Review of Integrated Conditional Moment Tests for Consistently Testing the Validity of Parametric Regression and Conditional Probability Models

Herman J. Bierens

Professor Emeritus of Economics
Pennsylvania State University

Introduction

As is well-known, misspecification of parametric econometric models may lead to incorrect inference about the model parameters.

In particular, if a parametric econometric model is misspecified the usual test for parameter restrictions such as the t-test, the Wald, Lagrange multiplier and likelihood ratio tests do no longer have their standard null distributions.

Moreover, if a misspecified model is used for forecasting, their conditional forecasts will become biased and inefficient.

Therefore, it is of utmost importance to test for model misspecification before using the model for inference and forecasting.

Regression models aim to represent conditional expectations.

Therefore, a cross-section regression model is correctly specified if the conditional expectation of the error term given the regressors is zero with probability 1.
Thus, strictly speaking, the model is misspecified if the latter probability is less than 1.

In the time series case a regression model is correctly specified if the conditional expectation of the error term given the exogenous variables *and all* lagged dependent and exogenous variables is zero with probability 1.
Again, the model is misspecified if the latter probability is less than 1.

In other words, time series regression errors need to be *martingale differences*.

Now the following questions arise:

- Can we (asymptotically) distinguish the hypothesis that a conditional expectation equals zero with probability 1 and the alternative hypothesis that this probability is less than 1 (for example 0.999)?
- If so, how can we devise tests for the null hypothesis that a regression model is correctly specified against the alternative that the null hypothesis is false, such that these tests are consistent, i.e. their asymptotic power is 1?

The answer to these questions is: **Yes we can!**

My paper

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

and its companion paper

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

are to the best of my knowledge the first papers ever to propose consistent tests of the null hypothesis that the functional form of a regression model is correctly specified as a conditional expectation, against all deviations from the null hypothesis.

However, at that time I did not know how to derive the limiting null distributions of the test statistics involved, but I did know how to consistently estimate their expectations under the null hypothesis.

Therefore, in these papers I proposed to use upper bounds of the critical values based on Chebyshev's inequality for first moments.

In

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

I finally figured out what the null distribution of the test in Bierens (1982) looks like, but I was still not able to derive its critical values.

Up to the early nineties the only papers on consistent model specification testing were Bierens (1982, 1984, 1990).
After 1990 two strands of econometric and statistical literature emerged:

(1) De Jong (1996), Bierens and Ploberger (1997), Stute (1997), Stinchcombe and White (1998), Boning and Sowell (1999), Fan and Li (2000) and Escanciano (2006) for integrated conditional moment (ICM) and related tests.

(2) Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Gozalo (1993), Horowitz and Härdle (1994), Hong and White (1995), Li and Wang (1998), Zheng (1996) and Lavergne and Vuong (2000), among others, for tests based on comparisons of parametric functional forms with corresponding nonparametric or semi-parametric estimates.

In

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

Werner Ploberger and I showed for a version of the integrated conditional moments (ICM) test in Bierens (1982) that

- the ICM test has nontrivial power against $\sqrt{n}$ local alternatives, i.e., alternatives of the form
  $Y = f(X, \theta_0) + g(X)/\sqrt{n} + U$, with $\Pr(E[U|X] = 0) = 1$.
- the ICM test is admissible, i.e., there does not exist an uniformly more powerful test, and
- the null distribution of the ICM test takes the form of a weighted sum of independent χ_1^2 distributed random variables, but with case-dependent weights.

With the results in Bierens and Ploberger (1997) at hand, my Ph.D. student Li Wang and I have been able to extend the approach in Bierens (1982,1984) to consistently testing the correctness of parametric conditional distribution models, for cross-section models in:

Bierens, H. J., and Wang, L., 2012, Integrated Conditional Moment Tests for Parametric Conditional Distributions, *Econometric Theory* 28, 328-362,

and for time series models in

Bierens, H. J., and Wang, L., 2017, Weighted Simulated Integrated Conditional Moment Tests for Parametric Conditional Distributions of Stationary Time Series Processes, *Econometric Reviews* 36, 103-135.

In this talk I will focus on the papers Bierens (1982,1984) and Bierens and Wang (2012,2017), together with their updates in:

Bierens, H.J. 2017, *Econometric Model Specification: Consistent Model Specification Tests and Semi-Nonparametric Modeling and Inference.* World Scientific Publishers, Singapore.

This book contains reprints of the relevant published papers up to 2014, where the papers on consistent model specification tests come with extensive addendums, bringing these papers up to date.

Cross-section regression models

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

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

where

- $E[Y^2] < \infty$;
- θ_0 is an unknown parameter vector contained in the interior of a given convex and compact parameter space $\Theta \subset \mathbb{R}^m$,
- $f(x, \theta)$ is an a priori chosen continuous function on $\mathbb{R}^k \times \Theta$ satisfying $E[(f(X, \theta))^2] < \infty$ for all $\theta \in \Theta$,
- U is the error term.

Without loss of generality we may define

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

regardless whether the model is misspecified or not.

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

$$H_0 : \Pr(E[U|X] = 0) = 1,$$

which is equivalent to

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

and the model is misspecified if

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

which is equivalent to

$$H_1 : \Pr(E[Y|X] = f(X, \theta_0)) < 1.$$

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.

My paper Bierens (1982) is the first paper ever to address this problem.

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

The Fourier transform of the function

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

takes the form

$$\begin{aligned} \varphi(\tau) &= E\left[g(X)\exp\left(\mathbf{i}.\tau' X\right)\right] \\ &= E\left[U\exp\left(\mathbf{i}.\tau' X\right)\right],\ \tau \in \mathbb{R}^k,\ \mathbf{i} = \sqrt{-1}. \end{aligned}$$

Then

$$\sup_{\tau\in\mathbb{R}^k} |E[U\exp(\mathbf{i}.\tau' X)| \;=\; 0 \text{ under } H_0,$$
$$\sup_{\tau\in\mathbb{R}^k} |E[U\exp(\mathbf{i}.\tau' X)| \;>\; 0 \text{ under } H_1.$$

Question:

Where to look for a $\tau \in \mathbb{R}^k$ such that $E[U\exp(\mathbf{i}.\tau' X)] \neq 0$ if H_1 is true?

My answer in 1982:

- If X is bounded then under H_1,

$$\forall\delta > 0, \sup_{||\tau||\leq\delta} |E[U.\exp(\mathbf{i}.\tau' X)]| > 0.$$

- If X is not bounded, let $\Phi : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ be a bounded one-to-one mapping with Borel measurable inverse Φ^{-1}, so that $E[U|X] = E[U|\Phi(X)]$ with probability 1. For example, let

$$\Phi(x) = (\arctan(x_1), ..., \arctan(x_k))'.$$

  Then under H_1,

$$\forall\delta > 0, \sup_{||\tau||\leq\delta} |E[U.\exp(\mathbf{i}.\tau'\Phi(X))]| > 0.$$

Thus under H_1,

$$E[U.\exp(\mathbf{i}.\tau'\Phi(X))] \neq 0$$

for a τ in an arbitrary neighborhood of the origin of $\mathbb{R}^k$.

