Integrated Conditional Moment Tests for Parametric Conditional Distributions

Herman J. Bierens and Li Wang

This paper extends the Integrated Conditional Moment (ICM) test for the functional form of nonlinear regression models to tests for the correctness of parametric conditional distribution specifications.

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.

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.

1 Literature

1.1 ICM tests

1.1.1 Cross-section regression models

Bierens, H.J., 1982, Consistent Model Specification Tests, *Journal of Econometrics*, 20, 105-134.

Bierens, H.J.,1990, A Consistent Conditional Moment Test of Functional Form, *Econometrica*, 58, 1443-1458.

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

Stinchcombe, M. and White, H., 1998, Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative, *Econometric Theory*, 14, 295-325.

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

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

Escanciano, J.C., 2006, A Consistent Diagnostic Test for Regression Models Using Projections, *Econometric Theory*, 22, 1030-1051.

Escanciano, J.C., 2009, On the Lack of Power of Omnibus Specification Tests, *Econometric Theory*, 25, 162-194.

1.1.2 Time series regression models

Bierens, H.J., 1984, Model Specification Testing of Time Series Regressions, *Journal of Econometrics*, 26, 323-353.

De Jong, R., 1996, The Bierens Test Under Data Dependence, *Journal of Econometrics*, 72, 1-32.

1.2 Consistent tests for conditional distribution models

Andrews, D.W., 1997, A Conditional Kolmogorov Test, *Econometrica*, 65, 1097-1128.

Test statistic:

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

Problem:

If the dimension of X_j is large, the inequality $X_j < X_i$ may never happen, even in large samples.

Zheng, J.X., 2000, A Consistent Test of Conditional Parametric Distributions, *Econometric Theory*, 16, 667-691.

Test statistic:

$$T_n \propto \frac{1}{n} \sum_{j=1}^{n} \frac{\widehat{p}_2(Y_j, X_j) - f(Y_i|X_j, \widehat{\theta})\widehat{p}_1(X_j)}{f(Y_i|X_j, \widehat{\theta})}$$

where $\widehat{p}_2(y, x)$ and $\widehat{p}_1(x)$ are kernel density estimators.

Problems:

This test only applies if (Y_j, X_j) has an absolutely continuous distribution.

The local power rate is slower than $\sqrt{n}$, and decreases with the dimension of X_j.

2 The ICM test for conditional distribution models

Throughout we will assume that

(a) We observe a random sample $(Y_1, X_1), ..., (Y_n, X_n)$ from $(Y, X) \in \mathbb{R}^m \times \mathbb{R}^k$.

(b) The conditional distribution function of Y given X is assumed to belongs to a given parametric family

$$F(y|X;\theta),\ \theta \in \Theta,$$

where $\Theta \subset \mathbb{R}^p$ is a compact and convex parameter space.

(c) The conditions for convergence in probability and asymptotic normality of the quasi-maximum likelihood (QML) estimator $\widehat{\theta}$ hold.

Hypothesis to be tested:

$H_0 : \exists\theta_0 \in \Theta, \sup_{y\in\mathbb{R}^m} |\Pr[Y \leq y|X] - F(y|X;\theta_0)| = 0$ a.s.

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

These hypotheses are equivalent to:

$$H_0 : \exists\theta_0 \in \Theta : \sup_{\tau\in\mathbb{R}^m} \left| E\left[\exp(i.\tau' Y)|X\right] - \int \exp(i\tau' y) dF(y|X,\theta_0)\right| = 0 \text{ a.s.}$$

$$H_1 : \forall\theta \in \Theta : \sup_{\tau\in\mathbb{R}^m} \left| E\left[\exp(i.\tau' Y)|X\right] - \int \exp(i\tau' y) dF(y|X,\theta)\right| > 0 \text{ a.s.}$$

These hypotheses are equivalent to:

$$H_0 \Leftrightarrow$$
$$\exists \theta_0 \in \Theta : \sup_{(\tau,\xi)\in\mathbb{R}^m\times\mathbb{R}^k} \left| E\left[\exp(i.\tau'Y)\exp(i.\xi'X)\right] - E\left[\int \exp(i\tau'y)dF(y|X,\theta_0)\exp(i.\xi'X)\right]\right| = 0$$

$$H_1 \Leftrightarrow$$
$$\forall \theta \in \Theta : \sup_{(\tau,\xi)\in\mathbb{R}^m\times\mathbb{R}^k} \left| E\left[\exp(i.\tau'Y)\exp(i.\xi'X)\right] - E\left[\int \exp(i\tau'y)dF(y|X,\theta)\exp(i.\xi'X)\right]\right| > 0$$

Suppose that $Y \in \mathbb{R}^m$ and $X \in \mathbb{R}^k$ are *bounded* random vectors.

