A Note on Consistent Conditional Moment Tests

Wang, Xuexin

he Wang Yanan Institute for Studies in Economics(WISE), Xiamen University, Xiamen, Fujian, China

1 July 2015

A NOTE ON CONSISTENT CONDITIONAL MOMENT TESTS

Xuexin WANG*

This Version: July 2015

Abstract

In this paper we propose a new consistent conditional moment test, which synergizes Bierens' approach with the consistent test of overidentifying restrictions. It relies on a transformation-based empirical process combining both approaches. This new empirical process enjoys some advantages. Firstly it is not affected by the uncertainty from the parameter estimation. Moreover this estimation-effect-free property requires much less restrictive rate condition than in the consistent test of overidentifying restrictions alone. Furthermore the integrated conditional moment (ICM) test based on the new empirical process have power against Pitman local alternatives. We prove, under some regularity conditions, the admissibility of the ICM test based on this transformation-based empirical process in the case that there exists heteroskedasticity of unknown form, extending the result in Bierens and Ploberger (1997). The new consistent test also allows us to propose a much simpler bootstrap procedure than the standard ones. A version of Bierens (1990) test based on the new empirical process is also discussed, and its asymptotic properties are analyzed. Monte Carlo simulations show that Bierens (1990) test based on the new empirical process is more powerful for a large number of alternatives when heteroskedasticity of unknown form is presented.

JEL Classification: C12 C21

Keywords: Consistent Conditional Moment Test; Consistent Test of Overidentifying Restrictions ; ICM Test; Admissibility

*Address: The Wang Yanan Institute for Studies in Economics(WISE), Xiamen University, Xiamen, Fujian, China. E-mail: *xuexinwang@xmu.edu.cn*. The financial support from National Natural Science Funds of China (Grant No.71301136) is gratefully acknowledged.

1 Introduction

Models based on conditional moment restrictions are quite prevalent in econometrics, arising in many econometric settings. For example the rational expectations and dynamic asset pricing models in macroeconomics and finance give rise to conditional moment restrictions in the form of stochastic Euler equations. Other cases include panel data models and instrumental variable regressions. In order to conduct convincing estimation and inference of these models, it is crucial to check the validity of these conditional moment restrictions.

There is a vast amount of literature on consistently testing the correct specification of conditional moment restrictions. Generally, these tests can be grouped into two classes. The first class is based on smoothing methods, comparing the fitted parametric regression function with a nonparametric function estimator, see, Härdle and Mammen (1993), Gozalo (1993), Hong and White (1995), Fan and Li (1996), and Zheng (1996), to mention but a few. The smoothing-based tests only require some consistent parameter estimator, and typically lead to asymptotic pivotal test statistics under the null. However, they depend on a smoothing parameter, have no power in Pitman local alternative generally, and there has been much concern over their small sample properties. The second class of tests avoid smoothing estimation by means of converting the conditional moment restriction into an infinite number of unconditional moment restrictions, see, for example, Bierens (1982, 1990), de Jong and Bierens (1994), Bierens and Ploberger (1997), Stute (1997), Donald, Imbens and Newey (2003) and Escanciano (2006a). One strand of literature in this class, which we would like to label as Bierens' approach, considers a continuum of unconditional moment restrictions implied by the conditional moment restrictions. It relies on functionals of an empirical process, has to handle the uncertainty from the parameters estimation, and leads to case-dependent limiting distributions. But it has power against Pitman local alternatives. On the other hand, de Jong and Bierens (1994) and Donald et al. (2003) propose consistent tests of overidentifying restrictions with the discrete number of unconditional moment restrictions increasing with the sample size. Under some regularity conditions, the normalized consistent tests of overidentifying restrictions are not affected by the estimation effects. However they have no power against Pitman local alternatives, and they depend on a nuisance parameter similar to the smoothing parameter in smoothing methods. Finally, Carrasco and Florens (2000) and Dominguez and Lobato (2015) propose consistent tests of overidentifying restrictions with a continuum of unconditional moment restrictions, which can be regarded as a synergy of consistent tests of Bierens and the consistent tests of overidentifying restrictions.

In this paper we propose a new consistent conditional moment test, which synergizes Bierens' approach with the consistent test of overidentifying restrictions. It relies on a transformation-based empirical process combining both approaches. This new empirical process enjoys some advantages. Firstly it is not affected by the uncertainty from the parameter estimation. Moreover this estimation-effect-free property requires much less restrictive rate condition than in the consistent test of overidentifying restrictions alone. Furthermore the integrated conditional moment (ICM) test based on the new empirical process have power against Pitman local alternatives. We prove, under some regularity conditions, the admissibility of the ICM test based on this transformation-based empirical process in the case that there exists heteroskedasticity of unknown form, extending the result in Bierens and Ploberger (1997). The new consistent test also allows us to propose a much simpler bootstrap procedure than the standard ones. A version of Bierens (1990) test based on the new empirical process is also discussed, and its asymptotic properties are analyzed. Monte Carlo simulations show that Bierens (1990) test based on the new empirical process is more powerful for a large number of alternatives when heteroskedasticity of unknown form is presented.

The outline of the paper is as follows. In Section 2, we establish the preliminaries. In Section 3, we define the class of weighting functions, the new residual empirical process and study its properties. Section 4 discusses the asymptotic theory of the ICM test. Section 5 establishes the asymptotic admissibility of the ICM test when there exists heteroskedasticity of unknown form. Section 6 propose a simple bootstrap for the ICM test. Section 7 discusses Bierens (1990)'s test based on the new empirical process. Section 8 conducts Monte Carlo simulations. Section 9 concludes.

2 Preliminaries

let Z denote a single observation, θ a $p \times 1$ vector of parameters, and X is a $d \times 1$ subvector of Z. For a unique value $\theta_0 \in \Theta \in \mathbb{R}^p$, the following conditional moment restrictions hold

$$E\left[\rho(Z, \theta_0)|X\right] = 0 \text{ a.s.} \tag{1}$$

where $\rho(Z, \theta_0)$ is a $J \times 1$ vector of functions. It often can be thought of as residuals. In this paper, for simplicity, we will consider the case that $J = 1$.

