi}_j\right], \ \overline{\xi}_j > 0 \qquad (2)$$

and let μ be the uniform probability measure on $\Upsilon \times \Xi$:

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

Bonning and Sowell (1999) have show that in the case of the ICM test of Bierens and Ploberger (1997) this uniform measure is optimal. Then it follows from Lemma 2 that under H_1,

$$\int_{\Upsilon}\int_{\Xi}\left|\varsigma\left(\tau,\xi;\theta\right)\right|^2 d\mu(\tau,\xi) > 0 \text{ for all } \theta \in \Theta$$

whereas of course under H_0,

$$\int_{\Upsilon}\int_{\Xi}\left|\varsigma\left(\tau,\xi;\theta_0\right)\right|^2 d\mu(\tau,\xi) = 0$$

This suggests that similar to Bierens and Ploberger (1997) the null hypothesis can be tested consistency by an ICM test of the form

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

where

$$Z_n(\tau,\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp\left(i\tau' Y_j\right) - \int \exp\left(i\tau' y\right) dF(y|X_j,\hat{\theta})\right)\exp\left(i\xi' X_j\right). \tag{3}$$

is a complex-valued continuous empirical process.

3.3 Asymptotic Properties

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

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

$$\varphi\left(\tau|X;\theta\right) = \int \exp\left(i.\tau' y\right) dF(y|X,\theta), \tag{4}$$

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

$$E\left[\sup_{\theta\in\Theta_0}\|\Delta\varphi(\tau|X;\theta)\|\right] < \infty.$$

Then

Lemma 3. *Under Assumptions 1-2,*

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

pointwise in (τ,ξ), where

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

Proof: Appendix.

Consequently, denoting

$$\begin{aligned}\phi(\tau,\xi|Y,X) &= (\exp(i.\tau'Y) - \varphi(\tau|X;\theta_0))\exp(i.\xi'X) \\ &\quad + b(\tau,\xi)'A^{-1}\Delta\ell(Y|X;\theta_0) \qquad (5)\\ \widetilde{Z}_n(\tau,\xi) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi(\tau,\xi|Y_j,X_j), \qquad (6)\\ \widetilde{T}_n &= \int_{\Upsilon}\int_{\Xi}|\widetilde{Z}_n(\tau,\xi)|^2 d\mu(\tau,\xi)\end{aligned}$$

it follows from Lemma 3 that

Lemma 4. *Under Assumptions 1-2,*

$$\widehat{T}_n = \widetilde{T}_n + o_p(1) \qquad (7)$$

r*egardless 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,*

$$\widetilde{Z}_n \Rightarrow Z \text{ on } \Upsilon \times \Xi, \tag{8}$$

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

$$\Gamma\left((\tau_1, \xi_1), (\tau_2, \xi_2)\right) = E\left[\phi(\tau_1, \xi_1 | Y, X)\overline{\phi(\tau_2, \xi_2 | Y, X)}\right], \tag{9}$$

hence by Lemma 3 and the continuous mapping theorem,

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

Moreover, under H_1,

$$p \lim_{n\to\infty} \widetilde{Z}_n(\tau, \xi)/\sqrt{n} = E\left[\left(\exp\left(i.\tau' Y\right) - \varphi\left(\tau | X; \theta_0\right)\right)\exp\left(i.\xi' X\right)\right] = \eta(\tau, \xi)$$

say, where the set $\{(\tau, \xi) \in \Upsilon \times \Xi : \eta(\tau, \xi) > 0\}$ has positive Lebesgue measure, hence by Lemma 3 and bounded convergence,

$$p \lim_{n\to\infty} \widehat{T}_n/n = \int_\Upsilon \int_\Xi |\eta(\tau, \xi)|^2 d\mu(\tau, \xi) > 0. \tag{10}$$

Proof: Appendix.

Remark 1. The condition that Y and X are bounded random vectors is not essential, because in view of Lemma 2 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 $\varphi\left(\tau | X; \theta\right)$, $b(\tau, \xi)$ and $\phi(\tau, \xi | Y, X)$ as

$$\begin{aligned}
\varphi\left(\tau | X; \theta\right) &= \int \exp\left(i.\tau' \Phi_1(y)\right) dF(y | X, \theta), && (11)\\
b(\tau, \xi) &= E\left[\Delta\varphi\left(\tau | X; \theta_0\right)\exp\left(i.\xi' \Phi_2(X)\right)\right] && (12)\\
\phi(\tau, \xi | Y, X) &= \left(\exp\left(i.\tau' \Phi_1(Y)\right) - \varphi\left(\tau | X; \theta_0\right)\right)\exp\left(i.\xi' \Phi_2(X)\right) \\
&\quad - b(\tau, \xi)' A^{-1} \sum_{j=1}^{n} \Delta\ell\left(Y | X; \theta_0\right). && (13)
\end{aligned}$$

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 the next section. 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.

Remark 2. If $\int \exp(i.\tau' y)dF(y|X,\theta)$ is analytic,[4] no bounded transformation of Y is needed for the consistency result (10), although we still need a bounded 1-1 transformation $\Phi_2(X)$ of X if X is not already bounded. However, for proving (8) we need a bounded 1-1 transformation $\Phi_1(Y)$ of Y if Y is not already bounded, except if the moment generating function of $||Y||^2$ is everywhere finite. See the proof of Theorem 1 for the latter.

3.4 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. In particular, let $\Phi_1(Y) = \arctan\left(\sigma_n^{-1}(Y-\mu_n)\right)$, where μ_n and $\sigma_n > 0$ are location and scale parameters. For example, choose for μ_n the sample mean and for σ_n the sample standard error of Y. Alternatively, choose μ_n and σ_n such that most of the observations $\sigma_n^{-1}(Y_j - \mu_n)$ fall in the interval $[-1, 1]$, because in this interval the $\arctan(y)$ function has still substantial variation. In particular, let

$$\mu_n = \frac{1}{2}\left(Q_n(0.95) + Q_n(0.05)\right),\ \sigma_n = \frac{1}{2}\left(Q_n(0.95) - Q_n(0.05)\right), \qquad (14)$$

[4] An analytic function is one that has a Taylor series expansion at a neighborhood of every point and is everywhere infinitely differentiable. If a characteristic function $\varphi(t)$, $t \in \mathbb{R}^m$, is analytic, it is completely determined by its shape in an arbitrary open neighborhood of the origin of $\mathbb{R}^m$. Examples of distributions with analytic characteristic functions are the normal distribution, the Gamma distribution, and the Poisson distribution, among others. See Lukacs (1970).

where

$$Q_n(\alpha) = \arg \max_{\frac{1}{n}\sum_{j=1}^{n} I(Y_j \leq y) \leq \alpha} y \tag{15}$$

is the $\alpha \times 100\%$ sample quantile of Y. Then $\frac{1}{n}\sum_{j=1}^{n} I\left(\left|\sigma_n^{-1}(Y_j - \mu_n)\right| \leq 1\right) \approx 0.9$.