My answer in 1990:

Similar to Bierens (1990), we have the more general result that under H_1 the set

$$S = \left\{\tau \in \mathbb{R}^k : E\left[U.\exp\left(\mathbf{i}.\tau'\Phi(X)\right)\right] = 0\right\}$$

has Lebesgue measure zero and is nowhere dense, whereas of course under H_0,

$$S = \mathbb{R}^k.$$

This implies that for any compact subset Υ of $\mathbb{R}^k$ with positive Lebesgue measure, and with $\mu(\tau)$ be the uniform probability measure on Υ, for example,

$$\int_\Upsilon \left|E\left[U.\exp\left(\mathbf{i}.\tau'\Phi(X)\right)\right]\right|^2 \mathrm{d}\mu(\tau) \ = \ 0 \text{ under } H_0,$$

$$\int_\Upsilon \left|E\left[U.\exp\left(\mathbf{i}.\tau'\Phi(X)\right)\right]\right|^2 \mathrm{d}\mu(\tau) \ > \ 0 \text{ under } H_1.$$

These results suggest that, given a random sample $\{(Y_j, X_j)\}_{j=1}^{n}$ from (Y, X), a consistent test can be based on the integrated conditional moment (ICM) statistic

$$\begin{aligned}\widehat{T}_n &= \int_{\Upsilon} \left|\widehat{W}_n(\tau)\right|^2 \mathrm{d}\mu(\tau), \text{ where} \\ \widehat{W}_n(\tau) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \widehat{U}_j \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right)\end{aligned}$$

with $\widehat{U}_j$ the NLLS residual.

In Bierens (1982) I showed that

$$\widehat{T}_n = \int_{\Upsilon} \left|\widehat{W}_n(\tau)\right|^2 \mathrm{d}\mu(\tau) \xrightarrow{\mathrm{d}} T \text{ under } H_0,$$

$$\widehat{T}_n/n \xrightarrow{\mathrm{p}} \eta > 0 \text{ under } H_1.$$

However, at that time I was only able to derive a consistent estimate of $E[T]$ but I could not derive the limiting null distribution T itself.

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

This is how far I got in 1982.

The null distribution of the ICM test

It took me until 1990 to figure out what the nature of T is, namely, similar to Bierens (1990) it follows that under H_0 the empirical process

$$\widehat{W}_n(\tau) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \widehat{U}_j \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right),$$

converges weakly to a zero-mean complex-valued Gaussian process $W(\tau)$, so that by the continuous mapping theorem,

$$\widehat{T}_n = \int_\Upsilon \left|\widehat{W}_n(\tau)\right|^2 \mathrm{d}\mu(\tau) \overset{\mathrm{d}}{\to} T = \int_\Upsilon |W(\tau)|^2 \mathrm{d}\mu(\tau)$$

The zero mean complex-valued Gaussian process $W(\tau)$ is characterized by its covariance function

$$\Gamma(\tau_1, \tau_2) = E\left[W(\tau_1)\overline{W(\tau_2)}\right],$$

where the bar denotes the complex-conjugate.

Similar to symmetric positive semi-definite matrices, this covariance function

$$\Gamma(\tau_1, \tau_2) = E\left[W(\tau_1)\overline{W(\tau_2)}\right]$$

has countable many nonnegative eigenvalues and corresponding orthonormal eigenfunctions.

This eigenvalue problem reads:
Find an eigenvalue λ and corresponding nonzero eigenfunction $\varphi(\tau)$ such that

$$\int_{\Upsilon} \Gamma(\tau_1, \tau_2)\varphi(\tau_2)\mathrm{d}\mu(\tau_2) = \lambda\varphi(\tau_1) \text{ for all } \tau_1 \in \Upsilon.$$

This problem has countable many real valued nonnegative solutions λ_i, $i \in \mathbb{N}$, with corresponding orthonormal eigenfunctions $\varphi_i(\tau)$,

$$\int_{\Upsilon} |\varphi_i(\tau)|^2\mathrm{d}\mu(\tau) = 1,\ \int_{\Upsilon} \varphi_{i_1}(\tau)\overline{\varphi_{i_2}(\tau)}\mathrm{d}\mu(\tau) = 0 \text{ if } i_1 \neq i_2.$$

According to the complex version of Mercer's theorem,

- The covariance function $\Gamma(\tau_1, \tau_2)$ can be written as

$$\Gamma(\tau_1, \tau_2) = \sum_{m=1}^{\infty} \lambda_m \varphi_m(\tau_1) \overline{\varphi_m(\tau_2)},$$

  hence $\sum_{m=1}^{\infty} \lambda_m = \int_\Upsilon \Gamma(\tau, \tau) \mathrm{d}\mu(\tau) < \infty$.

- The eigenfunctions $\varphi_i(\tau)$ form a complete orthonormal sequence in the complex Hilbert space $L^2(\mu)$.

Since $W \in L^2(\mu)$, the latter property implies that we can write

$$W(\tau) = \sum_{i=1}^{\infty} a_i \varphi_i(\tau),$$

where

$$a_i = \int_\Upsilon W(\tau) \overline{\varphi_i(\tau)} d\mu(\tau),$$

which are zero mean complex valued Gaussian random variables.

Then by the first Mercer property,

$$\begin{aligned} E[\overline{a}_j a_i] &= \int_\Upsilon \int_\Upsilon \varphi_j(\tau_1) E[\overline{W(\tau_1)} W(\tau_2)] \overline{\varphi_i(\tau_2)} \mathrm{d}\mu(\tau_1) \mathrm{d}\mu(\tau_2) \\ &= \sum_{m=1}^{\infty} \lambda_m \int_\Upsilon \varphi_j(\tau) \overline{\varphi_m(\tau)} \mathrm{d}\mu(\tau) \int_\Upsilon \varphi_i(\tau) \overline{\varphi_m(\tau)} \mathrm{d}\mu(\tau) \\ &= \sum_{m=1}^{\infty} \lambda_m I(j=m) I(i=m) = \lambda_i I(i=j) \end{aligned}$$

Denoting

$$g_m = \frac{a_m}{\sqrt{\lambda_m}} = \frac{\int_\Upsilon W(\tau) \overline{\varphi_m(\tau)} \mathrm{d}\mu(\tau)}{\sqrt{\lambda_m}} \text{ if } \lambda_m > 0$$

we can now write

$$W(\tau) = \sum_{m=1}^{\infty} \sqrt{\lambda_m} g_m \varphi_m(\tau),$$

where the g_m's are independent zero mean complex valued normal random variables with variances

$$E\left[g_m \overline{g_m}\right] = E\left[|g_m|^2\right] = 1.$$

Therefore,

$$
\begin{aligned}
T &= \int_\Upsilon |W(\tau)|^2 \, \mathrm{d}\mu(\tau) \sim \sum_{m=1}^{\infty} \lambda_m |g_m|^2, \\
E[T] &= \int_\Upsilon \Gamma(\tau,\tau) \mathrm{d}\mu(\tau) = \sum_{i=1}^{\infty} \lambda_i < \infty.
\end{aligned}
$$

It can be shown that

$|g_m|^2 \sim \kappa_m e_{1,m}^2 + (1-\kappa_m)\, e_{2,m}^2$ for some $\kappa_m \in [0,1]$,

where the $e_{i,m}$'s are i.i.d. $\mathcal{N}(0,1)$, and κ_m and $1-\kappa_m$ are the eigenvalues of Var$\left((\mathrm{Re}[g_m], \mathrm{Im}[g_m])'\right)$.

Thus,

$$
\begin{aligned}
T &= \int_\Upsilon |W(\tau)|^2 \, \mathrm{d}\mu(\tau) \sim \sum_{m=1}^{\infty} \lambda_m |g_m|^2 \\
&\sim \sum_{m=1}^{\infty} \lambda_m \kappa_m e_{1,m}^2 + \sum_{m=1}^{\infty} \lambda_m (1-\kappa_m)\, e_{2,m}^2 = \sum_{m=1}^{\infty} \omega_m e_m^2, \text{ say,}
\end{aligned}
$$

where the e_m's are i.i.d. $\mathcal{N}(0,1)$.