Example 1. *Let* $Z = (Y, X')'$ *be a random vector in a* $(1+d)$*-dimensional Euclidean space, where* X *is a* $d \times 1$ *vector and* Y *is a scalar. When* $E(|Y|) < \infty$*, there exists a Borel measurable function* f *such that* $E(Y|X) = f(X)$*. In parametric modeling,* $f(X)$ *is assumed to belong to a parametric family* $\mathcal{G} = \left\{ f(X, \theta) : \mathbb{R}^d \to \mathbb{R} | \theta \in \Theta \subset \mathbb{R}^p \right\}$*. In this case* $\rho(Z, \theta_0) = Y - f(X, \theta_0)$*.*

In order to test whether model (1) is correctly specified, we need to test the following null hypothesis

$$H_0 : E\left[\rho(Z, \theta_0)|X\right] = 0 \text{ a.s., for } \theta_0 \in \Theta, \tag{2}$$

and the alternative hypothesis is

$$H_1 : P\left(E\left[\rho(Z, \theta)|X\right] = 0\right) < 1 \text{ a.s., for all } \theta \in \Theta. \tag{3}$$

The idea of Bierens' approach is to convert the conditional moment restriction into an infinite number of unconditional moment restrictions, i.e,

$$E\left[\rho(Z, \theta_0)|X\right] = 0 \text{ a.s} \Leftrightarrow E\left[\rho(Z, \theta_0) w(X, t)\right] = 0, \text{ for almost all } t \in \mathcal{T}, \tag{4}$$

where $\mathcal{T} \subset \mathbb{R}^h$, $h \in \mathbb{N}$, and $w(X, t)$ is a proper weighting function such that the equivalence (4) holds. There are many weighting functions meeting the requirement of (4). One example is $w(X, t) = \exp(it'X)$ where $i = \sqrt{-1}$, $\mathcal{T} = \mathbb{R}^d$, which is employed by Bierens (1982). Bierens (1990) proposes $w(X, t) = \exp(t'X)$, $\mathcal{T} = \mathbb{R}^d$. Stute (1997) proposes the indicator function $w(X, t) = I(X < t)$, $\mathcal{T} = \mathbb{R}^d$. Escanciano (2006a) introduces the weighting function $w(X, t) = I(\beta' X \leq u)$, with $t = (\beta', u)' \in \mathcal{T} = S^d \times (-\infty, \infty)$, where $S^d = \left\{ \beta \in \mathbb{R}^d : |\beta| = 1 \right\}$.

Given a sample Z_j, $j = 1, \cdots, n$, and a $\sqrt{n}$-consistent estimator $\hat{\theta}$, the scaled sample analog of $E\left[\rho(Z, \theta_0) w(X, t)\right]$, which forms a residual empirical process, is

$$\hat{M}\left(\hat{\theta}, t\right) = n^{-1/2} \sum_{j=1}^{n} \rho(Z_j, \hat{\theta}) w(X_j, t), \ t \in \Pi \subset \mathcal{T}.$$

Stinchcombe and White (1998) coin this class of specification tests as the one with a nuisance parameter present only under the alternative, given the presence of t in $\hat{M}\left(\hat{\theta}, t\right)$. Bierens (1982) proposes to integrate t out. The so-called integrated conditional moment (ICM) test statistic has the form

$$\widehat{ICM} = \int_{\Pi} \left| \hat{M}\left(\hat{\theta}, t\right) \right|^2 d\mu\left(t\right),$$

where $\mu(t)$ is a probability measure on Π that is absolutely continuous with respect to Lebesgue measure on $\Pi \subset \mathcal{T}$. Or we can maximize $\hat{M}\left(\hat{\theta}, t\right)$ over Π, resulting in a Kolmogorov-Smirnov type statistic

$$\widehat{KS} = \sup_{t \in \Pi} \left| \hat{M}\left(\hat{\theta}, t\right) \right|^2.$$

In contrast with smoothing-based tests, the bierens' approach has to handle the uncertainty from the parameter estimation. More specifically, it could be shown, under some regularity conditions, that

$$\hat{M}\left(\hat{\theta}, t\right) = n^{-1/2} \sum_{j=1}^{n} \rho(Z_j, \theta_0) w\left(X_j, t\right) + \hat{b}(\theta_0, t) n^{1/2} \left(\hat{\theta} - \theta_0\right) + o_p\left(1\right),$$

where

$$\hat{b}(\theta, t) = \frac{1}{n} \sum_{j=1}^{n} w\left(X_j, t\right) \frac{\partial \rho(Z_j, \theta)}{\partial \theta'}.$$

In most cases, $\hat{b}(\theta_0, t) \neq 0$. The existence of the so-called "estimation effect" $\hat{b}(\theta_0, t) n^{1/2} \left(\hat{\theta} - \theta_0\right)$ makes the asymptotic theory more complicated, and limit results dependent on the parameter estimator employed.

While Bierens' approach exploits a continuum of the unconditional moment restrictions, de Jong and Bierens (1994) and Donald et al. (2003) show that it is sufficient to employ a set of discrete countable unconditional moment restrictions in efficient estimation of parameters and consistent test in the conditional moment restrictions models.[1] More specifically, let U be the support of distribution of X, define L_2 be the space of measurable functions $\varphi : U \to \mathbb{R}$ with $E[\varphi^2(X)] < \infty$. We say a sequence of $\{q_j(X)\}_{j=1}^{\infty}$ in L_2 is L_2-complete if for any $\varepsilon > 0$, and any $\varphi \in L_2$, there exists a positive integer K and a $K \times 1$ vector γ_K such that

$$\left\{ E\left[\left\{ \varphi(X) - q^K\left(X\right)' \gamma_K \right\}^2 \right] \right\}^{1/2} < \varepsilon, \tag{5}$$

where $q^K\left(X\right) = (q_1(X), \cdots, q_K(X))'$ is a $K \times 1$ vector. Chamberlain (1987, 1992) show that

[1] Carrasco and Florens (2000) consider the continuum of unconditional moment restrictions in efficient estimation of the conditional moment restrictions models, however the singularity of the covariance matrix has to be handled. Furthermore, the indexed parameter t has to be a scalar.

the asymptotic variance of the GMM estimator based on the unconditional moment restrictions $E\left[q^K(X)\rho(Z,\theta_0)\right]=0$, where $q^K(X)$ satisfies (5), comes arbitrarily close to the semiparametric efficiency bound as $K\to\infty$. Intuitively, since the conditional moment restriction is equivalent to a sequence of unconditional moment restrictions, as K grows with the sample size, all of the information of the conditional moment restriction is eventually accounted for. The GMM estimator is obtained by minimizing the following objective function

$$\hat{\theta}_{GMM}=\arg\min_{\theta\in\Theta}\left(\sum_{j=1}^{n}q^K(X_j)\rho(Z_j,\theta)\right)'\hat{\Omega}\left(\bar{\theta},K\right)^{-1}\left(\sum_{j=1}^{n}q^K(X_j)\rho(Z_j,\theta)\right),$$

where

$$\hat{\Omega}(\theta,K)=\frac{1}{n}\sum_{i=1}^{n}\rho(Z_i,\theta)^2q^K(X_i)\,q^K(X_i)',$$

and $\bar{\theta}$ is some preliminary estimator of θ_0. Hahn (1997) and Donald et al. (2003) establish the rate condition of K for different choices of $q^K(X)$, for example splines and power series, in efficient estimation of θ_0, reaching semiparametric efficient bound.