More generally, standardize each component $Y_{\ell_1,j}$ of $Y_j \in \mathbb{R}^m$ and $X_{\ell_2,j}$ of $X_j \in \mathbb{R}^k$ by $\widehat{Y}_{\ell_1,j} = \widehat{\sigma}_{1,\ell_1}^{-1}\left(Y_{\ell_1,j} - \widehat{\mu}_{1,\ell_1}\right)$ and $\widehat{X}_{\ell_2,j} = \widehat{\sigma}_{2,\ell_2}^{-1}\left(X_{\ell_2,j} - \widehat{\mu}_{2,\ell_2}\right)$, respectively, and denote $\overline{Y}_{\ell,j} = \sigma_{1,\ell}^{-1}\left(Y_{\ell,j} - \mu_{1,\ell}\right)$, $\overline{X}_{\ell,j} = \sigma_{2,\ell}^{-1}\left(X_{\ell,j} - \mu_{2,\ell}\right)$, where

Assumption 3. *For $i = 1, 2$, μ_{i,ℓ_i} and $\sigma_{i,\ell_i} > 0$ are constants such that $\sqrt{n}(\widehat{\mu}_{i,\ell_i} - \mu_{i,\ell_i}) = O_p(1)$ and $\sqrt{n}(\widehat{\sigma}_{i,\ell_i} - \sigma_{i,\ell_i}) = O_p(1)$.*

These conditions hold for sample means and sample standard errors provided that the variables involved have finite fourth moments, and under mild additional conditions for quantiles as well. As to the latter, let $Q(\alpha) = \arg\max_{\Pr(Y_j \leq y) \leq \alpha} y$ be the $\alpha \times 100\%$ quantile of the marginal distribution $F_Y(y) = \Pr(Y_j \leq y)$. It is not hard to verify that $\sqrt{n}\left(Q_n(\alpha) - Q(\alpha)\right) = O_p(1)$ if $F_Y(y)$ is continuously differentiable in $y = Q(\alpha)$ with positive derivative $F_Y'(Q(\alpha))$, or if $F_Y(y)$ is discrete.

Next, denote

$$\begin{aligned}
\Phi_1(Y_j) &= \left(\arctan\left(\overline{Y}_{1,j}\right), ..., \arctan\left(\overline{Y}_{m,j}\right)\right)', \\
\widehat{\Phi}_1(Y_j) &= \left(\arctan\left(\widehat{Y}_{1,j}\right), ..., \arctan\left(\widehat{Y}_{m,j}\right)\right)', \\
\Phi_2(X_j) &= \left(\arctan\left(\overline{X}_{1,j}\right), ..., \arctan\left(\overline{X}_{k,j}\right)\right)', \\
\widehat{\Phi}_2(X_j) &= \left(\arctan\left(\widehat{X}_{1,j}\right), ..., \arctan\left(\widehat{X}_{k,j}\right)\right)'.
\end{aligned}$$

Finally, redefine $Z_n(\tau, \xi)$ as

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

and denote

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

$$\times \exp\left(i\xi'\widehat{\Phi}_2\left(X_j\right)\right).$$

Then the following results hold.

Lemma 5. *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.

3.5 The Null Distribution

To analyze the limiting null distribution of $\hat{T}_n$ along the lines in Bierens and Ploberger (1997) we need a generalized version of Mercer's theorem for complex-valued symmetric positive 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 $\mathbb{R}^m$ is a Borel measurable function $\Gamma(\beta_1,\beta_2):\mathbb{R}^q\times\mathbb{R}^q\to\mathbb{C}$, such that for any complex-valued Borel measurable function $\psi(\beta)$,

$$\int\int\psi(\beta_1)\Gamma(\beta_1,\beta_2)\overline{\psi(\beta_2)}d\mu(\beta_1)d\mu(\beta_2)\geq 0.$$

where the bar of $\overline{\psi(\beta_2)}$ denotes the complex-conjugate of $\psi(\beta_2)$. The covariance function (9) 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 (9) 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 6. *Let μ be a probability measure with compact support $\mathbf{B}\subset\mathbb{R}^q$, and let $\Gamma(\beta_1,\beta_2):\mathbf{B}\times\mathbf{B}\to\mathbb{C}$ be a symmetric and continuous positive semi-definite function relative to μ. Consider the eigenvalue equation $\lambda.\psi(\beta_1)=$*

$\int \Gamma(\beta_1, \beta_2)\overline{\psi(\beta_2)}d\mu(\beta_2)$, where λ is an eigenvalue with corresponding eigenfunction $\psi(.)$. This eigenvalue equation has countable-infinite many solutions, $\lambda_j.\psi_j(\beta_1) = \int \Gamma(\beta_1, \beta_2)\overline{\psi_j(\beta_2)}d\mu(\beta_2)$, $j = 1, 2, 3, ...$ The eigenvalues λ_j are real-valued[5] *and nonnegative and satisfy $\sum_{j=1}^{\infty} \lambda_j < \infty$. The eigenfunctions $\psi_j(.)$ are complex-valued and continuous on $\mathbf{B}$, and can be chosen orthonormal, i.e., $\int \psi_{j_1}(\beta)\overline{\psi_{j_2}(\beta)}d\mu(\beta) = I(j_1 = j_2)$. The function Γ has the series representation $\Gamma(\beta_1, \beta_2) = \sum_{j=1}^{\infty} \lambda_j \psi_j(\beta_1)\overline{\psi_j(\beta_2)}$.*[6] *Moreover, every complex-valued continuous function ϕ on $\mathbf{B}$ can be written as $\phi(\beta) = \sum_{j=1}^{\infty} g_j \psi_j(\beta)$, where $g_j = \int \phi(\beta)\overline{\psi_j(\beta)}d\mu(\beta)$ satisfying $\sum_{j=1}^{\infty} |g_j|^2 < \infty$.*

Proof: See Hadinejad-Mahram et al. (2002) and Krein (1998).

Following a similar arguments as in Bierens and Ploberger (1997), it can be shown on the basis of the results in Lemma 6 that

Lemma 7. *Under H_0 and the conditions of Theorem 1,*

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

where the ε_j's are i.i.d. $N(0, 1)$ and the λ_j's are the eigenvalues of the covariance function (9).

Proof: Appendix.

Consequently, the distribution of T depends on the covariance function Γ, which in its turn depends on the conditional distribution function $F(y|X; \theta_0)$ of Y and the distribution of X. Nevertheless, critical values can be obtained via a parametric bootstrap method, as discussed in the last subsection of this section.

An alternative approach is to construct a standardized version of the ICM test for which the limiting null distribution has a case independent upper bound, similar to the ICM test of Bierens and Ploberger (1997). However, that approach requires a consistent estimator of

$$\int_{\Upsilon} \int_{\Xi} \Gamma((\tau, \xi), (\tau, \xi)) d\mu(\tau, \xi) = \int_{\Upsilon} \int_{\Xi} E[|Z(\tau, \xi)|^2] d\mu(\tau, \xi) = \sum_{j=1}^{\infty} \lambda_j,$$

[5] Due to the symmetry of Γ.