Hence,

$$\frac{T}{E[T]} \sim \frac{\sum_{m=1}^{\infty} \omega_m e_m^2}{\sum_{m=1}^{\infty} \omega_m} \leq \sup_{n\geq 1} \frac{1}{n} \sum_{m=1}^{n} e_m^2 = \overline{\chi}_1^2, \text{ say,}$$

where the inequality follows from a result in Bierens and Ploberger (1997).

Upper bounds of the critical values

Therefore, given a consistent estimator $\widehat{\Gamma}_n(\tau,\tau)$ of $\Gamma(\tau,\tau)$, and denoting

$$\widetilde{T}_n = \frac{\int_\Upsilon \left|\widehat{W}_n(\tau)\right|^2 \mathrm{d}\mu(\tau)}{\int_\Upsilon \widehat{\Gamma}_n(\tau,\tau)\mathrm{d}\mu(\tau)}$$

we have

$$\limsup_{n\to\infty} \Pr\left[\widetilde{T}_n > y\right] \leq \Pr\left[\overline{\chi}_1^2 > y\right].$$

Thus, upper bounds of the critical values of $\widetilde{T}_n$ can be based on the quantiles of the distribution of $\overline{\chi}_1^2$.

These upper bounds, $\overline{c}(\alpha)$ say, of the $\alpha \times 100\%$ critical values for $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$ have been calculated in Bierens and Ploberger (1997), i.e.,

$$\overline{c}(0.01) = 6.81,\ \overline{c}(0.05) = 4.26,\ \overline{c}(0.10) = 3.23.$$

Bootstrap critical values

Instead of using upper bounds of the critical values, it is possible to approximate the actual critical values of

$$T = \int_{\Upsilon} |W(\tau)|^2 \, \mathrm{d}\mu(\tau)$$

via a parametric bootstrap method, as follows.

First, we need to eliminate the estimation error

$$f(X_j, \widehat{\theta}_n) - f(X_j, \theta_0)$$

from the empirical process

$$\begin{aligned}
\widehat{W}_n(\tau) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \widehat{U}_j \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right) \\
&\quad - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Big(f(X_j, \widehat{\theta}_n) - f(X_j, \theta_0) \Big) \exp\left(\mathbf{i}.\tau'\Phi(X_j)\right)
\end{aligned}$$

In particular, construct a complex empirical process

$$W_n(\tau) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \phi_{j,n}(\tau),$$

such that $\widehat{W}_n(\tau) = W_n(\tau) + o_p(1)$ uniformly in $\tau \in \Upsilon$, where the new weight functions $\phi_{j,n}(\tau)$ depend on the X_i's in the sample only, next to τ of course.

For example, if the null model is linear:

$$Y_j = (1, X_j')\theta_0 + U_j$$

then $\widehat{W}_n(\tau) = W_n(\tau) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_j \phi_{j,n}(\tau)$ where

$$\phi_{j,n}(\tau) = \exp\left(\mathbf{i}.\tau' \Phi(X_j)\right) - b_n(\tau)' A_n^{-1} \begin{pmatrix} 1 \\ X_j \end{pmatrix}, \text{ with}$$

$$A_n = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ X_i \end{pmatrix} (1, X_i'),\ b_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ X_i \end{pmatrix} \exp\left(\mathbf{i}.\tau' \Phi(X_i)\right).$$

Next, for given bootstrap sample size M and $m = 1, 2, ..., M$, let

$$\widetilde{W}_{m,n}(\tau) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_{m,j}\widehat{U}_j\phi_{j,n}(\tau),\ \widetilde{T}_{m,n} = \int_{\Upsilon}\left|\widetilde{W}_{m,n}(\tau)\right|^2 \mathrm{d}\mu(\tau).$$

where the $\varepsilon_{m,j}$'s are i.i.d. $\mathcal{N}(0,1)$.

Then under H_0,

$$\left(\widehat{T}_n, \widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}\right)' \xrightarrow{\mathrm{d}} (T_0, T_1, T_2, ..., T_M)'$$

where the T_m's for $m = 0, 1, ..., M$ are i.i.d. T, whereas under H_1,

$$\left(\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}\right)' \xrightarrow{\mathrm{d}} (T_1^*, T_2^*, ..., T_M^*)'$$

where the T_m^*'s for $m = 1, ..., M$ are i.i.d. (but no longer distributed as T).

The bootstrap critical values of the ICM test can now be based on the quantiles of the empirical distribution function of $\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}$.

Standardization of X *in* $\Phi(X)$

As suggested before, a suitable choice for $\Phi(X)$ is

$$\Phi(X) = (\arctan(X_1), ..., \arctan(X_k))'$$

where X_i is component i of X.

However, if X_i takes large positive values then for these values $\arctan(X_i) \approx \pi/2$.

For example suppose that the data come from a household survey, where X_i is the monthly dollar income, and suppose that $X_i > 100$ for all observations. Then $\arctan(X_i) > 1.56$ whereas $\pi/2 \approx 1.57$, so that $\arctan(X_i)$ is virtually constant.

But in this case conditioning on X is effectively no longer equivalent to conditioning on $\Phi(X)$, which is detrimental to the finite sample power and even the consistency of the ICM test.

To avoid this problem, it is recommended to standardize the X_i's in Φ as

$$\widetilde{X}_{i,n} = (X_i - \overline{X}_{i,n})/S_{i,n}$$

where $\overline{X}_{i,n}$ is the sample mean and $S_{i,n}$ is the sample standard error of X_i, before taking the $\arctan(.)$ transformations.

Then the effect on the asymptotic properties of the ICM test is the same as in the case where the X_i's would have been standardized as

$$\overline{X}_i = (X_i - E[X_i])/\sqrt{\text{var}(X_i)}.$$

Time series regression models

Time series regression models take the form

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

where the response function $f_{t-1}(\theta)$ is a parametric specification of the conditional expectation

$$E(Y_t | Z_{t-1}, Z_{t-2}, Z_{t-3}, ...), \text{ where } Z_t = (Y_t, X_t')',$$

with X_t a possible vector of exogenous variables.

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

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

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

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

only, for some fixed ℓ.

However, these tests are not consistent.

The AR(1) model as benchmark model

In this talk I will explain the Weighted ICM (WICM) test in the context of an AR(1) model.

Thus, given a strictly *and* covariance stationary time series process Y_t, the null hypothesis to be tested is that

$$H_0 : E[Y_t|Y_{t-1}, Y_{t-2}, Y_{t-3}, ...] = \alpha_0 + \beta_0 Y_{t-1} \text{ a.s.},$$

where

$$\theta_0 = (\alpha_0, \beta_0)' = \arg \min_{(\alpha,\beta)' \in \mathbb{R}\times(-1,1)} E\left[(Y_t - \alpha - \beta Y_{t-1})^2\right],\ |\beta_0| < 1.$$

Denoting the error term by $U_t = Y_t - \alpha_0 - \beta_0 Y_{t-1}$ and the σ-algebra generated by the sequence $\{Y_{t-i}\}_{i=1}^{\infty}$ by

$$\mathcal{F}_{-\infty}^{t-1} = \sigma\left(\{Y_{t-i}\}_{i=1}^{\infty}\right),$$

the null hypothesis to be tested is that U_t is a martingale difference process w.r.t. the filtration $\mathcal{F}_{-\infty}^{t-1}$:

$$H_0 : \Pr\left(E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right] = 0\right) = 1.$$

Now suppose that this null hypothesis is false:

$$H_1 : \Pr\left(E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right] = 0\right) < 1.$$

How can we distinguish between H_0 and H_1 in practice?

The problem is that both hypotheses involve conditional expectations relative to infinitely many lagged Y_t's, whereas in practice we only observe a finite sample from $\{Y_t\}_{t=-\infty}^{\infty}$.