When it comes to consistent tests, the test statistic is in the form

$$\hat{T}_{\theta}=n^{-1}\left(\sum_{j=1}^{n}q^K(X_j)\rho(Z_j,\theta)\right)'\hat{\Omega}(\theta,K)^{-1}\left(\sum_{j=1}^{n}q^K(X_j)\rho(Z_j,\theta)\right).$$

For a fixed K, $\hat{T}_{\hat{\theta}_{GMM}}$ is known to be chi-squared distribution with $K-p$ degrees of freedom asymptotically under the null hypothesis that the conditional moment restrictions are satisfied. As K grows with the sample size this approximation continues to hold. Normally the asymptotic normal approximation to the chi-square for large degrees of freedom is used. In this case, under some rate condition of K, a consistent estimator $\hat{\theta}$, it has been shown that

$$\frac{\hat{T}_{\hat{\theta}}-(K-p)}{\sqrt{2(K-p)}}\xrightarrow{d}N(0,1).$$

It should be noticed that the rate condition of K for consistent test is slower that the one for efficient estimation, see Donald et al. (2003). Furthermore it has no power against Pitman local alternatives.

3 A Class of Weighting Functions and A New Empirical Process

In this section, we will synergize the Bierens' approach with the consistent test of overidentifying restrictions in a class of weighting functions. We focus on a class of weighting functions $\mathcal{W}$ such that

$$\mathcal{W} = \left\{ w\left(t'X\right), t \in \mathbb{R}^d, w \text{ is an analytic function that is nonpolynomial} \right\}.$$

For any weighting function in this class, we have the following lemma.

Lemma 1. *Let X be a random vector in $\mathbb{R}^d$, $\Phi(\cdot)$ a bounded one-to-one mapping from $\mathbb{R}^d$ into $\mathbb{R}^d$, for any weighting function $w(t'\Phi(X))$ in $\mathcal{W}$, the equivalence in (4) holds.*

Proof. See Stinchcombe and White (1998) Theorem 2.3. □

Remark: Bierens and Ploberger (1997) give an alternative version of conditions of the equivalence.

Examples of families satisfying this lemma are $w\left(t'\Phi(X)\right) = \exp\left(it'\Phi(X)\right)$ and $w\left(t'\Phi(X)\right) = \exp\left(t'\Phi(X)\right)$.

For $w \in \mathcal{W}$ and any $t \in \Pi \subset \mathbb{R}^d$, we have a residual empirical process such that

$$\hat{M}\left(\hat{\theta}, t\right) = n^{-1/2} \sum_{j=1}^{n} \rho(Z_j, \hat{\theta}) w\left(t'\Phi(X_j)\right).$$

On the other hand, $w\left(t'\Phi(X)\right)$ also forms a basis for the efficient parameters estimation and consistent test of overidentifying restrictions. For any fixed sequence $\{t_j\}_{j=1}^{\infty}$, which is dense in some subset of $\mathbb{R}^d$, $q_j(X) = w(t_j'\Phi(X))$, $j = 1, 2, \cdots$, and for each positive integer K, define the $K \times 1$ vector

$$q^K(X) = \left(w\left(t_1'\Phi(X)\right), \cdots w\left(t_K'\Phi(X)\right)\right)'. \tag{6}$$

Note that we omit in the notation $q^K(X)$ the dependence on $\{t_j\}_{j=1}^{K}$ sequence. We have the following corollary:

Corollary 1. *For any $\varepsilon > 0$, $\varphi \in L_2$, and for each $K \times 1$ vector $q^K(X)$ defined in (6), there exist $K \times 1$ vectors γ_K such that (5) holds.*

Proof. See Appendix. □

Now we present the assumptions:

Assumption 1. *The data are i.i.d. The parameter space* Θ *is a compact subset of* $\mathbb{R}^p$. $\theta_0 \in int(\Theta)$.

Assumption 2. $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$.

Assumption 3. $E\left[\sup_{\theta\in\Theta} \rho(Z,\theta)^2|X\right] < \infty$, $\sigma^2(\theta, X) = E\left[\rho(Z,\theta)^2|X\right]$ *is bounded away from zero. There is* $\delta(X)$ *and* $\alpha > 0$ *such that for all* $\bar{\theta}$, $\theta \in \Theta$, $\left|\rho(Z,\bar{\theta}) - \rho(Z,\theta)\right| \leq \delta(X)\left|\left|\bar{\theta} - \theta\right|\right|^{\alpha}$ *and* $E\left[\delta(X)^2\right] < \infty$.

Assumption 4. $\rho(Z,\theta)$ *is twice continuously differentiable in a open and convex neighborhood* Δ *of* θ_0. $E\left[\left|\left|\frac{\partial^2\rho(Z,\theta)}{\partial\theta\partial\theta'}\right|\right|\right]$ *is bounded,* $E\left[\frac{\partial\rho(Z,\theta)}{\partial\theta}\frac{\partial\rho(Z,\theta)}{\partial\theta'}\right]$ *is nonsingular.* $E\left[\sup_{\theta\in\Delta}\left|\left|\frac{\partial\rho(Z,\theta)}{\partial\theta}\right|\right|^2\right] < \infty$, $E\left[\sup_{\theta\in\Delta}|\rho(Z,\theta)|^4|X\right] < \infty$, *and for all* $\theta \in \Delta$, $|\rho(Z,\theta) - \rho(Z,\theta_0)| \leq \delta(X)\,||\theta - \theta_0||$ *and* $E\left[\delta(X)^2\right] < \infty$. $E\left[\sup_{t\in\mathcal{T}}|w(t'\Phi(X))|^4\right] < \infty$. $K \geq p$, $E\left[q^K(X)\frac{\partial\rho(Z,\theta)}{\partial\theta'}\right]$ *is of full rank.*

Assumption 5. *Denote* U *as the support of* X, *for each* K *there is a constant scalar* $\xi(K)$ *and matrix* B *such that* $\tilde{q}^K(X) = Bq^K(X)$ *for every* $X \in U$, $sup_{X\in U}\left|\left|\tilde{q}^K(X)\right|\right| \leq \xi(K)$, $\sqrt{K} \leq \xi(K)$, *and* $E\left(\tilde{q}^K(X)\,\tilde{q}^K(X)'\right)$ *has smallest eigenvalue bounded away from zero.*

Assumptions 1 is a standard regularity condition. It restricts our analysis to an i.i.d context. It is possible to extend it to dependent data following De Jong (1996). Assumption 2 shows that we only need a $\sqrt{n}$-consistent estimator. Since we are dealing with a testing problem, we do not present the identification conditions of parameter estimation explicitly. To obtain a $\sqrt{n}$-consistent estimator, only an identification condition as weak as Dominguez and Lobato's (2004) is needed. Assumption 3 imposes some restrictions on second moment condition of $\rho(Z,\theta)$ and the smoothness of the function $\rho(Z,\theta)$. Assumption 4 is essential for asymptotic normality when the number of moment conditions is growing with the sample size. Assumption 5 imposes a normalization on the approximate function, bounds the second moment restriction away from singularity and restricts the magnitude of the series terms. The magnitude of the series terms is important, playing a crucial role in the asymptotic theory of GMM estimation and test when K increases with sample size n. Primitive conditions for this assumption are given in the case of $w(\cdot) = \exp(\cdot)$ when we discuss the improved Bierens (1990) statistic in Section 6.