[6] This result is the actual Mercer's theorem.

which in the case under review is much more complicated to construct than in the regression case considered in Bierens and Ploberger (1997). Therefore, we will not explore that route.

3.6 Local Power

Let $Q(y|X)$ be a conditional distribution function that is not identically equal to $F(y|X,\theta_0)$, and consider the $\sqrt{n}$-local alternative

$$\begin{aligned} F_n(y|X,\theta_0) &= F(y|X,\theta_0) + (Q(y|X) - F(y|X,\theta_0))/\sqrt{n} \\ &= \left(1 - n^{-1/2}\right) F(y|X,\theta_0) + n^{-1/2} Q(y|X) \end{aligned}$$

It follows straightforwardly from (5) that under this local alternative,

$$E\left[\phi(\tau,\xi|Y,X)\right] = n^{-1/2}\left(\varphi_Q(\tau,\xi) - \varphi_F(\tau,\xi)\right)$$

where

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

hence,

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

On the other hand, the covariance function $\Gamma(.,.)$ of Z under the local alternative is the same as under the null hypothesis.

By the same argument as in Bierens and Ploberger (1997) it follow now from Lemma 6 that under the local alternative,

$$\int_\Upsilon \int_\Xi |Z(\tau,\xi)|^2 d\mu(\tau,\xi) \sim \sum_{j=1}^\infty (\eta_j + \varepsilon_j \sqrt{\lambda_j})^2 = T_{alt},$$

say, where the η_j's are the Fourier coefficient of $E[Z(\tau,\xi)]$ under the local alternative:

$$\eta_j = \int_\Upsilon \int_\Xi \left(\varphi_Q(\tau,\xi) - \varphi_F(\tau,\xi)\right) \overline{\psi}_j(\tau,\xi) d\mu(\tau,\xi).$$

Since $\varphi_Q(\tau,\xi)$ and $\varphi_F(\tau,\xi)$ are not identical on $\Upsilon \times \Xi$, at least one η_j is nonzero. Then by Corollary 1 in Bierens and Ploberger (1997), $\Pr(T_{alt} > \omega) > \Pr(T > \omega)$ for all $\omega > 0$, which implies that the ICM test has nontrivial power against $\sqrt{n}$-local alternatives.

3.7 Maximizing the ICM Test over the Integration Domain

The choice of the hypercubes Υ and Ξ defined by (1) and (2), 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 (1) and $\underline{\Xi}$ and $\overline{\Xi}$ are given hypercubes in $\mathbb{R}^k$ of the type (2), provided that it can be shown that under the null hypothesis,

$$\sup_{\underline{\Upsilon}\subset\Upsilon\subset\overline{\Upsilon},\ \underline{\Xi}\subset\Xi\subset\overline{\Xi}} \frac{\int_\Upsilon \int_\Xi |Z_n(\tau,\xi)|^2 d\tau d\xi}{\int_\Upsilon 1 d\tau \int_\Xi 1 d\xi} \xrightarrow{d} \sup_{\underline{\Upsilon}\subset\Upsilon\subset\overline{\Upsilon},\ \underline{\Xi}\subset\Xi\subset\overline{\Xi}} \frac{\int_\Upsilon \int_\Xi |Z(\tau,\xi)|^2 d\tau d\xi}{\int_\Upsilon 1 d\tau \int_\Xi 1 d\xi} \tag{16}$$

Note that the denominator $\int_\Upsilon 1 d\tau \int_\Xi 1 d\xi$ is the Lebesgue measure of $\Upsilon \times \Xi$.

Indeed, (16) is true, as will be shown for the following special case:

Lemma 8. *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*

$$\begin{aligned} \widehat{T}_n(c) &= \frac{1}{(2c)^{m+k}} \int_{\Upsilon(c)} \int_{\Xi(c)} |Z_n(\tau,\xi)|^2 d\tau d\xi, \qquad (17) \\ T(c) &= \frac{1}{(2c)^{m+k}} \int_{\Upsilon(c)} \int_{\Xi(c)} |Z(\tau,\xi)|^2 d\tau d\xi \end{aligned}$$

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

Proof: Appendix

Although it is too much of a computational burden to compute this supremum exactly, let alone the supremum in (16), 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. However, in the Monte Carlo study and the empirical application we will apply the ICM test for a few values of c separately.

3.8 Parametric Bootstrap

Since the seminal work by Efron (1979), bootstrap has become a popular method for deriving null distributions of tests, especially if the null distribution cannot be derived analytically or is case dependent. 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 $\widehat{\theta}$, generate M bootstrap samples

$$\left\{(\widetilde{Y}_{b,1}, X_1), ..., (\widetilde{Y}_{b,n}, X_n)\right\},\ b = 1, ..., M,$$

where $\widetilde{Y}_{b,j}$ is a random drawing from $F(y|X_j;\widehat{\theta})$ in bootstrap sample b. The vectors X_j of covariates are the same as in the actual sample. Let $\widetilde{\theta}_b$ be the ML estimator on the basis of this bootstrap sample, i.e.,.

$$\widetilde{\theta}_b = \arg\max_{\theta\in\Theta} \ln L_{b,n}(\theta)$$

where

$$\ln L_{b,n}(\theta) = \sum_{j=1}^{n} \ell(\widetilde{Y}_{b,j}|X_j;\theta)$$

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

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

where

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

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

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

Assumption 4. *Let $G(x)$ be the distribution function of X_j. The function*

$$\kappa(\theta_1,\theta_2) = \int\int \ell(y|x;\theta_1)dF(y|x;\theta_2)dG(x)$$

is continuous on $\Theta \times \Theta$. Moreover,

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

and

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

Furthermore, $\theta_0 = \arg\max_{\theta\in\Theta}\kappa(\theta,\theta_0)$,[7] *where $\theta_0 = p\lim_{n\to\infty}\widehat{\theta}$.*

Then it follows from Assumption 1 that

$$p\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|n^{-1}\ln L_{b,n}(\theta) - \kappa(\theta,\theta_0)\right| = 0$$

which in its turn implies[8] that

$$\widetilde{\theta}_b = \arg\max_{\theta\in\Theta}\ln L_{b,n}(\theta)/n \xrightarrow{p} \theta_0$$

[7] Recall from maximum likelihood theory that in general, $\kappa(\theta_2,\theta_2) \geq \kappa(\theta_1,\theta_2)$ for all $(\theta_1,\theta_2) \in \Theta\times\Theta$. See for example Bierens (2004, Theorem 8.1).

