

Consistent Model Specification Tests

Herman J. Bierens

Pennsylvania State University

1 Conditional expectation models

1.1 Cross-section regression models

Given a dependent variable Y and a vector $X \in \mathbb{R}^k$ of explanatory variables, a parametric nonlinear regression model takes the form

$$Y = f(X, \theta_0) + U,$$

where

- θ_0 is an unknown parameter vector contained in a parameter space $\Theta \subset \mathbb{R}^m$,
- U is an error term satisfying $E[U^2] < \infty$,
- $f(x, \theta)$ is a known function on $\mathbb{R}^k \times \Theta$.

For example, in the case of a linear regression model,

$$f(x, \theta) = \alpha + \beta'x, \quad \theta = (\alpha, \beta)', \quad \Theta = \mathbb{R}^{k+1}.$$

The model $Y = f(X, \theta_0) + U$ is correctly specified if

$$H_0 : \exists \theta_0 \in \Theta : \Pr (f (X, \theta_0) = E[Y|X]) = 1,$$

and it is misspecified if

$$H_1 : \forall \theta \in \Theta : \Pr (f (X, \theta) = E[Y|X]) < 1.$$

In either case, let

$$\theta_0 = \arg \min_{\theta \in \Theta} E \left[(Y - f (X, \theta))^2 \right].$$

Then

$$H_0 \Leftrightarrow \Pr (E[U|X] = 0) = 1.$$

$$H_1 \Leftrightarrow \Pr (E[U|X] = 0) < 1.$$

1.2 The integrated conditional moment (ICM) test

The question is: How can we test the correctness of the functional specification of the model $Y = f(X, \theta_0) + U$ such that the test has asymptotic power 1 against H_1 .

The first paper to address this problem is:

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

The approach in that paper is based on the uniqueness of the Fourier transform of a function.

Let

$$g(X) = E[U|X],$$

with Fourier transform

$$\begin{aligned}\varphi(\xi) &= E[g(X) \exp(i.\xi'X)] \\ &= E[U \exp(i.\xi'X)], \quad t \in \mathbb{R}^k, \quad i = \sqrt{-1}.\end{aligned}$$

Then

$$H_0 \Leftrightarrow \sup_{\xi \in \mathbb{R}^k} |\varphi(\xi)| = 0$$

$$H_1 \Leftrightarrow \sup_{\xi \in \mathbb{R}^k} |\varphi(\xi)| > 0$$

Question: Where to look for a $\xi \in \mathbb{R}^k$ such that $|\varphi(\xi)| > 0$?

Answer:

- If X is bounded then

$$H_1 \Leftrightarrow \forall \delta > 0, \sup_{|\xi| \leq \delta} |\varphi(\xi)| > 0$$

- If X is not bounded, let $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}^k$ be a bounded one-to-one mapping and let

$$\varphi(\xi) = E [U \exp (i \cdot \xi' \Phi(X))], \quad \xi \in \mathbb{R}^k.$$

Then

$$H_1 \Leftrightarrow \forall \delta > 0, \sup_{|\xi| \leq \delta} |\varphi(\xi)| > 0$$

Denote

$$\Xi = \times_{l=1}^k [-\bar{\xi}_l, \bar{\xi}_l], \bar{\xi}_l > 0,$$

and let $\mu(\xi)$ be the uniform probability measure on Ξ . Then

$$H_0 \Leftrightarrow \int_{\Xi} |E[U \cdot \exp(i \cdot \xi' \Phi(X))]|^2 d\mu(\xi) = 0$$

$$H_1 \Leftrightarrow \int_{\Xi} |E[U \cdot \exp(i \cdot \xi' \Phi(X))]|^2 d\mu(\xi) > 0$$

These results suggest that, given a random sample

$$(Y_1, X_1), \dots, (Y_n, X_n)$$

from (Y, X) , a consistent test can be based on the integrated conditional moment (ICM) statistic

$$\hat{T}_n = \int_{\Xi} \left| \hat{Z}_n(\xi) \right|^2 d\mu(\xi),$$

where

$$\hat{Z}_n(\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \hat{U}_j \exp(i \cdot \xi' \Phi(X_j)),$$

with

$$\hat{U}_j = Y_j - f(X_j, \hat{\theta}_n)$$

the NLLS residuals, and $\hat{\theta}_n$ the NLLS estimator of θ_0 .