In stead of focusing on the residual empirical process $\hat{M}\left(\hat{\theta}, t\right)$ on some interval Π, we propose a new empirical process, combining $w(t'\Phi(X_j))$ and $q^K(X)$. The idea is to form a new weighting functions by a proper linear combination of $w(t'\Phi(X))$ and $q^K(X)$ such that there does not exist the estimation effect for the transformed empirical process. More specifically, given $q^K(X)$ defined in (6), for any $t \in \Pi \subset \mathbb{R}^d$, the new empirical process is

$$\tilde{M}(\theta,t,K)=n^{-1/2}\sum_{j=1}^{n}\rho(Z_j,\theta)\left[\begin{array}{c} w(t'\Phi(X_j))- \\ \hat{b}(\theta,t)\hat{A}\left(\theta,K\right)\hat{\Lambda}\left(\theta,K\right)'\hat{\Omega}\left(\theta,K\right)^{-1}q^{K}\left(X_j\right)\end{array}\right], \tag{7}$$

where the new weighting function depends on

$$\hat{\Lambda}\left(\theta,K\right)=\frac{1}{n}\sum_{i=1}^{n}q^{K}\left(X_i\right)\frac{\partial\rho(Z_i,\theta)}{\partial\theta'},$$

$$\hat{A}\left(\theta,K\right)=\left[\hat{\Lambda}\left(\theta,K\right)'\hat{\Omega}\left(\theta,K\right)^{-1}\hat{\Lambda}\left(\theta,K\right)\right]^{-1}.$$

Remark: We will show in Theorem 1 that $\tilde{M}(\theta,t,K)$ does not suffer from the estimation effect. The intuition behind the removal of the estimation effect is that the transformed weighting function $w(t'\Phi(X))-\hat{b}(\theta,t)\hat{A}\left(\theta,K\right)\hat{\Lambda}\left(\theta,K\right)'\hat{\Omega}\left(\theta,K\right)^{-1}q^{K}\left(X_j\right)$ is orthogonal to $\frac{\partial\rho(Z_j,\theta)}{\partial\theta'}$ now.

Remark: The new empirical process combines Bierens' approach and the consistent tests of overidentifying restrictions. To see this point, we rewrite (7) into

$$\begin{aligned}\tilde{M}(\theta,t,K) &= n^{-1/2}\sum_{j=1}^{n}\rho(Z_j,\theta)w(t'\Phi(X_j)) \\ &\quad -\hat{b}(\theta,t)\hat{A}\left(\theta,K\right)\hat{\Lambda}\left(\theta,K\right)'\hat{\Omega}\left(\theta,K\right)^{-1}n^{-1/2}\sum_{j=1}^{n}q^{K}\left(X_j\right)\rho(Z_j,\theta).\end{aligned}$$

$n^{-1/2}\sum_{j=1}^{n}\rho(Z_j,\theta)w(t'\Phi(X_j))$, $t\in\Pi\subset\mathcal{T}$ is the element of Bierens' approach, $n^{-1/2}\sum_{j=1}^{n}q^{K}\left(X_j\right)\rho(Z_j,\theta)$ is the essential part of consistent tests of overidentifying restrictions.

Remark: The form of $\tilde{M}(\theta,t,K)$ also has a connection with the martingale transformation approach employed by Stute et al. (1998). In their case, $w\left(X,t\right)=I\left(X<t\right)$, but this function does not fall into the class $\mathcal{W}$. While martingale transformation approach focuses on obtaining asymptotic distribution-free statistics, our transformation focuses on obtaining optimal statistics under conditional moment restrictions.

In the following theorem, we will show the estimation-effect free property of the new empirical process for any fixed $K\geq p$.

Theorem 1. *When Assumptions 1 to 4 hold, K is fixed and $K\geq p$, under H_0, for any $t\in\Pi\subset\mathbb{R}^d$, given a $\sqrt{n}$ consistent test $\hat{\theta}$, under the null*

$$\tilde{M}\left(\hat{\theta},t,K\right)=\tilde{M}(\theta_0,t,K)+o_p\left(1\right), \tag{8}$$

and

$$\tilde{M}(\hat{\theta}, t, K) \xrightarrow{d} N\left[0, s^2(\theta_0, t, K)\right], \tag{9}$$

where

$$s^2(\theta_0, t, K) = E\left\{\rho(Z, \theta_0)^2\left[w(t'\Phi(X)) - b(\theta_0, t)A(\theta_0, K)\Lambda(\theta, K)'\Omega(\theta, K)^{-1}q^K(X)\right]^2\right\},$$

$$b(\theta_0, t) = E\left(w(X, t)\frac{\partial\rho(Z, \theta_0)}{\partial\theta'}\right),$$

$$\Omega(\theta_0, K) = E\left(\rho(Z, \theta_0)^2 q^K(X) q^K(X)\right),$$

$$\Lambda(\theta_0, K) = E\left(q^K(X)\frac{\partial\rho(Z, \theta_0)}{\partial\theta'}\right),$$

$$A(\theta_0, K) = \left[\Lambda(\theta_0, K)'\Omega(\theta_0, K)^{-1}\Lambda(\theta_0, K)\right]^{-1}.$$

Proof. See Appendix □

Equation (8) shows that, unlike $\hat{M}\left(\hat{\theta}, t\right)$, $\tilde{M}(\hat{\theta}, t, K)$ does not suffer from the estimation effect for a fixed K-the empirical process evaluated at any $\sqrt{n}$-consistent estimator is asymptotically the same as the empirical process evaluated at the true parameters. Note that, to reach the estimation-effect-free property in consistent tests of overidentifying restrictions, it is required that K should follow the rate condition $K \to \infty$ and $\xi(K)^2 K^2/n \to 0$, see Donald et al. (2003).

In order to obtain more powerful test we would like to allow K to increase with sample size n. We observe $\hat{A}(\theta, K)$ is an estimate of the asymptotic variance matrix of the optimal GMM estimator, which will generally be bounded below by the semiparametric efficiency bound. So We establish the asymptotic boundary of the new empirical process when K increases with sample size n in the following lemma. Denote $D(\theta, X) = E\left(\frac{\partial\rho(Z,\theta)}{\partial\theta}|X\right)$,

Lemma 2. *When Assumptions 1 to 5 hold, under H_0, for any $t \in \Pi \subset \mathbb{R}^d$, when $K \to \infty$ and $\xi(K)^2K/n \to 0$,*

$$\hat{A}(\theta_0, K) \xrightarrow{p} A_*(\theta_0)$$

where $A_*(\theta_0) = \left\{E\left[D(\theta_0, X)\sigma^{-2}(\theta_0, X)D(\theta_0, X)'\right]\right\}^{-1}$.

$$\hat{\Lambda}(\theta_0, K)'\hat{\Omega}(\theta_0, K)^{-1}n^{-1/2}\sum_{j=1}^{n}q^K(X_j)\rho(Z_j, \theta_0) \xrightarrow{p} n^{-1/2}\sum_{j=1}^{n}D(\theta_0, X_j)\sigma^{-2}(\theta_0, X_j)\rho(Z_j, \theta_0).$$

Proof. See Appendix □

Remark: The rate condition of K here leads to the asymptotic boundaries in $\tilde{M}(\hat{\theta}, t, K)$. This rate condition is the same as the one in efficient estimation. In consistent tests of overidentifying restrictions, some stronger rate condition of K is required to obtain estimation-effect-free property.