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

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

$$\int \Delta\ell\,(y|X;\theta)\,dF(y|X;\theta) = (\partial/\partial\theta')\int \ell\,(y|X;\theta)\,dF(y|X;\theta) = 0 \qquad (18)$$

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

$$E\left[\sup_{\theta\in\Theta_0}\left\|\int \Delta\ell\,(y|X_j;\theta)\,dF(y|X_j;\theta)\right\|^{2+\delta}\right] < \infty \qquad (19)$$

$$E\left[\sup_{\theta\in\Theta_0}\left\|\int \Delta^2\ell\,(y|X_j;\theta)\,dF(y|X_j;\theta)\right\|\right] < \infty \qquad (20)$$

Note that Assumption 1 and part (18) of Assumption 5 imply that

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

Next, let

$$\mathcal{D}_n = \sigma\left(\{(Y_j, X_j)\}_{j=1}^n\right)$$

be the σ-algebra generated by the current sample. Then conditional on $\mathcal{D}_n$,

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

is a double array of independent random vectors, for which Liapounov's central limit theorem applies. See for example Chung (1974, p. 200). This is the reason for the δ in (19). 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 $\sigma^2 = p\lim_{n\to\infty}\sigma_n^2 = \xi' B\xi$. This result can also be proved using a martingale difference central limit theorem. Thus,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Delta\ell\left(\widetilde{Y}_{b,j}|X_j;\widehat{\theta}\right)\xrightarrow{d}N_p[0,B], \tag{21}$$

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

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

where A is the same as in Assumption 1.

It is now easy to verify that similar to Lemma 3 the following result holds.

Lemma 9. *Under Assumptions 1-2 and 4-5,*

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

pointwise in (τ,ξ), *where* $b(\tau,\xi)$ is the same as in Lemma 3.

Consequently, denoting

$$\begin{aligned}\phi_{b,j}(\tau,\xi|\widehat{\theta}) &= \left(\exp\left(i.\tau'\widetilde{Y}_{b,j}\right)-\varphi\left(\tau|X_j;\widehat{\theta}\right)\right)\exp\left(i.\xi'X_j\right) \qquad (22)\\ &\quad +b(\tau,\xi)'A^{-1}\Delta\ell\left(\widetilde{Y}_{b,j}|X_j;\widehat{\theta}\right)\\ \widetilde{Z}_{b,n}(\tau,\xi) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_{b,j}(\tau,\xi|\widehat{\theta}), \qquad (23)\\ \widetilde{T}_{b,n} &= \int_{\Upsilon}\int_{\Xi}|\widetilde{Z}_{b,n}(\tau,\xi)|^2d\mu(\tau,\xi)\end{aligned}$$

it follows from Lemma 9 that

Lemma 10. *Under Assumptions 1-2 and 4-5,* $\widehat{T}_{b,n}=\widetilde{T}_{b,n}+o_p\left(1\right)$.

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

Let

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

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

Since $b(\tau,\xi)$ is uniformly continuous on $\Upsilon \times \Xi$, the tightness of $\widetilde{Z}_{2,b,n}$ follows from (21). The tightness of $\widetilde{Z}_{1,b,n}$ follows from the proof of Theorem 1, simply by replacing the expectations involved by the corresponding conditional expectations $E[.|\mathcal{D}_n]$. Moreover, similar to it (21) can be shown that the finite distributions of $\widetilde{Z}_{1,b,n}(\tau,\xi)$ converge to a multivariate normal distribution. Thus, $\widetilde{Z}_{b,n} \Rightarrow Z_*$. Furthermore, it is easy to verify that the covariance function of this limiting process takes the form

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

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

Theorem 2. *Let Y and X be bounded random vectors.*[9] *Then under Assumptions 1-2 and 4-5, $\widetilde{Z}_{b,n} \Rightarrow Z_b$ on $\Upsilon \times \Xi$, where the Z_b's have the same distribution as the process Z in Theorem 1 and are independent. Hence by the continuous mapping theorem,*

$$\left(\widehat{T}_{1,n},...,\widehat{T}_{M,n}\right)' \xrightarrow{d} (T_1,...,T_M)' \tag{24}$$

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

[9] Note that then by Assumption 1 the bootstrap $\widetilde{Y}_{b,j}$'s are bounded too.

Note that (24) carries over if the components of these random vectors are sorted in decreasing order. If so it follows that for $\alpha \in (0,1)$,

$$\widehat{T}_{[\alpha M],n} \xrightarrow{d} T_{[\alpha M]},$$

where $[\alpha M]$ is the largest integer $\leq \alpha M$. The statistic $\widehat{T}_{[\alpha M],n}$ is the $\alpha \times 100\%$ bootstrap critical value, and $T_{[\alpha M]}$ is approximately the actual asymptotic $\alpha \times 100\%$ critical value of T.

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

4 The Simulated ICM Test

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 if Y is multi-dimensional.

To cope with this problem, a Simulated Integrated Conditional Moment (SICM) test is proposed, in which the process $Z_n(\tau,\xi)$ in the exact ICM test statistic is replaced by either

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

if Y_j and X_j are bounded random vectors, or

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

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 $\Phi_1(.)$ and $\Phi_2(.)$ are bounded one-to-one mappings. The SICM test statistic is then

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

Theorem 3. *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* (3) *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, $\widehat{T}_n^{(s)} \xrightarrow{d} T_s = \int_\Upsilon\int_\Xi |Z(\tau,\xi) - Z_s(\tau,\xi)|^2 d\mu(\tau,\xi)$, *where* Z *is the same as in Theorem 1 and* Z_s *is a complex-valued zero mean Gaussian process with covariance function*

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

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

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

which is the same as (10).

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 (9). 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 feasible to generate random drawings $\widetilde{Y}$ from it. Another advantage is that $\widehat{T}_n^{(s)}$ has a closed form. In particular, with $Y_{\ell,j}$ and $X_{\ell,j}$ components ℓ of Y_j and X_j, respectively, we have

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

$$
\begin{aligned}
&- \prod_{\ell=1}^{m} \frac{\sin\left(c(Y_{\ell,j_1} - \widetilde{Y}_{\ell,j_2})\right)}{c(Y_{\ell,j_1} - \widetilde{Y}_{\ell,j_2})} - \prod_{\ell=1}^{m} \frac{\sin\left(c(\widetilde{Y}_{\ell,j_1} - Y_{\ell,j_2})\right)}{c(\widetilde{Y}_{\ell,j_1} - Y_{\ell,j_2})}\Bigg) \\
&\qquad \times \left(\prod_{\ell=1}^{k} \frac{\sin\left(c(X_{\ell,j_1} - X_{\ell,j_2})\right)}{c(X_{\ell,j_1} - X_{\ell,j_2})}\right) \\
&+ \frac{2}{n}\sum_{j=1}^{n}\left(1 - \prod_{\ell=1}^{m} \frac{\sin\left(c(Y_{\ell,j} - \widetilde{Y}_{\ell,j})\right)}{c(Y_{\ell,j} - \widetilde{Y}_{\ell,j})}\right) \qquad (25)
\end{aligned}
$$

as is not hard to verify.