In Bierens (1982) I showed that

$$\begin{aligned} H_0 &\Rightarrow \hat{T}_n \xrightarrow{d} T \\ H_1 &\Rightarrow \hat{T}_n/n \xrightarrow{p} \eta > 0 \end{aligned}$$

However, at that time I was only able to derive an expression for $E[T]$ but not for the limiting null distribution T itself.

Therefore, I proposed to derive critical values of the ICM test on the basis of Chebyshev inequality for first moments.

It took me until 1990 to figure out what the nature of T is, for the real $\exp(\cdot)$ case:

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

In this paper I showed that under H_0 the empirical process

$$\widehat{Z}_n(\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \widehat{U}_j \exp(\xi' \Phi(X_j)),$$

is tight on Ξ , and converges weakly to a zero-mean Gaussian process $Z(\xi)$:

$$\widehat{Z}_n \Rightarrow Z.$$

Hence under H_0 ,

$$\widehat{T}_n = \int_{\Xi} \widehat{Z}_n(\xi)^2 d\mu(\xi) \xrightarrow{d} T = \int_{\Xi} Z(\xi)^2 d\mu(\xi)$$

whereas under H_1 ,

$$\widehat{T}_n/n \xrightarrow{p} \eta > 0.$$

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

In this paper it has been shown that my previous results carry over to more general empirical processes of the type

$$\widehat{Z}_n(\xi) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \widehat{U}_j w(\xi' \Phi(X_j)),$$

where $w(u)$ is a power function in an open neighborhood of $u = 0$ such that

$$(d/du)^s w(u) |_{u=0} \neq 0$$

for all but a finite number of natural numbers s . See also

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

Moreover, for the case of real valued weight functions $w(u)$ Bierens and Ploberger (1997) showed that

$$T = \int_{\Xi} Z(\xi)^2 d\mu(\xi) = \sum_{i=1}^{\infty} \lambda_i \varepsilon_i^2$$

where

- the ε_i 's are i.i.d. $N(0, 1)$ distributed
the λ_i 's are the eigenvalues of the covariance function

$$\Gamma(\xi_1, \xi_2) = E [Z(\xi_1)Z(\xi_2)]$$

i.e., the λ_i 's are the solutions of the eigenvalue problem

$$\lambda \cdot \psi(\xi_1) = \int_{\Xi} \Gamma(\xi_1, \xi_2) \psi(\xi_2) d\mu(\xi_2)$$

with corresponding eigenfunctions $\psi_i(\xi)$ satisfying

$$\int_{\Xi} \psi_i(\xi) \psi_j(\xi) d\mu(\xi) = I(i = j).$$

Moreover, by Mercer's theorem

$$\Gamma(\xi_1, \xi_2) = E [Z(\xi_1)Z(\xi_2)] = \sum_{i=1}^{\infty} \lambda_i \psi(\xi_1) \psi(\xi_2)$$

Bierens and Ploberger (1997) proposed to use the ICM statistic

$$\tilde{T}_n = \frac{\int_{\Xi} \widehat{Z}_n(\xi)^2 d\mu(\xi)}{\int_{\Xi} \widehat{\Gamma}(\xi, \xi) d\mu(\xi)}$$

where $\widehat{\Gamma}(\xi, \xi)$ is a consistent estimator of

$$\Gamma(\xi, \xi) = E [Z(\xi)^2] = \sum_{i=1}^{\infty} \lambda_i \psi(\xi)^2.$$

Then under H_0

$$\tilde{T}_n \xrightarrow{d} \frac{\sum_{i=1}^{\infty} \lambda_i \varepsilon_i^2}{\sum_{i=1}^{\infty} \lambda_i} \leq \sup_{\ell \geq 1} \frac{1}{\ell} \sum_{i=1}^{\ell} \varepsilon_i^2 = \bar{T}$$

Upper bounds of the critical values can now be based on \bar{T} .