Based on the previous lemma, we can establish the following theorem.

Theorem 2. *When Assumptions 1 to 5 hold, under H_0, for any $t \in \Pi \subset \mathbb{R}^d$, when $K \to \infty$ and $\xi(K)^2 K/n \to 0$,*

$$\tilde{M}(\hat{\theta}, t, K) = n^{-1/2} \sum_{j=1}^{n} \rho(Z_j, \theta_0) \phi_*(\theta_0, X_j, t) + o_p(1),$$

where

$$\phi_*(\theta_0, X, t) = w(t'\Phi(X)) - b(\theta_0, t) A_*(\theta_0) D(\theta_0, X) \sigma^{-2}(\theta_0, X),$$

and

$$\tilde{M}(\hat{\theta}, t, K) \xrightarrow{d} N\left[0, s_*^2(\theta_0, t)\right],$$

where

$$s_*^2(\theta_0, t) = E\left[\rho(Z, \theta_0)^2 \phi_*(\theta_0, X, t)^2\right].$$

Proof. See Appendix □

Theorem 2 establishes the properties of $\tilde{M}(\hat{\theta}, t, K)$, when we allow K to increase with sample size n.

4 The Limit Distribution of ICM Test Under Local Alternatives

In the section, we will discuss the asymptotic properties of the ICM test based on the new residual empirical process, which is denoted as $\widehat{ICM}\left(\hat{\theta}, K\right)$ such that

$$\widehat{ICM}\left(\hat{\theta}, K\right) = \int_{\Pi} \left|\tilde{M}(\hat{\theta}, t, K)\right|^2 d\mu(t), \tag{10}$$

where $\mu(t)$ is a probability measure on Π that is absolutely continuous with respect to Lebesgue measure on $\Pi \subset \mathcal{T}$. We will show that $\widehat{ICM}\left(\hat{\theta}, K\right)$ has power against Pitman local alternatives, when K increases with sample size n.

We consider the following local alternative:

$$H_1^L : \rho(Z, \theta_0) = \frac{g(X)}{\sqrt{n}}, \tag{11}$$

where the $\rho(Z, \theta_0)$ is the same as under the null hypothesis.

Under this local alternative, we assume that

$$\tilde{M}(\hat{\theta}, t, K) = n^{-1/2} \sum_{j=1}^{n} \rho(Z_j, \theta_0)[w(t'\Phi(X_j)) - \hat{b}(\theta, t)\hat{A}(\theta, K)\hat{\Lambda}(\theta, K)'\hat{\Omega}(\theta, K)^{-1} q^K(X_j)] \tag{12}$$
$$- \frac{1}{n} \sum_{j=1}^{n} g(X_j)[w(t'\Phi(X_j)) - \hat{b}(\theta, t)\hat{A}(\theta, K)\hat{\Lambda}(\theta, K)'\hat{\Omega}(\theta, K)^{-1} q^K(X_j)] + o_p(1)$$

holds.

Let $C(\mathcal{T})$ be the metric space of all continuous real functions on $\mathcal{T}$ with metric $\lambda(z_1, z_2) = sup_{t\in\mathcal{T}} E|z_1(t) - z_2(t)|$. The Borel sets of $C(\mathcal{T})$ are the members of the σ–algebra generated by the open sets in $C(\mathcal{T})$. In order to guarantees the tightness of the process $\tilde{M}(\hat{\theta}, t, K)$ and the asymptotic normality of the finite distribution of $\tilde{M}(\hat{\theta}, t, K)$, when K increases with sample size n, we need the following assumption:

Assumption 6. *$w(t'\Phi(X))$ are differentiable, on $t \in \mathcal{T}$. $E\left[\sup_{t\in\mathcal{T}} \|\partial w(t'\Phi(X))/\partial t'\|^4\right] < \infty$, $Eg^2(X) < \infty$, $\sup_{t\in\mathcal{T},\theta\in\Theta} \|\partial b(\theta, t)/\partial t'\| < \infty$.*

Theorem 3. *Let Assumptions 1-6 and (12) hold, under H_1^L, when $K \to \infty$ and $\xi(K)^2K/n \to 0$, $\tilde{M}(\hat{\theta}, t, K)$ converges to z_*, where z_* is a Gaussian process on Π with mean function $\eta_*(t) = -E[g(X)\phi_*(\theta_0, X, t)]$ and covariance function $\Gamma_*(t_1, t_2) = E\left\{\rho(Z, \theta_0)^2\phi_*(\theta_0, X, t_1)\phi_*(\theta_0, X, t_2)\right\}$, $t_1, t_2 \in \mathcal{T}$. Moreover*

$$\widehat{ICM}\left(\hat{\theta}, K\right) \to \int_{\Pi} z_*^2(t)\, d\mu(t) \quad \textit{in distribution.}$$

Proof. See Appendix. □

5 Admissibility of the ICM Test under Unknown Conditional Heteroskedasticity

Intuitively $\widehat{ICM}$ or $\widehat{KS}$ statistics based on $\tilde{M}(\hat{\theta}, t, K)$ will be more powerful, since not only a continuum of unconditional moment restrictions but also extra K dimensional unconditional moment restrictions are considered, where K increases with sample size n. In this section, we will show the correctness of this intuition by proving the admissibility of the ICM test in the sense that under conditional heteroskedasticity there does not exist a test that is uniformly more powerful than the ICM test, provided the errors $\rho(Z, \theta_0)$ are conditionally normally distributed. To be specific, we assume the following condition:

Assumption 7. *$\rho(Z, \theta_0)|X \sim N\left(0, \sigma^2(\theta_0, X)\right)$ under the null (2); $\rho(Z, \theta_0) - \frac{g(X)}{\sqrt{n}}|X \sim N\left(0, \sigma^2(\theta_0, X)\right)$ under the local alternative (11).*

Remark: In Bierens and Ploberger (1997), conditional homoskedasticity has to be assumed; see Assumption C in the Appendix of Bierens and Ploberger (1997).