Note that

$$\lim_{c\downarrow 0} \widehat{T}_n^{(s)}(c) = \lim_{c\to\infty} \widehat{T}_n^{(s)}(c) = 0$$

so that choosing too small or too large a c will destroy the small sample power of the test.

5 Monte Carlo Simulations

This section presents size and power performances of the SICM test $\widehat{T}_n^{(s)}(c)$ [c.f. (25)] in small samples, for $c = 4, 5, 6$, and various data generating processes, where the critical values are obtained by the bootstrap method. We adopt the data generating processes used in Zheng (2000). Each data generating process (DGP), indexed by DGP(k), corresponds to a conditional distribution $F_k(y|X)$, where X is standard normally distributed, and $y \in \mathbb{R}$. Only DGP(1) is correctly parametrized, as $F(y|X;\theta_0) = F_1(y|X)$ for some θ_0. This specification is the null hypothesis for all DGP's.

The bootstrap critical values have been computed according to the procedure described in section 3.5, with bootstrap sample size $M = 500$. The number of replications in each case is 1000. The percentage rejection frequencies are reported in Tables 1 and 2. For model 1 this is the size, and the power for the other models.[10]

The bootstrap sample size is 500, and the number of replications is 1000. The variables Y and X are transformed to bounded random variables using

[10] The program code is written in Fortran. In the case $n = 200$ it takes about half a minute for each simulation including 500 bootstrap repetitions, and about a minute in the case $n = 300$.

the arctan(.) function, after standardizing them by subtracting the sample mean first and then dividing the demeaned variable by the sample standard error. Thus, X has been transformed to

$$\Phi_2(X) = \arctan\left(\left(X - \overline{X}\right)/\widehat{S}_X\right)$$

where $\overline{X}$ is the sample mean of the X_j's and $\widehat{S}_X$ is the sample standard error, and the variable Y has been transformed similarly to

$$\Phi_1(Y) = \arctan\left(\left(Y - \overline{Y}\right)/\widehat{S}_Y\right).$$

The simulation settings are exactly the same as in Zheng (2000), except that the constant c in Zheng (2000) is the constant of the bandwidth, whereas in out case c is the constant in (25). The parameter c is set at either $c = 4$, 5, or 6.

Table 1: Percentage of rejections in non-truncated models

	$c=4$			$c=5$			$c=6$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
model 1a: $Y = 1 + X + e,\quad U \sim N(0,1)$									
200	0.8	4.9	10	1.1	5.0	9.2	1.1	4.3	9.8
300	1.6	5.7	11.8	1.6	5.6	11.7	1.6	5.6	10.8
model 2a: $Y = 1 + X + U,\quad U \sim$ Logistic									
200	2.6	8.1	16.0	3.0	9.1	15.7	3.1	8.5	15.5
300	4.4	10.9	18.7	3.6	11.0	17.9	3.3	10.3	16.8
model 3a: $Y = 1 + X + U,\quad U \sim t_5$									
200	99.8	99.9	100	99.9	99.9	100	99.9	99.9	100
300	100	100	100	100	100	100	99.9	100	100
model 4a: $Y = 1 + X + X^2 + U,\quad U \sim N(0,1)$									
200	100	100	100	100	100	100	100	100	100
300	100	100	100	100	100	100	100	100	100
model 5a: $Y = 1 + X + X.U,\quad U \sim N(0,1)$									
200	99.9	99.9	99.9	99.9	100	100	100	100	100
300	100	100	100	100	100	100	100	100	100

Table 2: Percentage of rejections in truncated models

	$c=4$			$c=5$			$c=6$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
model 1b: $Y = \max(0, 1 + X + U)$, $U \sim N(0,1)$									
200	1.4	4.4	9.8	1.1	4.3	9.3	1.0	4.9	9.9
300	1.7	5.5	12.0	1.6	6.8	11.0	1.8	6.9	11.5
model 2b: $Y = \max(0, 1 + X + U)$, $U \sim$ Logistic									
200	2.9	7.7	14.0	2.3	7.9	14.0	2.5	8.0	13.7
300	2.6	8.5	15.0	2.4	8.3	14.4	2.0	8.0	13.9
model 3b: $Y = \max(0, 1 + X + U)$, $U \sim t_5$									
200	5.0	13.0	18.9	4.6	13.1	21.2	6.2	15.1	23.1
300	6.4	15.9	24.0	6.2	15.1	23.1	5.4	12.7	20.8
model 4b: $Y = \max(0, 1 + X + X^2 + U)$, $U \sim N(0,1)$									
200	100	100	100	100	100	100	100	100	100
300	100	100	100	100	100	100	100	100	100
model 5b: $Y = \max(0, 1 + X + X.U)$, $U \sim N(0,1)$									
200	99.8	99.9	100	100	100	100	100	100	100
300	100	100	100	100	100	100	100	100	100

Compared with Zheng's (2000) simulation results, for both cases the actual size of the SICM test is much closer to the nominal size. The highest 10% size value in Zheng (2000) is only 5.5 for group model 1a, and 4.9 for model 1b. For some value of c, the 10% size is just 2.8 for $n = 300$. However, Zheng used asymptotic critical values, whereas we use bootstrap critical values. For model 2, the power is in general higher than in Zheng's case, especially for the truncated model, for which Zheng's test has power of 7.1 at the 10% level. In the case of model 3, the power of the ICM test is comparable with the power of Zheng's test for $n = 200$, and slightly better for $n = 300$. The SICM test has power 1 for both model 4 and 5, for all n's and c's, whereas in Zheng's case some powers are below 90%. Thus, in general the SICM test performs better than Zheng's test.

6 Application to Count Data Models in Health Economics

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 ICM 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 exclusive a property of the Poisson distribution, and therefore such a test is not consistent.

Table 3: Poisson model variables

variable name	meaning	mean	std
ofp	# of visit to physicians in an office setting	5.8	6.8
exclhlth	self-perceived health condition: excellent	0.1	0.3
poorhlth	self-perceived health condition: poor	0.1	0.3
numchron	# of chronicle diseases and conditions	1.5	1.3
adldiff	a measure of disability status	0.2	0.4
noreast	region: northeast	0.2	0.4
midwest	region: midwest	0.3	0.4
west	region: west	0.2	0.4
age	age in years (divided by 10)	7.4	0.6
black	= 1 if black	0.1	0.3
male	= 1 if male	0.4	0.5
married	= 1 if married	0.6	0.5
school	years of schooling	10.3	3.7
faminc	family income (in 10,000)	2.5	2.9
employed	employment status	0.1	0.3
privins	private insurance status	0.8	0.4
medicaid	public insurance status	0.1	0.3

The data source is the 1987-1988 National Medical Expenditure Survey (NMES) used by Deb and Trivedi (1997). There are 4406 observations of individuals over age 66. The variable Y of interest is the number of physician visits by elderly, which is explained by a vector X of various variables of health conditions and demographic characteristics, as listed in Table 3. Note that variable 1 in Table 3 is the dependent variable Y.