Then for arbitrary $\varepsilon > 0$,

$$\begin{aligned} & H_1 \Leftrightarrow \\ & \forall \theta \in \Theta : \sup_{||\tau|| \leq \varepsilon, ||\xi|| \leq \varepsilon} \Bigg| E\left[\exp(i.\tau' Y)\exp(i.\xi' X)\right] \\ & \quad - E\left[\int \exp(i\tau' y) dF(y|X,\theta)\exp(i.\xi' X)\right]\Bigg| > 0 \end{aligned}$$

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 for arbitrary $\varepsilon > 0$,

$$
\begin{aligned}
&H_1 \Leftrightarrow \\
&\forall \theta \in \Theta : \sup_{||\tau|| \leq \varepsilon, ||\xi|| \leq \varepsilon} \Bigg| E\left[\exp(i.\tau'\Phi_1(Y)\exp(i.\xi'\Phi_2(X))\right] \\
&\quad - E\left[\int \exp(i\tau'\Phi_1(y)dF(y|X,\theta)\exp(i.\xi'\Phi_2(X))\right]\Bigg| > 0
\end{aligned}
$$

For the time being we will assume that Y and X are bounded random vectors.

Denote

$$\begin{aligned}
\varsigma(\tau,\xi;\theta) &= E\left[\left(\exp(i\tau'Y) - \int \exp(i\tau'y)\, dF(y|X,\theta)\right)\right. \\
&\quad \times \exp(i\xi'X) \\
\Upsilon &= \times_{j=1}^{m}\left[-\overline{\tau}_j, \overline{\tau}_j\right],\ \overline{\tau}_j > 0, \\
\Xi &= \times_{j=1}^{k}\left[-\overline{\xi}_j, \overline{\xi}_j\right],\ \overline{\xi}_j > 0 \\
d\mu(\tau,\xi) &= \left(\prod_{j=1}^{m}(2\overline{\tau}_j)^{-1}\right)\left(\prod_{j=1}^{k}(2\overline{\xi}_j)^{-1}\right) d\tau d\xi
\end{aligned}$$

Then

$$H_0 \Leftrightarrow \exists\theta_0 \in \Theta : \int_{\Upsilon\times\Xi} \left|\varsigma(\tau,\xi;\theta_0)\right|^2 d\mu(\tau,\xi) = 0$$

$$H_1 \Leftrightarrow \forall\theta \in \Theta : \int_{\Upsilon\times\Xi} \left|\varsigma(\tau,\xi;\theta)\right|^2 d\mu(\tau,\xi) > 0$$

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

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

where

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

is a complex-valued continuous empirical process on $\Upsilon \times \Xi$, with $\hat{\theta}$ the QML estimator of θ_0.

3 Asymptotic properties

3.1 Main result

Theorem 1. *Let* Y *and* X *be bounded random vectors. Under* H_0,

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

where Z *is a zero mean complex-valued Gaussian process, hence by the continuous mapping theorem,*

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

Under H_1,

$$p\lim_{n\to\infty} \widehat{T}_n/n > 0,$$

hence

$$p\lim_{n\to\infty} \widehat{T}_n = \infty.$$

3.2 Standardization

The condition that Y and X are bounded random vectors is not essential, because in we may without loss of generality replace Y and X by bounded one-to-one mappings $\Phi_1(Y)$ and $\Phi_2(X)$, respectively.
However, it is important to standardize Y and X before taking any bounded transformation. In particular, in the case $Y \in \mathbb{R}$ let

$$\Phi_1(Y) = \arctan\left(\sigma_n^{-1}\left(Y - \mu_n\right)\right),$$

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.
As long as we choose μ_n and σ_n such that

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

this standardization does not effect the asymptotic properties.

3.3 The null distribution

The asymptotic null distribution

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

depend on the zero-mean complex valued Gaussian process $Z(\tau, \xi)$, which in its turn is determined by the covariance function

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

Consider the eigenvalue problem:
Find an eigenvalue λ and corresponding eigenfunction $\psi(\tau, \xi)$ such that

$$\lambda\psi(\tau_1, \xi_1) = \int_{\Upsilon \times \Xi} \Gamma\left((\tau_1, \xi_1), (\tau_2, \xi_2)\right) \overline{\psi(\tau_2, \xi_2)} d\mu(\tau_2, \xi_2).$$

This problem has countable many solutions $\lambda_j, \psi_j(\tau, \xi), j = 1, 2, 3, \ldots\ldots$