Theorem 4. *Under Assumptions 1-7, when $K \to \infty$ and $\xi(K)^2 K/n \to 0$, the ICM test in the form (10) is admissible.*

Proof. We rewrite (11) as

$$E\left[\frac{\rho(Z, \theta_0)}{\sigma(\theta_0, X)} - \frac{g(X)}{\sigma(\theta_0, X)\sqrt{n}}|X\right] = 0. \tag{13}$$

By Assumption 7, $\frac{\rho(Z,\theta_0)}{\sigma(\theta_0,X)} - \frac{g(X)}{\sigma(\theta_0,X)\sqrt{n}} \sim N(0,1)$. When $K \to \infty$ and $\xi(K)^2 K/n \to 0$, under (11)

$$\begin{aligned}
\tilde{M}(\hat{\theta}, t, K) &= n^{-1/2} \sum_{j=1}^{n} \frac{\rho(Z, \theta_0)}{\sigma(\theta_0, X_j)} \left[\sigma(\theta_0, X_j) w(t'\Phi(X_j)) - b(\theta_0, t) A_*(\theta_0) D(\theta_0, X) \sigma^{-1}(\theta_0, X_j)\right] \\
&\quad - \frac{1}{n} \sum_{j=1}^{n} \frac{g(X_j)}{\sigma(\theta_0, X_j)} [\sigma(\theta_0, X_j) w(t'\Phi(X_j)) - b(\theta_0, t) A_*(\theta_0) D(\theta_0, X) \sigma^{-1}(\theta_0, X_j)] + o_p(1).
\end{aligned}$$

It is equivalent to Bierens and Ploberger (1997) in the sense that for alternative model (13), we use weighting function $\sigma(\theta_0, X_j) w(t'\Phi(X_j))$ and the NLS estimator, so by Theorem 6 in Bierens and Ploberger (1997), we get to the conclusion. □

6 A Simple Bootstrapping Procedure for the ICM Test

In the previous section, we have shown $\widehat{ICM}\left(\hat{\theta}, K\right)$ is more powerful, when allowing K to increase with the sample size. The new consistent test also allows us to propose a much simpler bootstrap procedure than the standard ones. Note that the asymptotic distribution of $\widehat{ICM}\left(\hat{\theta}, K\right)$ is nonpivotal, some bootstrapping procedure has to be used. In this context, wild bootstrap (WB) introduced by Wu (1986) has been used extensively, see Stute et. al. (1998). Here we propose a version of WB bootstrapping procedure as follows

1. Given a $\sqrt{n}$-consistent estimator $\hat{\theta}$, obtain the residuals $\rho(Z_i, \hat{\theta})$, $1 \leq i \leq n$.

2. Generate WB residuals according to $\rho^*(Z_i, \hat{\theta}) = \rho(Z_i, \hat{\theta})V_i$ for $1 \leq i \leq n$, with V_i a sequence of iid random variables with mean 0, unit variance, and bounded support and that are independent of the sequence Z. Given a proper value of K,

3. Compute

$$\tilde{M}^*(\hat{\theta}, t, K) = n^{-1/2} \sum_{j=1}^{n} \rho^*(Z_i, \hat{\theta}) \left[w(t'\Phi(X_j)) - \hat{b}(\theta, t) \hat{A}(\theta, K) \hat{\Lambda}(\theta, K)' \hat{\Omega}(\theta, K)^{-1} q^K(X_j) \right]$$

$$\widehat{ICM}^*\left(\hat{\theta}, K\right) = \int_{\Pi} \left| \tilde{M}^*(\hat{\theta}, t, K) \right|^2 d\mu(t).$$

4. Repeat step 2 and 3 B times. where B is an integer, to obtain $\widehat{ICM}^{*1}\left(\hat{\theta}, K\right), ..., \widehat{ICM}^{*B}\left(\hat{\theta}, K\right)$.

5. Let the estimator of the distribution of $\widehat{ICM}(\theta_0, K)$ be the discrete distribution with $\Pr\left(\widehat{ICM}(\theta_0, K) = \widehat{ICM}^{*b}\left(\hat{\theta}, K\right)\right) = 1/B$.

Let $\hat{q}_{\alpha}^{B}$ here be the $1-\alpha$ quantile of the distribution estimator from step 5, A test that rejects if

$$\widehat{ICM}\left(\hat{\theta}, K\right) \geq \hat{q}_{\alpha}^{B}.$$

The standard bootstrap for ICM statistic requires that one solve B nonlinear optimization problems to obtain B bootstrap estimators. These estimators then used to construct bootstrap ICM statistic. The boostrap proposed here even does not need to solve a nonlinear optimization problem, given a $\sqrt{n}$-consistent estimator $\hat{\theta}$, which simplifies the computation greatly.

Samples of $\{V_i\}_{i=1}^{n}$ sequences are iid Bernoulli variates with $P(V_i = a_1) = p_{1,}$ $P(V_i = a_2) = 1 - p_1$, where $a_1 = 0.5\left(1-\sqrt{5}\right)$, $a_2 = 0.5\left(1+\sqrt{5}\right)$, and $\left(1+\sqrt{5}\right)/2\sqrt{5}$. See Mammen (1993) for motivation of the chosen numbers.

Theorem 5. *Let Assumptions 1-7 and (12) hold, under H_1^L, when $K \to \infty$ and $\xi(K)^2 K/n \to 0$, $\widehat{ICM}^* \left(\hat{\theta}, K\right) \to \int z_*^2(t)\, d\mu(t)$ in distribution.*

Proof. See Appendix. □

7 Bierens (1990) Test Based On The New Empirical Process

In the case of KS tests, Bierens (1990)'s procedure is attractive, since its null asymptotic distribution is tractable, and time-consuming bootstrap procedures are avoided.

Bierens (1990) considers the conditional model defined in Example 1, so $\rho(Z, \theta_0) = Y - f(X, \theta_0)$. Moreover $w(\cdot) = \exp(\cdot)$, so $q^K(X) = (\exp(t_1'\Phi(X)), \cdots, \exp(t_K'\Phi(X)))'$. In this case, we can give primitive conditions for Assumption 6.

Assumption 8. *Choose $(t_1', \cdots, t_K')'$ such that $t_j \in \mathbb{R}^d \backslash \Pi$ for $j = 1, \cdots, K$, and $t_j \neq t_i$ for any $j, i = 1, \cdots, K$. For $\Phi(X)$, the Borel measurable bounded one-to-one mapping from $\mathbb{R}^d$ into $\mathbb{R}^d$, has a probability density function that is bounded away from zero. When $K \geq p$, $E\left[q^K(X) \frac{\partial f(X,\theta_0)}{\partial \theta'}\right]$ is of full rank.*

This assumption imposes restrictions on the probability density function of X. Similar assumption has been used by Newey (1997) in the case of series estimation of nonparametric and semiparametric models. This assumption also sets the rules of how to choose $(t_1', \cdots, t_K')'$ and the subset Π, but they are hardly restrictive.

Lemma 3. *Assumption 9 implies Assumption 5, furthermore $\xi(K) = CK^{1/2}$, where $C > 0$ is a constant. Finally, for any $t \in \Pi$, $s^2(\theta_0, t, K) > 0$.*

Proof. See Appendix. □

Our approach conveniently avoids the extreme condition that $s^2(\theta_0, t, K) = 0$ for any proper K. In Bierens (1990), an additional assumption has to be imposed, and it only could be established that set $S_{NLS}^* = \left\{t \in \mathbb{R}^d : s_{NLS}^2(\theta_0, t) = 0\right\}$, where

$$s_{NLS}^2(\theta_0, t) = E\left\{\rho(Z, \theta_0)^2 \left[\exp(t'\Phi(X)) - b(\theta_0, t) A_{NLS}(\theta_0) \frac{\partial f(X, \theta_0)}{\partial \theta}\right]^2\right\},$$

$A_{NLS}(\theta_0) = \left[E\left(\frac{\partial f(X,\theta_0)}{\partial \theta} \frac{\partial f(X,\theta_0)}{\partial \theta'}\right)\right]^{-1}$, has Lebesgue measure zero and is not dense in $\mathbb{R}^d$.