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

We will use the SICM test to test the Poisson hypothesis, where Y and the X variables have been standardized and transformed in the same way as in the simulation study. To generate random drawings from the Poisson distribution we have use the fact that for independent random drawings e_j from the standard exponential distribution, the smallest Y for which $\sum_{j=1}^{Y} e_j > \lambda$ has a the Poisson(λ) distribution. See the appendix.

The SICM test has been conducted in the form (25), for $c = 1, 2, ..., 6$. Instead of bootstrap critical values we have computed bootstrap p-values, using 500 bootstrap samples. For all these values of c the bootstrap p-values are virtually zero, so that the null hypothesis is strongly rejected. This result supports the need for ongoing efforts in health economic research to develop more general models for count data than the Poisson model.

7 Conclusions

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_1^2 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 the same asymptotic properties as the latter test.

The small sample power and size of the test is better than competing tests, as shown via Monte Carlo simulations. 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. In further research, we will apply the SICM test to check whether more general count data models are appropriate for this data.

The extension of the ICM test to time series models, where one has to condition on infinitely many lagged variables, will be done in a follow-up paper, Bierens and Wang (2008).

8 Appendix

8.1 Proof of Lemma 2

For notational convenience, assume that Y and X are scalar variables. Denote

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

Using the well-known series expansion of the complex exp(.) function we can write

$$\begin{aligned} D(\tau,\xi) &= E\left[\left(\sum_{r=0}^{\infty}\frac{(i.\tau Y)^r}{r!} - \int\left(\sum_{r=0}^{\infty}\frac{(i\tau Y)^r}{r!}\right)dF(y|X,\theta_0)\right)\right. \\ &\qquad \left.\times\left(\sum_{s=0}^{\infty}\frac{(i\xi X)^s}{s!}\right)\right] \\ &= \sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{i^{r+s}\tau^r\xi^s}{r!s!}E\left[\left(Y^r - \int y^r dF(y|X,\theta_0)\right)X^s\right] \end{aligned}$$

where the second equality is due to the boundedness of Y and X. Since under H_1, $D(\tau,\xi) \neq 0$ for some (τ,ξ), it follows that $E\left[\left(Y^r - \int y^r dF(y|X,\theta_0)\right)X^s\right]$

$\neq 0$ for some $r \geq 1$ and $s \geq 0$, hence

$$\left.\frac{\partial^{r+s} D(\tau,\xi)}{(\partial\tau)^r(\partial\xi)^s}\right|_{\tau=\xi=0} = i^{r+s}\left[\left(Y^r - \int y^r dF(y|X,\theta_0)\right)X^s\right] \neq 0$$

This implies that in an arbitrary open neighborhood of $(0,0)$ there exists a (τ_*, ξ_*) such that $D(\tau_*, \xi_*) \neq 0$. By the continuity of $D(\tau,\xi)$ it then follows that $D(\tau,\xi) \neq 0$ in an open neighborhood of (τ_*, ξ_*).

The extension to the vector case is straightforward.

8.2 Proof of Lemma 3

By the mean value theorem and Assumption 1

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

and

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

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

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

Note that

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

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

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\left\|\widehat{R}_n(\tau|X_j)\right\| \\
&\quad\leq I\left(\left\|\widehat{\theta}-\theta_0\right\|\leq\delta\right)\left(\frac{1}{n}\sum_{j=1}^{n}\sup_{\|\theta-\theta_0\|\leq\delta}\|\Delta\varphi(\tau|X_j;\theta)-\Delta\varphi(\tau|X_j;\theta_0)\|\right) \\
&\quad+2.I\left(\left\|\widehat{\theta}-\theta_0\right\|>\delta\right)\frac{1}{n}\sum_{j=1}^{n}\sup_{\theta\in\Theta_0}\|\Delta\varphi(\tau|X_j;\theta_0)\| \\
&\quad\leq\frac{1}{n}\sum_{j=1}^{n}\sup_{\|\theta-\theta_0\|\leq\delta}\|\Delta\varphi(\tau|X_j;\theta)-\Delta\varphi(\tau|X_j;\theta_0)\|+o_p(1).
\end{aligned}$$

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,

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

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

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

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

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \varphi(\tau|X_j;\theta_0) \exp(i.\xi' X_j)$$

$$+ b(\tau,\xi)' A^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Delta\ell(Y_j|X_j;\theta_0) + o_p(1)$$

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

8.3 Proof of Lemma 4

Let $U_n(\tau,\xi) = Z_n(\tau,\xi) - \widetilde{Z}_n(\tau,\xi)$ and observe from Lemma 3 that $p\lim_{n\to\infty}$ $U_n(\tau,\xi) = 0$, pointwise in (τ,ξ), hence by bounded convergence

$$\int_\Upsilon \int_\Xi |U_n(\tau,\xi)|^2 \, d\mu(\tau,\xi) = o_p(1) \tag{26}$$

Next, observe that

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

hence

$$\int_\Upsilon \int_\Xi |\widetilde{Z}_n(\tau,\xi)|^2 d\mu(\tau,\xi) = O_p(1). \tag{27}$$

Finally, it is easy to verify that

$$\left| |Z_n(\tau,\xi)|^2 - \left|\widetilde{Z}_n(\tau,\xi)\right|^2 \right| \le |U_n(\tau,\xi)|^2 + 2\,|U_n(\tau,\xi)| \,.\, \left|\widetilde{Z}_n(\tau,\xi)\right|$$

and therefore

$$\begin{aligned}
\left|\widehat{T}_n - \widetilde{T}_n\right| &\le \int_\Upsilon \int_\Xi |U_n(\tau,\xi)|^2 \, d\mu(\tau,\xi) + 2\int_\Upsilon \int_\Xi |U_n(\tau,\xi)| \,.|\widetilde{Z}_n(\tau,\xi)| d\mu(\tau,\xi) \\
&\le \int_\Upsilon \int_\Xi |U_n(\tau,\xi)|^2 \, d\mu(\tau,\xi) \\
&\quad +2\sqrt{\int_\Upsilon \int_\Xi |\widetilde{Z}_n(\tau,\xi)|^2 d\mu(\tau,\xi)}\sqrt{\int_\Upsilon \int_\Xi |U_n(\tau,\xi)|^2 \, d\mu(\tau,\xi)} \\
&= o_p(1)
\end{aligned}$$

where the latter follows from (26) and (27).

8.4 Proof of Lemma 5

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

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

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

$$\sqrt{n}(\sigma_n-\sigma)=O_p(1),\ \sqrt{n}(\mu_n-\mu)=O_p(1) \tag{28}$$

Note that

$$\Psi'(x)=\frac{1}{1+x^2},\ \Psi''(x)=\frac{-2x}{(1+x^2)^2} \tag{29}$$

hence

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

say. Note that due to (29), $\Delta_1(Y_j|\sigma,\mu)$ and $\Delta_2(Y_j|\sigma,\mu)$ are bounded in Y_j.