However, by a well-known martingale convergence result we have

$$\Pr\left(\lim_{m\to\infty} E\left[U_t|\mathcal{F}_{t-m}^{t-1}\right] = E\left[U_t|\mathcal{F}_{-\infty}^{t-1}\right]\right) = 1,$$

where

$$\mathcal{F}_{t-m}^{t-1} = \sigma\left(\{Y_{t-i}\}_{i=1}^{m}\right)$$

is the σ-algebra generated by the finite sequence $Y_{t-1}, Y_{t-2}, ..., Y_{t-m}$.

Consequently, H_1 is equivalent to

$$H_1 : \exists k \in \mathbb{N} : \Pr\left(E\left[U_t|\mathcal{F}_{t-k}^{t-1}\right] = 0\right) < 1.$$

Of course, this k is unknown.

But given such a k,

$$H_1(k) : \Pr\left(E\left[U_t|\mathcal{F}_{t-k}^{t-1}\right] = 0\right) < 1$$

implies, similar to the i.i.d. case, that

$$E\left[U_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m Y_{t-m}\right)\right] \neq 0$$

for some $(\tau_1, \tau_2, ..., \tau_k)' \in \mathbb{R}^k$.

Moreover, given a bounded one-to-one mapping $\Phi : \mathbb{R} \to \mathbb{R}$ with Borel measurable inverse, for example $\Phi(y) = \arctan(y)$, $H_1(k)$ implies that

$$E\left[U_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m})\right)\right] \neq 0 \text{ a.e. on } \mathbb{R}^k.$$

Consequently, for any compact set $\Upsilon \subset \mathbb{R}$ with positive Lebesgue measure, for example, let

$$\Upsilon = [-c, c] \text{ for some constant } c > 0,$$

and with μ the uniform probability measure on Υ, $H_1(k)$ implies that

$$\eta_k = \int_{\Upsilon^k} \left| E\left[U_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m}) \right) \right] \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k) > 0$$

Therefore, given any sequence of positive constants γ_k satisfying $\sum_{k=1}^{\infty} \gamma_k < \infty$, and any subsequence L_n of n such that $L_n = o(n) \to \infty$ as $n \to \infty$, H_1 itself is equivalent to

$$H_1 : \lim \inf_{n \to \infty} \sum_{k=1}^{L_n} \gamma_k \eta_k > 0$$

whereas H_0 is equivalent to

$$H_0 : \sup_{n \in \mathbb{N}} \sum_{k=1}^{L_n} \gamma_k \eta_k = 0.$$

The WICM test

Without loss of generality we may assume that Y_t is observed for $t = 1 - L_n$ to $t = n$.

Then for $k \leq L_n$ the martingale difference null hypothesis can be tested against the specific alternative

$$H_1(k) : \Pr\left(E\left[U_t|\mathcal{F}_{t-k}^{t-1}\right] = 0\right) < 1$$

using the ICM test statistic

$$\widehat{B}_{n,k} = \int_{\Upsilon^k} \left|\widehat{W}_{k,n}(\tau_1, \tau_2, ...\tau_k)\right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k)$$

where

$$\widehat{W}_{k,n}(\tau_1, \tau_2, ...\tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{U}_t \exp\left(\mathbf{i}. \sum_{j=1}^{k} \tau_j \Phi(Y_{t-j})\right),$$

with $\widehat{U}_t$ the OLS residuals of the AR(1) null model.

Similar to the i.i.d. case, under H_0 and for each fixed k,

$$\widehat{W}_{k,n}(\tau_1, \tau_2, ...\tau_k) \Rightarrow W_k(\tau_1, \tau_2, ...\tau_k) \text{ on } \Upsilon^k,$$

where W_k is a zero mean complex valued Gaussian process.

Hence for each $k \in \mathbb{N}$,

$$\widehat{B}_{n,k} \xrightarrow{\text{d}} B_k \text{ under } H_0,$$

where

$$B_k = \int_{\Upsilon^k} |W_k(\tau_1, \tau_2, ...\tau_k)|^2 \, \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k)$$

Moreover, it can be shown that more generally,

$$\widehat{T}_n = \sum_{k=1}^{L_n} \gamma_k \widehat{B}_{n,k} \xrightarrow{\text{d}} T = \sum_{k=1}^{\infty} \gamma_k B_k \text{ under } H_0$$

for γ_k and L_n as before.

Furthermore, it can be shown that under H_1,

$$\widehat{T}_n/n = \sum_{k=1}^{L_n} \gamma_k \widehat{B}_{k,n}/n \xrightarrow{\text{p}} \sum_{k=1}^{\infty} \gamma_k \eta_k > 0$$

where

$$\eta_k = \int_{\Upsilon^k} \left| E\left[U_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m}) \right) \right] \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k).$$

Thus, the WICM test is consistent.

Bootstrap critical values in the AR(1) case

Similar to the i.i.d. case, denote

$$
\begin{aligned}
\widehat{b}_k(\tau_1, \tau_2, ...\tau_k) &= \frac{1}{n}\sum_{t=1}^{n}\left[\begin{pmatrix} 1 \\ Y_{t-1} \end{pmatrix} \exp\left(\mathbf{i}\sum_{j=1}^{k}\tau_j \Phi(Y_{t-j})\right)\right], \\
\widehat{A} &= \frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} 1 & Y_{t-1} \\ Y_{t-1} & Y_{t-1}^2 \end{pmatrix} \\
\widehat{\phi}_{k,t-1}(\tau_1, \tau_2, ...\tau_k) &= \exp\left(\mathbf{i}\sum_{j=1}^{k}\tau_j \Phi(Y_{t-j})\right) \\
&\quad -\widehat{b}_k(\tau_1, \tau_2, ...\tau_k)'\widehat{A}^{-1}\begin{pmatrix} 1 \\ Y_{t-1} \end{pmatrix}, \\
\phi_{k,t-1}(\tau_1, \tau_2, ...\tau_k) &= p\lim_{n\to\infty}\widehat{\phi}_{k,t-1}(\tau_1, \tau_2, ...\tau_k)
\end{aligned}
$$

Then

$$\widehat{W}_{k,n}(\tau_1, \tau_2, ...\tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{U}_t \exp\left(\mathbf{i}. \sum_{j=1}^{k} \tau_j \Phi(Y_{t-j})\right)$$
$$= W_{k,n}(\tau_1, \tau_2, ...\tau_k)$$

where

$$W_{k,n}(\tau_1, \tau_2, ...\tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \widehat{\phi}_{k,t-1}(\tau_1, \tau_2, ...\tau_k)$$

Hence,

$$\widehat{T}_n = \sum_{k=1}^{L_n} \gamma_k \widehat{B}_{n,k} = \sum_{k=1}^{L_n} \gamma_k \int_{\Upsilon^k} |W_{k,n}(\tau_1, \tau_2, ...\tau_k)|^2 \, \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)....\mathrm{d}\mu(\tau_k)$$

In order to generate bootstrap versions of $\widehat{T}_n$, similar to the i.i.d. case, we need to convert $\widehat{T}_n$ to an ICM test statistic in "single integral" form, as follows.