Properties:

- The eigenvalues λ_j are real-valued and non-negative.
- The eigenfunctions $\psi_j(\tau, \xi)$ are complex-valued and orthonormal:

$$\int_{\Upsilon \times \Xi} \psi_{j_1}(\tau, \xi) \overline{\psi_{j_2}(\tau, \xi)} d\mu(\tau, \xi) = I\left(j_1 = j_2\right)$$

- $Z(\tau, \xi) = \sum_{j=1}^{\infty} g_j \psi_j(\tau, \xi)$, where

$$g_j = \int_{\Upsilon \times \Xi} Z(\tau, \xi) \overline{\psi_j(\tau, \xi)} d\mu(\tau, \xi)$$

- Mercer's theorem:

$$\begin{aligned} \Gamma\left((\tau_1, \xi_1), (\tau_2, \xi_2)\right) &= E\left[Z(\tau_1, \xi_1)\overline{Z(\tau_2, \xi_2)}\right] \\ &= \sum_{j=1}^{\infty} \lambda_j \psi_j(\tau, \xi) \overline{\psi_j(\tau, \xi)} \end{aligned}$$

$Z(\tau,\xi) = \sum_{j=1}^{\infty} g_j \psi_j(\tau,\xi)$ implies

$$\begin{aligned}
T &= \int_{\Upsilon\times\Xi} |Z(\tau,\xi)|^2 d\mu(\tau,\xi) \\
&= \int_{\Upsilon\times\Xi} Z(\tau,\xi)\overline{Z(\tau,\xi)} d\mu(\tau,\xi) \\
&= \int_{\Upsilon\times\Xi} \left(\sum_{j_1=1}^{\infty} g_{j_1}\psi_{j_1}(\tau,\xi)\right) \left(\sum_{j_2=1}^{\infty} \overline{\psi_{j_2}(\tau,\xi)}\,\overline{g}_{j_2}\right) d\mu(\tau,\xi) \\
&= \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} g_{j_1} \left(\int_{\Upsilon\times\Xi} \psi_{j_1}(\tau,\xi)\overline{\psi_{j_2}(\tau,\xi)} d\mu(\tau,\xi)\right) \overline{g}_j \\
&= \sum_{j=1}^{\infty} g_j \overline{g}_j \\
&= \sum_{j=1}^{\infty} |g_j|^2, \text{ where } g_j = \int_{\Upsilon\times\Xi} Z(\tau,\xi)\overline{\psi_j(\tau,\xi)} d\mu(\tau,\xi)
\end{aligned}$$

Properties of $g_j = \int_{\Upsilon \times \Xi} Z(\tau, \xi)\overline{\psi_j(\tau, \xi)} d\mu(\tau, \xi)$

- The g_j's are complex-valued zero-mean Gaussian.
- $E[g_{j_1} \overline{g}_{j2}] = \lambda_j$ if $j = j_1 = j_2,\ E[g_{j_1} \overline{g}_{j2}] = 0$ if $j_1 \neq j_2$. The latter implies that the g_j's are independent.
- $(\mathrm{Re}(g_j), \mathrm{Im}(g_j))' \sim \sqrt{\lambda_j} e_j$ where $e_j \sim N_2 [0, I_2]$.

Therefore,

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

$$\sim \sum_{j=1}^{\infty} \lambda_j \chi^2_{2,j}$$

where the $\chi^2_{2,j}$'s are independently χ^2_2 distributed.

3.4 Parametric bootstrap

For $b = 1, ..., M$, generate random drawings $\widetilde{Y}_{b,j}$ from $F(y|X_j, \widehat{\theta})$, and compute the ICM test statistic $\widehat{T}_{b,n}$ for each bootstrap sample $(\widetilde{Y}_{b,1}, X_1), ..., (\widetilde{Y}_{b,n}, X_n)$. Then

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

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

Rather than computing bootstrap critical values, it is more convenient to compute bootstrap p-values

$$\widehat{p}_{n,M} = \frac{1}{M}\sum_{b=1}^{M} I\left(\widehat{T}_{b,n} > \widehat{T}_n\right)$$

and reject H_0 at the $\alpha \times 100\%$ significance level if $\widehat{p}_{n,M} < \alpha$.

3.5 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

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

Then

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

where $E\left[Z(\tau,\xi)\right] \neq 0$.

Denote

$$\begin{aligned} g_j &= \int_{\Upsilon\times\Xi} \left(Z(\tau,\xi) - E\left[Z(\tau,\xi)\right]\right) \overline{\psi_j(\tau,\xi)} d\mu(\tau,\xi) \\ \eta_j &= \int_{\Upsilon\times\Xi} E\left[Z(\tau,\xi)\right] \overline{\psi_j(\tau,\xi)} d\mu(\tau,\xi) \end{aligned}$$

Then

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

where at least one of the η_j's is nonzero.