Now by Taylor's theorem,

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

$$\times \begin{pmatrix} \sigma_n - \sigma \\ \mu_n - \mu \end{pmatrix}$$
$$= \sqrt{n}\left(\sigma_n - \sigma, \mu_n - \mu\right) \Delta_1\left(Y_j|\sigma,\mu\right) + O_p\left(1/\sqrt{n}\right)$$

and similarly,

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

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

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

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

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

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

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

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

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

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

8.5 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 $\beta_1,\beta_2,...,\beta_M$ in $\mathbf{B}$,

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

In the case of the empirical process (6) 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

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

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 4 is suffices to prove $\widetilde{Z}_n(\tau,\xi) \Rightarrow Z(\tau,\xi)$, where $\widetilde{Z}_n(\tau,\xi)$ defined by (6). Thus, in our case $\beta = (\tau,\xi)$, $\mathbf{B} = \Upsilon \times \Xi$, and $Z_n(\beta) = \widetilde{Z}_n(\tau,\xi)$.

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

$$\begin{aligned}\widetilde{Z}_{1,n}(\tau,\xi) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp(i.\tau' Y_j) - \int \exp(i.\tau' y)dF(y|X_j;\theta_0)\right)\exp(i.\xi' X_j) \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\exp(i.\tau' Y_j) - E[\exp(i.\tau' Y_j)|X_j]\right)\exp(i.\xi' X_j) \\ \widetilde{Z}_{2,n}(\tau,\xi) &= b(\tau,\xi)' A^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Delta\ell\left(Y|X;\theta_0\right)\end{aligned}$$

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

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

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

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

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

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

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

Then

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

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

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

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

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

Hence,

$$\begin{aligned} & \left(\sup_{||(\tau,\xi)' - (\rho,\zeta)'|| < \varepsilon} \left| \widetilde{Z}_{1,n}(\tau, \xi) - \widetilde{Z}_{1,n}(\rho, \zeta) \right| \right)^2 \\ & \quad \leq 16\varepsilon^2 \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{r!s!} \left(2c \right)^{r+s} \sum_{b=0}^{r} \binom{r}{b} \sum_{q=0}^{s} \binom{s}{q} \right. \end{aligned}$$

$$\times \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(Y_{1,j}^{b} Y_{2,j}^{r-b} - E[Y_{1,j}^{b} Y_{2,j}^{r-b} | X_j] \right) X_{1,j}^{q} X_{2,j}^{s-q} \right| \Bigg)^2$$

$$\leq 16\varepsilon^3 \left(\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{r!s!} \sum_{b=0}^{r} \binom{r}{b} \sum_{q=0}^{s} \binom{s}{q} (2c)^{r+s} \right)$$

$$\times \Bigg\{ \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{1}{r!s!} (2c)^{r+s} \sum_{b=0}^{r} \binom{r}{b} \sum_{q=0}^{s} \binom{s}{q}$$

$$\times \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(Y_{1,j}^{b} Y_{2,j}^{r-b} - E[Y_{1,j}^{b} Y_{2,j}^{r-b} | X_j] \right) X_{1,j}^{q} X_{2,j}^{s-q} \right)^2 \Bigg\}$$

and thus

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

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

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

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

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

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.

8.6 Proof of Lemma 7

Since $Z(\beta)$, $\beta = (\tau, \xi)$, is continuous on $\mathbf{B} = \Upsilon \times \Xi$ and $\mathbf{B}$ is compact, it follows from Lemma 6 that $Z(\beta)$ can be written as $Z(\beta) = \sum_{j=1}^{\infty} g_j \psi_j(\beta)$, where $g_j = \int Z(\beta) \overline{\psi}_j(\beta) d\mu(\beta)$, hence

$$T = \int |Z(\beta)|^2 \, d\mu(\beta) = \int Z(\beta) \overline{Z(\beta)} d\mu(\beta) = \sum_{j=1}^{\infty} |g_j|^2$$

The Fourier coefficients g_j are complex-valued jointly normally distributed, with zero expectation and covariances

$$
\begin{aligned}
E\left[g_{j_1}\overline{g}_{j_2}\right] &= E\left[\left(\int Z(\beta_1)\overline{\psi}_{j_1}(\beta_1)d\mu(\beta_1)\right)\left(\int \overline{Z}(\beta_2)\psi_{j_2}(\beta_2)d\mu(\beta_2)\right)\right] \\
&= \int\int E\left[Z(\beta_1)\overline{Z}(\beta_2)\right]\overline{\psi}_{j_1}(\beta_1)\psi_{j_2}(\beta_2)d\mu(\beta_1)d\mu(\beta_2) \\
&= \int\int \Gamma(\beta_1,\beta_2)\overline{\psi}_{j_1}(\beta_1)\psi_{j_2}(\beta_2)d\mu(\beta_1)d\mu(\beta_2) \\
&= \sum_{j=1}^{\infty}\lambda_j\int \psi_j(\beta_1)\overline{\psi}_{j_1}(\beta_1)d\mu(\beta_1)\int \psi_{j_2}(\beta_2)\overline{\psi_j}(\beta_2))d\mu(\beta_2) \\
&= \lambda_j \text{ if } j = j_1 = j_2 \\
&= 0 \text{ if } j_1 \neq j_2
\end{aligned}
$$

hence the g_j are independent normally distributed, with variance $E[g_j\overline{g}_j] = \lambda_j$ hence if $\lambda_j > 0$ then $|g_j|^2/\lambda_j$ has a χ_1^2 distribution.

8.7 Proof of Lemma 8

In view of (7) it suffices to show that

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

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

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

Then

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

hence for $0 < \delta < 1$,

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

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

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

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

8.8 Proof of Theorem 3

It follows similar to Lemma 4 that

$$\widehat{T}_n^{(s)} = \int_\Upsilon \int_\Xi |\widetilde{Z}_n(\tau, \xi) - \widetilde{Z}_n^{(s)}(\tau, \xi)|^2 d\mu(\tau, \xi) + o_p(1). \tag{30}$$

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, $\widetilde{Z}_n^{(s)} \Rightarrow Z_s$ conditional on $\mathcal{D}$, where $Z_s(\tau, \xi)$ is a zero-mean Gaussian process with covariance function

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

where $\varphi(\tau|X; \theta)$ is the conditional characteristic function of $F(y|X_1, \theta)$ [c.f. (4)]. Since Γ_s does not depend on $\mathcal{D}$ is follows now that unconditionally,

$$\widetilde{Z}_n^{(s)} \Rightarrow Z_s \text{ under } H_0 \text{ and } H_1, \tag{31}$$

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 (30), (31) and the continuous mapping theorem,

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

The result under H_1 is easy to verify from (31) and Theorem 1.