Denote

$$\begin{aligned} W^{+}_{k,n}(\tau_1, \tau_2, ..., \tau_k, \tau_{k+1}) &= \sqrt{\gamma_k} W_{k,n}(\tau_1, \tau_2, ...\tau_k)\rho_{k+1}(\tau_{k+1}) \\ \overline{W}_{n,L}(\tau_1, \tau_2, ..., \tau_L, \tau_{L+1}) &= \sum_{k=1}^{L} W^{+}_{k,n}(\tau_1, \tau_2, ...\tau_k, \tau_{k+1}) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \sum_{k=1}^{L} \sqrt{\gamma_k} \widehat{\phi}_{k,t-1}(\tau_1, ...\tau_k)\rho_{k+1}(\tau_{k+1}) \end{aligned}$$

where the functions $\rho_k(\tau)$ are continuous real functions on Υ such that

$$\int_{\Upsilon} \rho_m(\tau)\rho_k(\tau)\mathrm{d}\mu(\tau) = I(m = k), \ \int_{\Upsilon} \rho_k(\tau)\mathrm{d}\mu(\tau) = 0$$

Then

$$\widehat{T}_n = \int_{\Upsilon^{L_n+1}} \left|\overline{W}_{n,L_n}(\tau_1, \tau_2, ...\tau_{L_n}, \tau_{L_n+1})\right|^2 \mathrm{d}\mu(\tau_1)....\mathrm{d}\mu(\tau_{L_n})\mathrm{d}\mu(\tau_{L_n+1})$$

$$= \sum_{k=1}^{L_n} \gamma_k \int_{\Upsilon^k} |W_{k,n}(\tau_1, \tau_2, ...\tau_k)|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)....\mathrm{d}\mu(\tau_k)$$

Similarly, let

$$\widetilde{W}_{k,n}(\tau_1, \tau_2, ...\tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \widehat{U}_t \widehat{\phi}_{k,t-1}(\tau_1, \tau_2, ...\tau_k),$$

$$\widetilde{W}_{n,L}(\tau_1, ..., \tau_L, \tau_{L+1}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t \widehat{U}_t \sum_{k=1}^{L} \sqrt{\gamma_k} \widehat{\phi}_{k,t-1}(\tau_1, ...\tau_k)\rho_{k+1}(\tau_{k+1})$$

where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$.

Then as before,

$$\widetilde{T}_n = \int_{\Upsilon^{L_n+1}} \left| \widetilde{W}_{n,L_n}(\tau_1, \tau_2, ...\tau_{L_n}, \tau_{L_n+1}) \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)....\mathrm{d}\mu(\tau_{L_n})\mathrm{d}\mu(\tau_{L_n+1})$$

$$= \sum_{k=1}^{L_n} \gamma_k \int_{\Upsilon^k} \left| \widetilde{W}_{k,n}(\tau_1, \tau_2, ...\tau_k) \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)....\mathrm{d}\mu(\tau_k)$$

This motivates the following bootstrap procedure.

Let for $i = 1, 2, ..., M$, with M the bootstrap sample size,

$$\widetilde{W}_{i,k,n}(\tau_1, \tau_2, ...\tau_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{i,t}.\widehat{U}_t.\widehat{\phi}_{k,t-1}(\tau_1, \tau_2, ...\tau_k),$$

$$\widetilde{B}_{i.k,n} = \int_{\Upsilon^k} \left| \widetilde{W}_{i,k,n}(\tau_1, \tau_2, ...\tau_k) \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k),$$

$$\widetilde{T}_{i,n} = \sum_{k=1}^{L_n} \gamma_k \widetilde{B}_{i.k,n}$$

where the $\varepsilon_{i,t}$ are i.i.d. $\mathcal{N}(0, 1)$.

Then it can be shown, similar to the i.i.d. case, that under H_0,

$$(\widehat{T}_n, \widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n})' \xrightarrow{\text{d}} (T, T_1, T_2, ..., T_M)',$$

where $T, T_1, T_2, ..., T_M$ are i.i.d., whereas under H_1,

$$(\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n})' \xrightarrow{\text{d}} (T_1^*, T_2^*, ..., T_M^*)',$$

where $T_1^*, T_2^*, ..., T_M^*$ are i.i.d., but not as T.

The bootstrap critical values of the WICM test can now be based on the quantiles of the empirical distribution function

$$\widetilde{G}_{n,M}(x) = \frac{1}{M}\sum_{i=1}^{M} I\left(\widetilde{T}_{i,n} \leq x\right)$$

of $\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}$.

Standardization of the lagged conditioning variables
Another unresolved issue in Bierens (1984) is how to standardize the conditioning lagged variables of Y_t before taking the bounded transformation Φ in order to preserve enough variation in $\Phi(Y_t)$.

For example, let

$$\Phi(y) = \arctan(y).$$

Suppose that the actual data-generating process is

$$Y_t = 1000 + U_t - 0.5U_{t-1},\ U_t \sim \text{ i.i.d. } N(0,1).$$

Then

$$\Phi(Y_t) \approx \sup_{y\in\mathbb{R}} \arctan(y) = \pi/2,$$

which destroys the power of the WICM test.

Therefore, one should standardize the lagged Y_t before taking the $\arctan(.)$ transformation.

However, in doing this one should also preserve the martingale difference structure under the null hypothesis, as follows.
Given that Y_t is observed for $t = 1 - t_0$ to $t = n$, denote

$$\begin{aligned} \widehat{\mu}_t &= (t+t_0)^{-1} \sum_{i=1-t_0}^{t} Y_i, \\ \widehat{\sigma}_t &= \sqrt{(t+t_0)^{-1} \sum_{i=1-t_0}^{t} (Y_i - \widehat{\mu}_t)^2}, \text{ if } t > 1 - t_0, \\ \widehat{\mu}_t &= 0, \ \widehat{\sigma}_t = 1 \text{ if } t \leq 1 - t_0, \\ \overline{\mu} &= E[Y_t], \ \overline{\sigma} = \sqrt{E[(Y_t - \overline{\mu})^2]} \\ \widetilde{\underline{Y}}_t &= (Y_t - \widehat{\mu}_t)/\widehat{\sigma}_t, \ \underline{Y}_t = (Y_t - \overline{\mu})/\overline{\sigma}. \end{aligned}$$

Replace in the WICM test each $\Phi(Y_{t-j})$ by

$$\Phi(\widetilde{\underline{Y}}_{t-j}) = \arctan(\widetilde{\underline{Y}}_{t-j}).$$

Then the asymptotic results are the same as if each $\Phi(Y_{t-j})$ was replaced by $\Phi(\underline{Y}_{t-j}) = \arctan(\underline{Y}_{t-j})$.

The ICM and WICM tests as discussed so far in this talk are now incorporated in my free econometrics Windows software package *EasyReg International*, which can be downloaded from

http://www.personal/hxb11/EASYREG.HTM


Easy Regression International [User: Herman Bierens]
Registration Help Level File Menu Tours Tools WWW Exit
EasyReg
International
Herman J. Bierens
May 27, 2015
Pennsylvania State University

Choices to make

The WICM test requires to make a number of choices, namely regarding

- the absolutely continuous (with respect to Lebesgue measure) probability measure μ on Υ,
- the compact set Υ itself, and
- the positive sequence $\{\gamma_k\}_{k=1}^{\infty}$.

Under the null hypothesis that the time series regression model is correctly specified these choices do not matter too much. However, they do affect the finite sample power of the test.

The question is:
Can we choose Υ, μ and/or $\{\gamma_k\}_{k=1}^{\infty}$ such that the finite sample power of the WICM is "optimal" in some sense?

The probability measure μ

Boning and Sowell (1999) have shown that with μ the uniform probability measure on Υ the ICM test in Bierens and Ploberger (1997) is optimal in the sense of having the greatest weighted average local power.

Also, with μ the uniform probability measure and Υ a hypercube the WICM test statistic has a closed form expression.

Therefore, it is recommended to choose for μ the uniform probability measure on Υ.

The compact set Υ

In Bierens (1982, 1984) it was recommended to choose Υ around the origin of the Euclidean space involved.

In Bierens (1990) it was shown that the ICM test remains consistent for any compact set Υ with positive Lebesgue measure.