This implies that for all $K > 0$,

$$\Pr(T_{alt} > K) > \Pr(T > K).$$

Thus, the ICM test has non-trivial power against $\sqrt{n}$-local alternatives.

3.6 Integration domain

The choice of the hypercubes Υ and Ξ does not affect the consistency of the ICM test, 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 $\mathbb{R}^m$ and $\underline{\Xi}$ and $\overline{\Xi}$ are given hypercubes in $\mathbb{R}^k$, provided that it can be shown that under H_0,

$$\sup_{\underline{\Upsilon}\subset\Upsilon\subset\overline{\Upsilon},\ \underline{\Xi}\subset\Xi\subset\overline{\Xi}} \int_{\Upsilon\times\Xi} |Z_n(\tau,\xi)|^2 d\tau d\xi / \lambda\left(\Upsilon\times\Xi\right)$$

$$\overset{d}{\to} \sup_{\underline{\Upsilon}\subset\Upsilon\subset\overline{\Upsilon},\ \underline{\Xi}\subset\Xi\subset\overline{\Xi}} \int_{\Upsilon\times\Xi} |Z(\tau,\xi)|^2 d\tau d\xi / \lambda\left(\Upsilon\times\Xi\right),$$

where $\lambda\left(\Upsilon\times\Xi\right)$ is the Lebesgue measure of $\Upsilon\times\Xi$.

Indeed, this is true, as will be shown for the following special case.

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

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

Then under H_0,

$$\sup_{\underline{c} \leq c \leq \overline{c}} \widehat{T}_n(c) \xrightarrow{d} \sup_{\underline{c} \leq c \leq \overline{c}} T(c).$$

Although it is too much of a computational burden to compute this supremum exactly, 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.

4 The simulated ICM test

Quite a few 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.

To cope with this problem, a Simulated Integrated Conditional Moment (SICM) test is proposed, in which the conditional characteristic function

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

is replaced with $\exp(i\tau'\widetilde{Y}_j)$, where $\tilde{Y}_j$ is a random drawing from the estimated conditional distribution $F(y|X_j;\hat{\theta})$.

Thus, the empirical process $Z_n(\tau, \xi)$ is replaced with

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

The SICM test statistic is then

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

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:

With $Y_{\ell,j}$, $\widetilde{Y}_{\ell,j}$ and $X_{\ell,j}$ the components ℓ of Y_j, $\widetilde{Y}_j$ and X_j, respectively, we have

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

Theorem 2. *Let* $\widehat{Z}_n^{(s)}(\tau,\xi) = Z_n(\tau,\xi) - \widetilde{Z}_n^{(s)}(\tau,\xi)$, *where*

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

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 before,* Z_s *is a complex-valued zero mean Gaussian process, and* Z *and* Z_s *are independent. Under* H_1,

$$p\lim_{n\to\infty} \widehat{T}_n^{(s)}/n > 0.$$

4.1 Small sample performance

The null hypothesis is that the dependent variable Y is generated by the conditional Poisson model:

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

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

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

where

$$p(x) = (1 + \exp(-x))^{-1}, \ X \sim N(0, 1).$$

The sample size is $n = 200$, and the bootstrap sample size is $M = 500$.The number of replications is 200.

The SICM test involved is the MAXSICM test

$$\max\left\{\widehat{T}_n^{(s)}(5), \widehat{T}_n^{(s)}(10), \widehat{T}_n^{(s)}(15), \widehat{T}_n^{(s)}(20), \widehat{T}_n^{(s)}(25)\right\}$$

Both Y and X are first standardized by taking them in deviation of their sample means, dividing them by their sample standard errors, and then using the arctan transformation to make them bounded.

The simulated $\widetilde{Y}$ are transformed similarly, using the sample mean and sample standard error of the actual Y variable.

Table 1: MAXSICM test

Rejection %

	1%	5%	10%	DGP:
$H_1^{(0)}$	0	5	10	Poisson $(\exp(X))$
$H_1^{(1)}$	52	71	83	NB$(1, p(X))$
$H_1^{(2)}$	33	56	68	NB$(5, p(X))$
$H_1^{(3)}$	30	52	66	NB$(10, p(X))$

The empirical size is very good, and the small sample power against the negative binomial alternatives is what can be expected for such a small sample size.

Note that the decrease in power is do to the fact that the conditional distribution NB$(m, p(X))$ approaches the conditional Poisson distribution for $m \to \infty$.