8.9 Generating Poisson Random Variables

Let X_j be a random drawing from the exponential distribution $F(x) = 1 - \exp(-x)$. Then

$$\Pr[X_1 > \lambda] = \exp(-\lambda)$$

and

$$\begin{aligned}
\Pr[\{X_1 + X_2 > \lambda\} \cap \{X_1 \leq \lambda\}] &= E[I(X_2 > \lambda - X_1) I(X_1 \leq \lambda)] \\
&= E[E(I(X_2 > \lambda - X_1) | X_1) I(X_1 \leq \lambda)] \\
&= \int_0^\lambda \exp(-\lambda + x) \exp(-x) dx \\
&= \exp(-\lambda) \int_0^\lambda dx = \lambda \exp(-\lambda)
\end{aligned}$$

Denote $Z_k = \sum_{j=1}^k X_j$ for $k = 1, 2, 3, ...,$ $Z_0 = 0$. We have shown that for $m = 1, 2,$

$$\Pr[\{Z_m > \lambda\} \cap \{Z_{m-1} \leq \lambda\}] = \frac{\lambda^{m-1}}{(m-1)!} \exp(-\lambda) \tag{32}$$

Suppose that (32) holds for $m = 1, 2, ..., k.$. Then for $1 \leq m \leq k$,

$$\begin{aligned}
\Pr[Z_m > \lambda] &= \Pr[\{Z_m > \lambda\} \cap \{Z_{m-1} \leq \lambda\}] \\
&\quad + \Pr[\{Z_m > \lambda\} \cap \{Z_{m-1} > \lambda\}] \\
&= \frac{\lambda^{m-1}}{(m-1)!} \exp(-\lambda) + \Pr[\{Z_{m-1} > \lambda\}]
\end{aligned}$$

hence

$$\begin{aligned}
\Pr[Z_m > \lambda] - \Pr[Z_{m-1} > \lambda] &= \frac{\lambda^{m-1}}{(m-1)!} \exp(-\lambda) \\
m &= 1, 2, .., k
\end{aligned}$$

Summing up for $m = 1, 2, .., k$ yields

$$\Pr\left[Z_k > z\right] = \sum_{m=0}^{k-1} \frac{z^m}{m!} \exp\left(-z\right),$$

with density

$$f_k(z) = \sum_{m=0}^{k-1} \frac{z^m}{m!} \exp\left(-z\right) - \sum_{m=1}^{k-1} \frac{z^{m-1}}{(m-1)!} \exp\left(-z\right) = \frac{z^{k-1}}{(k-1)!} \exp\left(-z\right)$$

Now

$$\begin{aligned} \Pr\left[\{Z_{k+1} > \lambda\} \cap \{Z_k \leq \lambda\}\right] &= E\left[\exp\left(-\lambda + Z_k\right) I\left(Z_k \leq \lambda\right)\right] \\ &= \int_0^\lambda \exp\left(-\lambda + z\right) \frac{z^{k-1}}{(k-1)!} \exp\left(-z\right) dz \\ &= \exp\left(-\lambda\right) \int_0^\lambda \frac{z^{k-1}}{(k-1)!} dz = \frac{\lambda^k}{k!} \exp\left(-\lambda\right) \end{aligned}$$

Therefore by induction, (32) holds for all $k \geq 0$.

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., and Ploberger, W., 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.

Bickle, P. and Freedman, D., 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, U.K.: Cambridge University Press

Bierens, H. J. and Carvalho, J.R., 2007, Semi-Nonparametric Competing Risks Analysis of Recidivism, *Journal of Applied Econometrics*, 22, 971-993.

Bierens, H.J. and Ploberger, W., 1997, Asymptotic Theory of Integrated Conditional Moment Tests, *Econometrica*, 65, 1129-1151.

Bierens, H. J., and L. Wang, 2008, Consistent Integrated Conditional Moment Tests for Parametric Conditional Distributions of Stationary Time Series Processes (incomplete working paper, downloadable from http://econ.la.psu.edu/~hbierens/ICM_TS.PDF)

Billingsley, P., 1968, *Convergence of Probability Measures.* New York: John Wiley.

Boning, B. and Sowell, F., 1999, Optimality for the Integrated Conditional Moment Test, *Econometric Theory*, 15, 710-719.

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

Chernick, M., 1999, *Bootstrap Methods: A Practitioner's Guide*, New York: John Wiley.

Chung, K.L., 1976, *A Course in Probability Theory*, New York: Academic Press

Deb, P. and Trivedi, P., 1997, Demand for Medical Care by the Elderly, *Journal of Applied Econometrics*, 12, 313-336.

Efron, B., 1979, Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, 7, 1-26.

Fan, Y.Q. and Li, Q., 1996, Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms, *Econometrica*, 65, 865-890.

Fan, Y.Q., and Li, Q., 2000, Consistent Model Specification Tests: Kernel Based Tests Versus Bierens' ICM tests, *Econometric Theory*, 16, 1016-1041.

Hardle, W. and Mammen, E, 1993, Comparing Nonparametric Versus Parametric Regression Fits, *Annals of Statistics*, 21, 1926-1947.

Hadinejad-Mahram, H, Dahlhaus, D and Blomker, D, 2002, Karhunen-Loeve Expansion of Vector Random Processes, Technical Report No. IKT-

NT 1019, Communications Technological Laboratory, Swiss Federal Institute of Technology, Zurich.

Hausman, J., 1978, Specification Tests in Econometrics, *Econometrica*, 46, 1251-1271.

Holly, A., 1982, A Remark on Hausman's Specification Test, *Econometrica*, 50, 749-760.

Hong, Y. and White, H., 1995, Consistent Specification Testing Via Nonparametric Series Regression, *Econometrica*, 63, 1133-1159.

Jennrich, R. I. ,1969, Asymptotic Properties of Non-Linear Least Squares Estimators, *Annals of Mathematical Statistics*, 40, 633-643.

Justel, A., Pena, D. and Zamar, R., 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.

Lee, L.F., 1986, Specification Test for Poisson Regression Models, *International Economic Review*, 27, 689-706.

Li, F., and Tkacz, G., 1996, A Consistent Bootstrap Test for Conditional Density Functions with Time-Series Data, *Journal of Econometrics*, 133, 863-886.

Kullback, L., and Leibler, R. A., 1951, On Information and Sufficiency, *Annals of Mathematical Statistics,* 22, 79-86.

Lukacs, ,E., 1970, *Characteristic Functions*, Hafner Publishing, New York.

Newey, W.K., 1985, Maximum Likelihood Specification Testing and Conditional Moment Tests, *Econometrica*, 53, 1047-1070.

Pinkse, J., 1998, A Consistent Nonparametric Test for Serial Independence, *Journal of Econometrics*, 84, 205-232.

Stinchcombe,M. and White, H., 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.

Su, L., and White, H., 2006, :Testing Conditional Independence via Empirical Likelihood, Working paper, Department of Economics, UCSD.

Vuong, Q.H., 1989, Likelihood Ratio Tests for Model Selection and Non-Nested Hypothesis, *Econometrica*, 57, 307-333.

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*, Cambidge 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.