However, for linear and nonlinear regression models with an additive constant term (as is usual the case) it is well-know that the least squares residuals sum up to zero, regardless whether the model is correctly specified or not.

Consequently, in the AR(1) case the empirical processes

$$\widehat{W}_{k,n}(\tau_1, \tau_2, \ldots.\tau_k) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{U}_t \exp\left(\mathbf{i}.\sum_{j=1}^{k}\tau_j\Phi(Y_{t-j})\right),$$

are identically zero in $(\tau_1, \tau_2, \ldots.\tau_k)' = 0$.

Therefore, in general it seems better to choose Υ away from the origin of its Euclidean space.

In the AR(1) case, under H_1, we have:

$$\begin{aligned}
&\widehat{T}_n/n \\
&= \sum_{k=1}^{L_n} \gamma_k \int_{\Upsilon^k} \left| \frac{1}{n} \sum_{t=1}^{n} \widehat{U}_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m}) \right) \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k) \\
&\xrightarrow{\text{p}} \sum_{k=1}^{\infty} \gamma_k \int_{\Upsilon^k} \left| E\left[U_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m}) \right) \right] \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k) \\
&> 0
\end{aligned}$$

The latter expression is a function of Υ.

Can we choose Υ such that this expression is ”maximal”, in some sense?

For example, in the AR(1) case, let for given $c > 0$,

$$\Upsilon = \Upsilon(\xi) = [\xi - c, \xi + c],\ \xi \in \Xi$$

where the set Ξ is compact.

Denote

$$\widehat{T}_n(\xi) = \sum_{k=1}^{L_n} \gamma_k \int_{\Upsilon(\xi)^k} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widehat{U}_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m}) \right) \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k)$$

where μ is now the uniform probability measure on $\Upsilon(\xi)$.

Then under H_0,

$$\sup_{\xi \in \Xi} \widehat{T}_n(\xi) \xrightarrow{\mathrm{d}} \sup_{\xi \in \Xi} T(\xi)$$

where

$$T(\xi) = \sum_{k=1}^{\infty} \gamma_k \int_{\Upsilon(\xi)^k} \left| W_k(\tau_1, \tau_2, ...\tau_k) \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k)$$

Under H_1,

$$\sup_{\xi \in \Xi} \widehat{T}_n(\xi)/n \xrightarrow{\text{p}} \sup_{\xi \in \Xi} \eta(\xi) > 0$$

where

$$\eta(\xi) = \sum_{k=1}^{\infty} \gamma_k \int_{\Upsilon(\xi)^k} \left| E\left[U_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m}) \right) \right] \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k)$$

The bootstrap procedure can be adapted to $\sup_{\xi \in \Xi} \widehat{T}_n(\xi)$.

The sequence of weights

The ideal weight sequence $\{\gamma_k\}_{k=1}^{\infty}$ for the WICM test is such that under H_1, γ_k is maximal when η_k is maximal, where in the AR(1) case,

$$\eta_k = \int_{\Upsilon^k} \left| E\left[U_t \exp\left(\mathbf{i} \sum_{m=1}^{k} \tau_m \Phi(Y_{t-m}) \right) \right] \right|^2 \mathrm{d}\mu(\tau_1)\mathrm{d}\mu(\tau_2)...\mathrm{d}\mu(\tau_k)$$

But we don't know the η_k's.

However, what we can do is to make the γ_k's dependent on parameters.

For example, choose for γ_k the probability of the Poisson(ω) distribution for $k-1$, i.e.,

$$\gamma_k(\omega) = \exp(-\omega)\frac{\omega^{k-1}}{(k-1)!},\ k \in \mathbb{N},$$

with ω confined to a compact set Ω in $(0, \infty)$.

Then the WICM test statistic takes the form

$$\widehat{T}_n(\omega) = \sum_{k=1}^{L_n} \gamma_k(\omega)\widehat{B}_{n,k}.$$

It can be shown that

$$\sup_{\omega\in\Omega} \widehat{T}_n(\omega) \xrightarrow{\text{d}} \sup_{\omega\in\Omega} \sum_{k=1}^{\infty} \gamma_k(\omega) B_k \text{ under } H_0,$$

$$\sup_{\omega\in\Omega} \widehat{T}_n(\omega)/n \xrightarrow{\text{p}} \sup_{\omega\in\Omega} \sum_{k=1}^{\infty} \gamma_k(\omega)\eta_k > 0 \text{ under } H_1.$$

The bootstrap procedure can also be adapted to $\sup_{\omega\in\Omega} \widehat{T}_n(\omega)$.

The ICM test for conditional distributions: The i.i.d. case

A wide range of parametric econometric cross-section models take the form of a conditional distribution specifications

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

where X is a vector of stochastic covariates, Y is a multivariate or univariate dependent variable, and θ_0 is the vector of "true" parameters, to be estimated by maximum likelihood.

For example, if Y represents count data, with $\Pr[Y = y] > 0$ for all $y \in \{0\} \cup \mathbb{N}$, a convenient (and therefore popular) specification of its conditional distribution is the conditional Poisson model

$$\begin{aligned} \Pr[Y = y|X] &= \exp(-\exp((1, X')\theta_0)) \,.\, \exp(y.(1, X')\theta_0)/y!, \\ y &= 0, 1, 2, ... \end{aligned}$$

The literature on consistent testing of the validity of these kind of conditional distribution models is very limited, as it consists only of three papers:

Andrews, D.W., 1997, A conditional Kolmogorov test. *Econometrica* 65, 1097-1128.

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

Bierens, H. J., and Wang, L., 2012, Integrated conditional moment tests for parametric conditional distributions. *Econometric Theory* 28, 328-362.

Andrews' (1997) conditional Kolmogorov (CK) test statistic takes the form

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

where $F(y|X,\widehat{\theta})$ is the estimated conditional distribution model and $I(.)$ is the indicator function.

A problem with the CK test is that if the dimension of X is large then the events $X_j \leq X_i$ may never happen, even in large samples, so that then the CK test statistic becomes zero.

Zheng's (2000) test is based on comparison of estimated parametric conditional densities with their corresponding kernel density estimates, and is therefore confined to absolutely continuous distribution specifications only.

The Bierens-Wang (2012) paper avoids these limitations by comparing an estimate of the joint characteristic function of (Y, X) implied by the estimated model $F(y|X; \widehat{\theta})$ with the corresponding empirical characteristic function.

In this talk I will focus on the case $Y \in \mathbb{R}$, but the approach in Bierens-Wang (2012) carries over to the case $Y \in \mathbb{R}^m$ as well.

The conditional distribution function of Y given X is assumed to belong to a given parametric family $F(y|X;\theta)$, $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^p$ is a given compact and convex parameter space.

The support of $F(y|X;\theta)$ does not depend on θ.

The null hypothesis to be tested is that

H_0 : There exists a $\theta_0 \in \Theta$ such that

$$\Pr\left[\sup_y |\Pr[Y \le y|X] - F(y|X;\theta_0)| = 0\right] = 1$$

against the alternative hypothesis that H_0 is false, i.e.,

H_1 : For all $\theta \in \Theta$,

$$\Pr\left[\sup_y |\Pr[Y \le y|X] - F(y|X;\theta)| = 0\right] < 1$$

Throughout it will be assumed that the parameter vector θ_0 under H_0 is estimated consistently by maximum likelihood (ML), with ML estimator $\widehat{\theta}_n$, on the bases of a random sample $\{(Y_j, X_j)\}_{j=1}^n$ from $(Y, X) \in \mathbb{R} \times \mathbb{R}^k$.