5 Application to health economic count data models

A popular model for count data is the conditional Poisson distribution. This section applies the MAXSICM method to test whether a conditional Poisson model is correctly specified.

To the best of our knowledge, there is no other consistent specification test available for the conditional Poisson distribution.

The data source is the 1987-1988 National Medical Expenditure Survey. There are 4406 observations of individuals over the age of 66.

The variable Y of interest is the number of physician visits by elderly, which is explained by a vector of various variables of health conditions and demographic characteristics.

Y	# of visit to physicians in an office setting

X_1	health condition: excellent	X_9	$= 1$ if black
X_2	health condition: poor	X_{10}	$= 1$ if male
X_3	# of chronicle diseases	X_{11}	$= 1$ if married
X_4	disability status	X_{12}	years of schooling
X_5	region: northeast	X_{13}	family income
X_6	region: midwest	X_{14}	employment status
X_7	region: west	X_{15}	private insurance status
X_8	age	X_{16}	public insurance status

It is conceivable that the effects of the covariates X_3 through X_{16} are different for people with excellent health $(X_1 = 1)$ and poor health $(X_2 = 1)$.

Therefore, we have augmented the list of covariates with $X_1 \times X_j$ and $X_2 \times X_j$ for $j = 3, 4, .., 16$, so that the actual number of covariates is 44.

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

We will use the MAXSICM test

$$\max\left\{\widehat{T}_n^{(s)}(5), \widehat{T}_n^{(s)}(10), \widehat{T}_n^{(s)}(15), \widehat{T}_n^{(s)}(20), \widehat{T}_n^{(s)}(25)\right\}$$

to test the Poisson hypothesis, with bootstrap sample size 500.

It suffices to include only the original sixteen covariates as conditioning variables in the test .

The dependent variable Y and the sixteen X variables have been standardized and transformed in the same way as in the simulation study.

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

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

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

where $\widehat{\mu}_j = \exp((1, X_j')\widehat{\theta})$ with $\widehat{\theta}$ the ML estimator of θ_0.

Under the null hypothesis that the conditional expectation and the conditional variance of Y_j are equal the parameter α should be zero.

The test statistic involved is the t-value $\widehat{t}$ of the OLS estimate $\widehat{\alpha}$ of α. The results are

$$\widehat{\alpha} = 0.874068, \ \widehat{t} = 12.7497.$$

Thus, the Cameron-Trivedi test also strongly rejects the validity of the Poisson model.

As a further comparison we have also tried to conduct Andrews' (1997) Conditional Kolmogorov (CK) test

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

However, for the 16 covariates the inequality $X_j < X_i$ for $i\neq j$ never happened, so that the CK test statistic collapsed to

$$\max_{1\leq j\leq n}\left|1-F(Y_j|X_j,\widehat{\theta})\right|/\sqrt{n}<1/\sqrt{n}=0.015.$$

This problem does not occur for our ICM test.

If

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

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

$$Y|X \sim \text{NBL}\left(m\right),$$

If so, then

$$\widehat{\alpha} \xrightarrow{p} 1/m$$

where $\widehat{\alpha}$ is the OLS estimate of the parameter α in the Cameron-Trivedi model

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

Since $1/\widehat{\alpha} \approx 1.144$ is somewhat close to $m = 1$ we will now try a NBL(1) model.

The MAXSICM test statistic involved for the NBL(1) model is now 10.796, which is much lower than in the Poisson case.

However, the bootstrap p-value is still virtually zero, so that also the NBL(1) model is strongly rejected.

The same applies to the NBL(2) model: the MAXSICM test statistic is 15.990 with again virtually zero bootstrap p-value.

The estimation and test computations for these applications have been conducted via a modified version of *EasyReg International*, which can be downloaded freely from
http://econ.la.psu.edu/~hbierens/EASYREG.HTM
The modified EasyReg modules involved are available upon requests.

6 Concluding remarks

This paper extends the ICM specification test for the functional form of regression models to specification tests for parametric conditional distributions, on the basis of the integrated squared difference between the empirical characteristic function and the theoretical characteristic function.

This test is consistent, has $\sqrt{n}$-local power, and the conditional distributions to be tested can be of any type: continuous, discrete, or mixed.

The null distribution is case-dependent, so that the critical values have to be derived via a parametric bootstrap method.

To avoid numerical integration for computing the theoretical characteristic function, a 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.

The SICM 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 SICM test works well for models with a large number of covariates, contrary to Andrews' (1997) Conditional Kolmogorov test, and does not suffer from the curse of dimensionality as is the case with Zheng's (2000) test.