Other results:

- The ICM test is admissible, i.e., there does not exist a uniformly more powerful test.
- The ICM test has non-trivial power against \sqrt{n} local alternatives:

$$H_1^L : E[Y|X] = f(X, \theta_0) + g(X)/\sqrt{n},$$
$$\Pr [g(X) = 0] < 1.$$

1.3 Consistent ICM tests of the martingale difference hypothesis

Time series regression models take the form

$$Y_t = f_{t-1}(\theta_0) + U_t$$

where

$$f_{t-1}(\theta_0) = E(Y_t | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots)$$

with

$$Z_t = (Y_t, X_t)'$$

Therefore, to test the validity of the specification $f_{t-1}(\theta)$ consistently one has to test the martingale difference hypothesis

$$E(U_t | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots) = 0$$

Most papers in the literature "solve" this problem by testing

$$E(U_t | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots, Z_{t-\ell}) = 0$$

for some finite ℓ . However, these tests are not consistent.

There are only two papers that propose consistent tests of the martingale different hypothesis:

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

In this paper I propose to compute for each hypothesis

$$E(U_t | Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots, Z_{t-\ell}) = 0$$

an ICM test $\hat{T}_{n,\ell}$, and then use

$$\sum_{\ell=1}^{L_n} \alpha_\ell \hat{T}_{n,\ell}$$

as the actual test statistic, where α_ℓ is an a priori chosen sequence of positive constants satisfying $\sum_{\ell=1}^{\infty} \alpha_\ell < \infty$, and $L_n \rightarrow \infty$ as $n \rightarrow \infty$.

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

In this paper Robert de Jong extends the ICM test to infinite dimensional empirical processes $\hat{Z}_n(\xi)$.

1.4 Consistent tests based on kernel estimators

Up to the early nineties, my papers Bierens (1982, 1984, 1990) were the only literature on consistent model specification testing.

In the mid-nineties a related strand of statistics literature emerged, starting with

Stute, W., 1997, Nonparametric Model Checks for Regression, *Annals of Statistics*, 25, 613-641.

Stute's approach is based on the fact that

$$\Pr (E[U|X] = 0) < 1 \Leftrightarrow \exists x \in \mathbb{R}^k : E [U.I(X \leq x)] \neq 0.$$

Therefore, various consistent test can be based on the empirical process

$$I_n(x) = n^{-1/2} \sum_{j=1}^n \hat{U}_j I(X_j \leq x),$$

including an ICM type test.

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

combined the Stute approach with my ICM approach.

Another strand of literature on model specification testing emerged in the nineties in the statistics and econometric literature based on comparisons, in various ways, of parametric functional forms and corresponding nonparametric estimates.

However, these tests have only nontrivial power against local alternatives that approach the null at a slower rate than \sqrt{n} , and this the rate decreases with the dimension of X .

2 Consistent tests for conditional distribution models

2.1 Introduction

Maximum likelihood requires that the conditional distribution of a (vector of) dependent variable(s) $Y \in \mathbb{R}^m$ given a vector $X \in \mathbb{R}^k$ of explanatory variables is completely parametrized, as

$$\Pr [Y \leq y|X] = F(y|X, \theta_0)$$

for example, where

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

is a parametric family of conditional distribution functions, with $\Theta \subset \mathbb{R}^p$ a parameter space.

To test the validity of this specification consistently, we need to test the null hypothesis:

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

a.s.

against the general alternative hypothesis that H_0 is false:

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

a.s.,

given a random sample

$$(Y_1, X_1), \dots, (Y_n, X_n)$$

from (Y, X) .

2.2 Published literature

The published literature on consistently testing parametric conditional distribution specifications consists only of two papers:

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

proposes a consistent test by generalizing the Kolmogorov test to:

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

This test is consistent and has nontrivial \sqrt{n} local power. However, a practical problem is that 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

proposes a test for the validity of conditional densities by comparing a parametric conditional density with a corresponding nonparametric kernel estimator.