Under H_1 the estimator $\widehat{\theta}_n$ becomes a Quasi-ML (QML) estimator, which will also converge in probability to a point in Θ, and without loss of generality we may denote this point by θ_0 as well.

Thus, in either case, $p\lim_{n\to\infty} \widehat{\theta}_n = \theta_0$.

Hence, H_0 and H_1 now read

$$H_0 : \Pr\left[\sup_y |\Pr[Y \leq y|X] - F(y|X;\theta_0)| = 0\right] = 1,$$

$$H_1 : \Pr\left[\sup_y |\Pr[Y \leq y|X] - F(y|X;\theta_0)| = 0\right] < 1,$$

respectively.

Since conditional distributions are equal if and only if their conditional characteristic functions are equal, these hypotheses are equivalent to

$$H_0 : \Pr\left[\sup_{\tau\in\mathbb{R}}\left|E[\exp(\mathbf{i}.\tau.Y)|X] - \int \exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X;\theta_0)\right| = 0\right] = 1,$$

$$H_1 : \Pr\left[\sup_{\tau\in\mathbb{R}}\left|E[\exp(\mathbf{i}.\tau.Y)|X] - \int \exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X;\theta_0)\right| = 0\right] < 1,$$

respectively, where $\mathbf{i} = \sqrt{-1}$.

Similar to the ICM test for regression model these hypotheses are equivalent to

$$H_0 \;:\; E\left[\exp(\mathbf{i}.\tau.Y)\exp(\mathbf{i}.\xi'X)\right] = E\left[\int \exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X;\theta_0)\exp(\mathbf{i}.\xi'X)\right]$$

$$\text{for all } (\tau,\xi) \in \mathbb{R}\times\mathbb{R}^k,$$

$$H_1 \;:\; E\left[\exp(\mathbf{i}.\tau.Y)\exp(\mathbf{i}.\xi'X)\right] \neq E\left[\int \exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X;\theta_0)\exp(\mathbf{i}.\xi'X)\right]$$

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

Moreover, if Y and X are bounded then under H_1 the set

$$\begin{aligned} S &= \Big\{(\tau, \xi) \in \mathbb{R} \times \mathbb{R}^k : E\left[\exp(\mathbf{i}.\tau.Y)\exp(\mathbf{i}.\xi' X)\right] \\ &\quad = E\left[\int \exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X;\theta_0)\exp(\mathbf{i}.\xi' X)\right]\Big\} \end{aligned}$$

has zero Lebesgue measure and is nowhere dense, whereas under H_0,

$$S = \mathbb{R} \times \mathbb{R}^k.$$

If Y and/or X are not bounded then we may replace Y, y and X in the complex $\exp(.)$ functions by $\Psi(Y)$, $\Psi(y)$ and $\Phi(X)$, respectively, where $\Psi : \mathbb{R} \to \mathbb{R}$ and $\Phi : \mathbb{R}^k \to \mathbb{R}^k$ are bounded one-to-one mappings with Borel measurable inverses, so that S becomes

$$S = \Big\{(\tau, \xi) \in \mathbb{R} \times \mathbb{R}^k : E\left[\exp(\mathbf{i}.\tau.\Psi(Y))\exp(\mathbf{i}.\xi'\Phi(X))\right] \\ = E\left[\int \exp(\mathbf{i}.\tau.\Psi(y))\mathrm{d}F(y|X;\theta_0)\exp(\mathbf{i}.\xi'\Phi(X))\right]\Big\}$$

Again, under H_1 this set S has zero Lebesgue measure and is nowhere dense, whereas under H_0,

$$S = \mathbb{R} \times \mathbb{R}^k.$$

However, for the time being let us assume that Y and X are bounded, and that the conditional characteristic function

$$\varphi(\tau|X;\theta) = \int \exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X,\theta)$$

has a continuous closed form expression in τ and θ.

The result that under H_1 the set

$$\begin{aligned} S \;=\; \Big\{(\tau,\xi) \in \mathbb{R}\times\mathbb{R}^k : E\left[\exp(\mathbf{i}.\tau.Y)\exp(\mathbf{i}.\xi'X)\right] \\ = E\left[\varphi\left(\tau|X;\theta_0\right)\exp(\mathbf{i}.\xi'X)\right]\Big\} \end{aligned}$$

has zero Lebesgue measure and is nowhere dense now suggests that the validity of the null hypothesis can be consistently tested by an ICM test of the form

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

where

- $Z_n(\tau,\xi) = \frac{1}{\sqrt{n}}\sum_{j=1}^n(\exp(\mathbf{i}.\tau.Y_j) - \varphi(\tau|X_j;\widehat{\theta}_n))\exp(\mathbf{i}.\xi'X_j)$.
- Υ and Ξ are compact subsets of $\mathbb{R}$ and $\mathbb{R}^k$, respectively, with positive Lebesgue measure, and
- $\mu(\tau,\xi)$ is the uniform probability measure on $\Upsilon\times\Xi$.

It follows straightforwardly that under H_1,

$$\begin{aligned} Z_n(\tau,\xi)/\sqrt{n} &\xrightarrow{\text{p}} E\left[(\exp(\mathbf{i}.\tau.Y) - \varphi\left(\tau|X;\theta_0\right))\exp(\mathbf{i}.\xi'X)\right] \\ &= \varsigma(\tau,\xi), \text{ say,} \end{aligned}$$

uniformly on $\Upsilon\times\Xi$, where $\varsigma(\tau,\xi)\neq 0$ on $(\Upsilon\times\Xi)\backslash S$, hence

$$\widehat{T}_n/n = \int_{\Upsilon\times\Xi}|Z_n(\tau,\xi)/\sqrt{n}|^2\mathrm{d}\mu(\tau,\xi) \xrightarrow{\text{p}} \int_{\Upsilon\times\Xi}|\varsigma(\tau,\xi)|^2\mathrm{d}\mu(\tau,\xi) > 0$$

In order to derive the null distribution of the ICM statistic

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

write $Z_n(\tau,\xi)$ as

$$\begin{aligned} Z_n(\tau,\xi) \;&=\; \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(\exp(\mathbf{i}.\tau'Y_j) - \varphi(\tau|X_j;\theta_0))\exp(\mathbf{i}.\xi'X_j) \\ &\quad -\frac{1}{n}\sum_{j=1}^{n}\sqrt{n}\left(\varphi(\tau|X_j;\widehat{\theta}_n) - \varphi(\tau|X_j;\theta_0)\right)\exp(\mathbf{i}.\xi'X_j) \end{aligned}$$

Similar to the regression case we can write

$$\frac{1}{n}\sum_{j=1}^{n}\sqrt{n}\left(\varphi(\tau|X_j;\widehat{\theta}_n)-\varphi(\tau|X_j;\theta_0)\right)\exp(\mathbf{i}.\xi'X_j)$$

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

under H_0 and standard ML conditions, where

- $\ell\left(Y,X;\theta\right)$ is the log-likelihood function,
- $\Delta\ell\left(Y,X;\theta_0\right)=\partial\ell\left(Y,X;\theta\right)/\partial\theta'|_{\theta=\theta_0}$ is the score vector,
- $A=\mathrm{Var}(\Delta\ell\left(Y,X;\theta_0\right))$,
- $b(\tau,\xi)=E[\Delta\varphi\left(\tau|X;\theta_0\right)\exp(\mathbf{i}.\xi'X)]$,
- $\Delta\varphi\left(\tau|X;\theta\right)=\partial\varphi\left(\tau|X;\theta\right)/\partial\theta'|_{\theta=\theta_0}$, with
- $\varphi\left(\tau|X;\theta\right)=\int\exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X,\theta)$, and
- the $o_p(1)$ term is uniform on $\Upsilon\times\Xi$.