In practice, given any proper K, the function $s^2(\theta_0, t, K)$ can be consistently estimated by

$$\hat{s}^2(\hat{\theta}, t, K) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - f(X_j, \hat{\theta}))^2[\exp(t'\Phi(X_j)) - \hat{b}(\theta, t)\hat{A}\,(\theta, K)\,\hat{\Lambda}\,(\theta, K)'\,\hat{\Omega}\,(\theta, K)^{-1}\,q^K\,(X_j)]^2,$$

and

$$\tilde{W}\left(\hat{\theta}, t, K\right) = \frac{\left[\,\tilde{M}(\hat{\theta}, t, K)\right]^2}{\hat{s}^2\left(\hat{\theta}, t, K\right)}$$

is well defined for any $t \in \Pi$ for sample size large enough.

Lemma 1 of Bierens (1990) shows that under H_1

$$S_{NLS} = \left\{t \in \mathbb{R}^d : E\left[(Y - f(X, \theta_0))\exp(t'\Phi(X)) - b(\theta_0, t)A_{NLS}\,(\theta_0)\,\frac{\partial f(X, \theta_0)}{\partial \theta}\right] = 0\right\}$$

has Lebesgue measure zero. Following the same logic, we can obtain that under H_1, the set

$$S = \left\{t \in \mathbb{R}^d : E\left[(Y - f(X, \theta_0))\left(\exp(t'\Phi(X)) - b(\theta_0, t)A\,(\theta_0, K)\,\Lambda\,(\theta, K)'\,\Omega\,(\theta, K)^{-1}\,q^K(X)\right)\right] = 0\right\}$$

has Lebesgue measure zero for any K.

In terms of the choice of Π, Bierens (1990) has to assume that $\Pi \subset \mathbb{R}^d \backslash S_{NLS} \cup S^*_{NLS}$, however we only need to assume $\Pi \subset \mathbb{R}^d \backslash S$ here. This may have important impact on the testing power, since the testing power negatively relies on the size of the set $S_{NLS} \cup S^*_{NLS}$ or S. The new test statistic may have more power than the Bierens (1990) test, where the NLS estimator is employed, even in the case of conditional homoskedasticity.

Following Bierens (1990), we maximize $\tilde{W}\left(\hat{\theta}, t, K\right)$ over a subset Π of $\mathbb{R}^d$. The results are summarized in the following theorem.

Theorem 6. *Let Assumptions 1-5 and 7 hold, under H_0, when $K \to \infty$ and $K^2/n \to 0$, $\tilde{W}\left(\hat{\theta}, t, K\right)$ converges weakly to z_*^2 under H_0, where z_* is a Gaussian element of $C(\Pi)$ with mean zero, and covariance function*

$$\Gamma_*\,(t_1, t_2) = E\left\{[Y - f(X, \theta_0)]^2\,\phi_*(\theta_0, X, t_1)\phi_*(\theta_0, X, t_2)/\sqrt{s_*^2\,(\theta_0, t_1)}\sqrt{s_*^2\,(\theta_0, t_2)}\right\}. \qquad (14)$$

Moreover, $\tilde{W}\left(\hat{\theta}, \tilde{t}, K\right)$ with $\tilde{t} = \arg\max_{t\in\Pi}\tilde{W}\left(\hat{\theta}, t, K\right)$ converges in distribution to $\sup_{t\in\Pi} z_*^2\,(t)$.

Proof. See Appendix □

Note that the covariance function $\Gamma_*(t_1, t_2)$ depends on the DGP of the model, so does the distribution of $sup_{t\in\Pi} z_*^2(t)$. Then critical values should be tabulated for each model and each DGP. Normally some bootstrap procedure should be applied to overcome this problem. Bierens (1990) circumvents this by introducing some penalty functions. The alternative procedure similar to Bierens (1990) is the following. Choose independently of the data generating process real numbers $\gamma > 0$, $\rho \in (0,1)$, and a point $t_0 \in \Pi$. Let $\tilde{t} = \arg\max_{t\in\Pi} \tilde{W}\left(\hat{\theta}, t, K\right)$ and let

$$\bar{t} = t_0, \text{ if } \tilde{W}\left(\hat{\theta}, \tilde{t}, K\right) - \tilde{W}\left(\hat{\theta}, t_0, K\right) \leq \gamma n^{\rho};\ \bar{t} = \tilde{t}, \text{ if } \tilde{W}\left(\hat{\theta}, \tilde{t}, K\right) - \tilde{W}\left(\hat{\theta}, t_0, K\right) \geq \gamma n^{\rho}.$$

Then we have the following theorem.

Theorem 7. *Let Assumptions 1-5, 6 hold. then under H_0, when $K \to \infty$ and $K^2/n \to 0$, $\tilde{W}\left(\hat{\theta}, \bar{t}, K\right) \to \chi_1^2$ in distribution.*

Proof. Similar to the proof of Bierens (1990) Theorem 4. □

In practice, it may be quite laborious to determine $\tilde{t} = \arg\min_{t\in\Pi} \tilde{W}\left(\hat{\theta}, t, K\right)$ on the continuum set Π. We can simplify this problem by discretizing the maximum problem by the following theorem.

Theorem 8. *Choose a sequence of positive integers L converging to infinity with n, and choose a sequence (t_i) such that $\{t_1, t_2, t_3, \cdots\}$ is dense in Π. Replace $\tilde{t}$ by $t = argmax_{t\in\{t_1,\cdots,t_L\}} \tilde{W}(\hat{\theta}, t, K)$. Then the previous two theorems carry over for $\tilde{t}$.*

Proof. Similar to the proof of Bierens (1990) Theorem 5. □

In practice, we have to choose value K in a given sample. Monte Carlo simulations in section 8 shows that the size and power properties of the new Bierens' test is not quite sensitive to the choice of K.

8 Monte Carlo Simulations

We analyze in the following Monte Carlo simulations the finite sample properties of the improved test, comparing with the Bierens test where the NLS estimator is employed. Since we have choose value K in practice. We will check how sensitive is the size and power properties of new test to the different K value.

Let z, v_1, v_2, and u follow the independent standard normal distribution, and let the regressors be $X_1 = z + v_1$, $X_2 = z + v_2$. The dependent variable is generated according to

$$Y = 1 + X_1 + X_2 + e.$$

Under the null, when the homoskedasticity is assumed, $e = u$, under heteroskedasticity, $e = \left(0.1 + 0.5x_1^2\right)^{1/2} u$. In both cases, OLS is employed to obtain the parameters estimator. Based on the OLS estimator and residuals, we calculate Bierens (1990) test and our improved Bierens (1990) test. Following Bierens (1990), we choose $L = [n/10]-1$ and $\Pi = [1,5]\times[1,5]$. $(t_1, \cdots, t_L)'$ have been drawn randomly from the uniform distribution on Π. $(t_1, \cdots, t_K)'$ have been drawn randomly from the uniform distribution on subset $[-1,1]\times[-1,1]$. We use the weighting function with $\Phi(x_1, x_2) = \left(\tan^{-1}(x_1/2), \tan^{-1}(x_2/2)\right)'$. The Monte Carlo simulations have been conducted for sample size 200 and 400 with four sets of values of the penalty parameters

$$\begin{array}{cc} \{\gamma = 1, \rho = 0.5\} & \{\gamma = 0.5, \rho = 0.5\} \\ \{\gamma = 0.25, \rho = 0.5\} & \{\gamma = 0.25, \rho = 0.25\}. \end{array}$$

For both sample sizes, we report the results of choosing K starting from 3 to 20. Note that $K = 3$ is minimum dimension requirement for the model.