Thus, this test is only applicable to absolutely continuous conditional distribution models.

Zheng's test has non-trivial local power, but only against local alternatives that approach the null at a slower rate than \sqrt{n} .

2.3 The ICM test for conditional distribution models

As an alternative to Andrews' (1997) and Zheng's (2000) tests,

Bierens, H. J., and L. Wang, 2009, "Integrated Conditional Moment Tests for Parametric Conditional Distributions", submitted for publication,

propose an ICM test based on the integrated square difference of the empirical characteristic function of the data and the empirical characteristic function implied by the estimated model.

The hypotheses:

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

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

are equivalent to:

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

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

respectively

These hypotheses are equivalent to:

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

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

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

Then for arbitrary $\varepsilon > 0$,

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

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 $\varepsilon > 0$,

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

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

Denote

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

$$\Upsilon = \times_{j=1}^m [-\bar{\tau}_j, \bar{\tau}_j], \bar{\tau}_j > 0,$$

$$\Xi = \times_{j=1}^k [-\bar{\xi}_j, \bar{\xi}_j], \bar{\xi}_j > 0,$$

and let $\mu(\tau, \xi)$ be the uniform distribution on $\Upsilon \times \Xi$.

Then

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

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

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

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

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

2.4 Asymptotic properties

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 ,

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

hence

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

2.5 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 to preserve enough variation in $\Phi_1(Y)$ and $\Phi_2(X)$.

In particular, in the case $Y \in \mathbb{R}$, 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 .

As long as we choose μ_n and σ_n such that

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

this standardization does not effect the asymptotic properties.

2.6 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

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

Consider the eigenvalue problem:

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

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

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

Properties:

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

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

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

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

- Mercer's theorem:

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

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

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

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

Therefore,

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

where the $\chi_{2,j}^2$'s are independently χ_2^2 distributed.

2.7 Parametric bootstrap

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

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

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

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

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

2.8 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 $E[Z(\tau, \xi)] \neq 0$.

Denote

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

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

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.

2.9 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 ,

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

where $\lambda(\Upsilon \times \Xi)$ 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 $c \in [\underline{c}, \bar{c}]$, with $0 < \underline{c} < \bar{c} < \infty$ given constants, and let

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

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

Then under H_0 ,

$$\sup_{\underline{c} \leq c \leq \bar{c}} \hat{T}_n(c) \xrightarrow{d} \sup_{\underline{c} \leq c \leq \bar{c}} T(c).$$

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 $\hat{T}_n(c)$ for these values as the actual ICM test.

2.10 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

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

is replaced with $\exp(i\tau'\tilde{Y}_j)$, 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 is feasible to generate random drawings \widetilde{Y} from it.

Another advantage is that $\widehat{T}_n^{(s)}$ has a closed form:

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

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

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

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

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

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

Under H_1 ,

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

2.11 Small sample performance

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

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

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

$$H_1^{(0)} : Y|X \sim \text{Poisson}(\exp(X))$$

$$H_1^{(1)} : Y|X \sim \text{NB}(1, p(X))$$

$$H_1^{(2)} : Y|X \sim \text{NB}(5, p(X))$$

$$H_1^{(3)} : Y|X \sim \text{NB}(10, p(X))$$

where

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

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

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

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 \tilde{Y} are transformed similarly, using the sample mean and sample standard error of the actual Y variable.

Table 1: MAXSICM test

	Rejection %			DGP:
	1%	5%	10%	
$H_1^{(0)}$	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 $m \rightarrow \infty$.

2.12 Application to health economic count data models

A popular model for count data is the conditional Poisson distribution.

The MAXSICM test will be used to test whether a conditional Poisson model is correctly specified.

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 $j = 3, 4, \dots, 16$, 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 $\mu(X) = \exp((1, X')\theta_0)$.

We will use the MAXSICM test

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

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

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

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

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

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

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

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

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

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

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

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

This problem does not occur for our ICM test.

If

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

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:

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

If so, then

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

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

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

Since $1/\hat{\alpha} \approx 1.144$ 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.