Thus, denoting

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

where

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

it follows that under H_0

$$\widehat{T}_n = \int_{\Upsilon\times\Xi} |\widetilde{Z}_n(\tau,\xi)|^2 \mathrm{d}\mu(\tau,\xi) + o_p(1)$$

Thus, denoting

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

where

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

it follows that under H_0

$$\widehat{T}_n = \int_{\Upsilon\times\Xi} |\widetilde{Z}_n(\tau,\xi)|^2 \mathrm{d}\mu(\tau,\xi) + o_p(1)$$

Moreover, it can be shown that under H_0, $\widetilde{Z}_n(\tau,\xi)$ converges weakly to a zero mean complex valued Gaussian process $Z(\tau,\xi)$ on $\Upsilon\times\Xi$, so that

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

Bootstrap procedure

The bootstrap procedure in the present case is quite different than for regression models.

The current bootstrap procedure is an adaptation of the approach in:

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

as follows.

Given the bootstrap sample size M, and for each $m = 1, 2, ..., M$,

- Generate random drawings $Y_{m,j}$ from the estimated conditional model $F(y|X_j; \widehat{\theta}_n)$, $j = 1, 2, ..., n$, given the actual X_j's in the sample.
- Compute the ML estimator $\widetilde{\theta}_{m,n}$ on the basis of the bootstrap sample $\{(Y_{m,j}, X_j)\}_{j=1}^n$.
- Compute the corresponding ICM test statistic $\widetilde{T}_{m,n}$.

Then under H_0,

$$\left(\widehat{T}_n, \widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}\right)' \xrightarrow{\text{d}} (T, T_1, T_2, ..., T_M)',$$

where $T, T_1, T_2, ..., T_M$ are i.i.d., whereas under H_1,

$$\left(\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}\right)' \xrightarrow{\text{d}} (T_1^*, T_2^*, ..., T_M^*)'$$

where $T_1^*, T_2^*, ..., T_M^*$ are i.i.d. (but not as T)
As before, bootstrap critical values can now be based on the quantiles of the empirical distribution function of $\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}$.

The simulated ICM test

The theoretical conditional characteristic function

$$\varphi\left(\tau|X;\theta\right)=\int\exp(\mathbf{i}.\tau.y)\mathrm{d}F(y|X,\theta),$$

poses computational challenges in various 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.

Moreover, the need for numerical integration will slow down the bootstrap too much.

To cope with these problems, 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

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

if Y and X are bounded, or

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

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 $\Psi(.)$ and $\Phi(.)$ are bounded one-to-one mappings.

The SICM test statistic is then

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

A practical advantage of the SICM test that $\widehat{T}_n^{(s)}$ has a closed-form expression if Υ is an interval and Ξ is a hyper-cube.

All the previous results for the exact ICM test carry over to the SICM test, including the bootstrap, albeit with a different null distribution.

As to the latter, and assuming that Y and X are bounded, we can write

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

$$Z_n(\tau,\xi) = \tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^{n}\left(\exp(\mathbf{i}.\tau.Y_j) - \int\exp(\mathbf{i}.\tau' y)\mathrm{d}F(y|X_j,\hat{\theta})\right)\exp(\mathbf{i}.\xi' X_j),$$

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

Under H_0 the empirical process $Z_n(\tau, \xi)$ converges weakly to a zero-mean complex valued Gaussian process $Z(\tau, \xi)$ on $\Upsilon \times \Xi$, and the empirical process $\widehat{Z}_n^{(s)}(\tau, \xi)$ converges weakly to a zero-mean complex valued Gaussian process $Z^{(s)}(\tau, \xi)$ on $\Upsilon \times \Xi$, where $Z(\tau, \xi)$ and $Z^{(s)}(\tau, \xi)$ are independent.

Consequently, under H_0,

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

whereas in the case of the exact ICM test,

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

The previous bootstrap procedure can easily adapted to the SICM test.

The WICM test for conditional distributions: The stationary time series case

In

Bierens, H. J., and Wang, L., 2017, Weighted Simulated Integrated Conditional Moment Tests for Parametric Conditional Distributions of Stationary Time Series Processes, *Econometric Reviews* 36, 103-135,

we propose a consistent weighted simulated integrated conditional moment (WSICM) test of the validity of a parametric conditional distribution specification for time series data, by combining the WICM test for time series regression with the SICM test for conditional distributions in the i.i.d. case.

For example, let Y_t be a strictly and covariance stationary univariate time series process. Without loss of generality we may assume that Y_t is bounded, as otherwise we may replace Y_t by $\Psi(Y_t)$ with Ψ a bounded one-to-one transformation.

Let $F_{t-1}(y|\theta_0)$ be a parametric specification of the conditional distribution function

$$\Pr\left[Y_t \leq y|\mathcal{F}_{-\infty}^{t-1}\right]$$

where $\mathcal{F}_{-\infty}^{t-1}$ is the σ-algebra generated by the sequence $\{Y_{t-i}\}_{i=1}^{\infty}$.

In particular, suppose that $F_{t-1}(y|\theta_0)$ depends on a finite number of lagged Y_t's, say $X_{t-1,k} = (Y_{t-1}, Y_{t-2,....,}Y_{t-k})'$ so that we can write

$$F_{t-1}(y|\theta_0) = F(y|X_{t-1,k};\theta_0)$$

Moreover, suppose that Y_t is observed for $t = 1 - L_n$ to $t = n$, where $L_n = o(n) \to \infty$ as for the WICM test, with n so large that $L_n \geq k$

Furthermore, suppose that θ_0 is estimated by the ML or QML estimator $\widehat{\theta}_n$, where in both cases, $\theta_0 = p\lim_{n\to\infty} \widehat{\theta}_n$.

Then the null hypothesis to be tested is that

$$H_0 : \Pr\left[\sup_y \left|\Pr\left[Y_t \leq y|\mathcal{F}_{-\infty}^{t-1}\right] - F(y|X_{t-1,k};\theta_0)\right| = 0\right] = 1$$

against the alternative that H_0 is false.

The WSICM test

For each $t = 1, 2, ..., n$, draw randomly an $\widetilde{Y}_t$ from $F(y|X_{t-1,k}; \widehat{\theta}_n)$, given $X_{t-1,k}$.

Denote for $m = 1, 2, ..., L_n$,

$$Z_{n,m}(\tau, \xi_1, \xi_2, ..., \xi_m) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\exp(\mathbf{i}.\tau.Y_t) - \exp(\mathbf{i}.\tau.\widetilde{Y}_t) \right) \times \exp\left(\mathbf{i}. \sum_{j=1}^{m} \xi_j Y_{t-j} \right)$$

$$\widetilde{B}_{n,m} = \int_{\Upsilon \times \Xi^m} |Z_{n,m}(\tau, \xi_1, \xi_2, ..., \xi_m)|^2 \, \mathrm{d}\mu_\Upsilon(\tau) \mathrm{d}\mu_\Xi(\xi_1) ...., \mathrm{d}\mu_\Xi(\xi_m),$$

where Υ and Ξ are compact sets in $\mathbb{R}$ with positive Lebesgue measure, and μ_Υ and μ_Ξ are uniform probability measures on Υ and Ξ, respectively.

Then similar to the WICM test for time series regressions, the test statistic of the WSICM test takes the form

$$\widetilde{T}_n = \sum_{m=1}^{L_n} \gamma_m \widetilde{B}_{n,m}$$

where the γ_m's are positive and satisfy $\sum_{m=1}^{\infty} \gamma_m < \infty$.

The asymptotic properties of $\widetilde{T}_n$ are similar to the SICM test in the i.i.d. case, and so is the bootstrap procedure involved.

THANK YOU!