For the empirical size check, 10,000 replications are used. We report the results in Figures 1-4. Firstly note that in both homoskedasticity and heteroskedasticity cases, the empirical size of the new statistic is quite stable or becomes stable quickly as K increases. In the homoskedasticity case, its empirical size properties are comparable to Bierens (1990)'s statistic, even when $K = 3$. In the heteroskedasticity case, under reasonable penalty parameters situations, while Bierens (1990) statistic is undersized, the empirical size of the new statistic is a little bit undersized, when K is a small number; it becomes very close to the nominal size, when K increases. Note that when penalty parameters are too small ($\{\gamma = 0.25, \rho = 0.25\}$), both statistics are all heavily oversized.

For the power check, 1000 replications are used. We consider the following alternatives

DGP 1.1: $Y = 1 + X_1 + X_2 + v_1^2 v_2^2 + u$.

DGP 1.2: $Y = 1 + X_1 + X_2 + v_1^2 v_2^2 + \left(0.1 + 0.5x_1^2\right)^{1/2} u$.

DGP 2.1: $Y = 1 + X_1 + X_2 + (1 + X_1 + X_2)\exp\left[-0.01(1 + X_1 + X_2)^2\right] + u$.

DGP 2.2: $Y = 1 + X_1 + X_2 + (1 + X_1 + X_2)\exp\left[-0.01(1 + X_1 + X_2)^2\right] + \left(0.1 + 0.5x_1^2\right)^{1/2} u$.

DGP 3.1: $Y = 1 + X_1 + X_2 + \sin(1 + X_1 + X_2) + u$.

DGP 3.2: $Y = 1 + X_1 + X_2 + \sin(1 + X_1 + X_2) + \left(0.1 + 0.5x_1^2\right)^{1/2} u$.

DGP 4.1: $Y_j = 1 + X_1 + X_2 + \cos(1 + X_1 + X_2) + u$.

DGP 4.2: $Y_j = 1 + X_1 + X_2 + \cos(1 + X_1 + X_2) + \left(0.1 + 0.5x_1^2\right)^{1/2} u$.

Remark: The second alternative is the same as the alternative 3 in Escanciano (2006a); The third is similar to the alternative 4 in Escanciano (2006a). The fourth changes the sine function in the third alternative into a cosine function.

We report the results in Figures 5-12. To save space, results of sample size 200 and 400 are reported in one figure for each alternative. For the first alternative, our new statistic is better than Bierens (1990)'s in both homoskedasticity and heteroskedasticity cases when K is relatively large. For the alternative 2, 3 and 4, under homoskedasticity, the power of the new statistic is quite close to Bierens (1990)'s test for all the K we considered. In heteroskedasticity case, the new statistic has very good power properties even when K is small; as K increases, the difference of the power between the new statistic and Bierens (1990)'s test reaches as much as 20%.

All in all, the new statistic has good size properties and improves the power significantly when there exists heteroskedasticity of unknown form. The choice of K is not restrictive. For a large number of alternatives, the general pattern of the test results is that when K increases, we obtain better power properties.

9 Conclusion

In this paper, we have proposed a new consistent conditional test, combining the Bierens' approach and consistent test of overidentifying restrictions. It relies on a transformation-based empirical process. The new empirical process enjoys some advantages of both approaches. Firstly it is not affected by the uncertainty from the parameter estimation. Moreover This estimation-effect-free property requires much less restrictive rate condition than in consistent tests of overidentifying restrictions alone. Furthermore the ICM test based on the new empirical process has power against Pitman local alternatives. We proved, under some regularity conditions, the admissibility of the ICM test based on this transformation-based empirical process in the case that there exists heteroskedasticity of unknown form, extending the result in Bierens and Ploberger (1997). It would be useful to know how to choose K in practice. This will be left for future work.

Appendix

Proof of Corollary 1. Without loss of generality we may assume that X is bounded itself, so that we may choose $\Phi(X) = X$. We set $w(t_1'X) = 1$. It is always possible to normalize $q^K(X)$ into this case when $w(t_1'X) \neq 1$. Firstly it is easy to check that $q(t_jX) \in L_2$, for $j = 1, 2, \cdots$. For $K = 2, 3, \cdots$, let

$$\varpi_K(X) = \sum_{j=1}^{K} \alpha_{K,j} w\left(t_j'X\right),$$

where $\alpha_{K,K} = 1$, and the other $\alpha_{K,j}$ are chosen such that

$$E\left[\varpi_K(X) w\left(t_j'X\right)\right] = 0 \text{ if } j < K.$$

For $K = 1, 2, \cdots$, define function $\psi_K(X)$ on the range of X such that

$$\psi_1(X) = 1,$$

$$\psi_K(X) = \begin{cases} \varpi_K(X) / \left[E\varpi_K(X)^2\right]^{1/2}, & \text{if } \left[E\varpi_K(X)^2\right] > 0 \\ 0, & \text{if } \left[E\varpi_K(X)^2\right] = 0 \end{cases}$$

for $K > 1$. Then $\psi_K(X)$, $K = 1, 2, \cdots$ form an orthonormal system of the Hilbert space of H of Borel measurable functions φ on the range of X satisfying $E\left[\varphi(X)^2\right] < \infty$, with inner product $(\psi_K, \varphi) = E\left[\psi_K(X)\varphi(X)\right]$. Then by Theorem 2.4.2 of Brockwell and Davis (1991), for any ε, there exists a positive integer K and constant $c_1, \cdots, c_K$ such that

$$\left[E\left(\varphi(X) - \sum_{j=1}^{K} c_j \psi_j(X)\right)^2\right]^{1/2} < \varepsilon,$$

then the conclusion follows. $\square$

Proof of Theorem 1. To prove this theorem, we rewrite $\tilde{M}(\hat{\theta}, t, K)$ into

$$\begin{aligned}\tilde{M}(\hat{\theta}, t, K) &= n^{-1/2}\sum_{j=1}^{n}\rho(Z_j, \hat{\theta})w(t'\Phi(X_j)) \\ &\quad -\hat{b}(\hat{\theta}, t)\hat{A}\left(\hat{\theta}, K\right)\hat{\Lambda}\left(\hat{\theta}, K\right)'\hat{\Omega}\left(\hat{\theta}, K\right)^{-1} n^{-1/2}\sum_{j=1}^{n} q^K\left(X_j\right)\rho(Z_j, \hat{\theta}). \\ &= A_1 - A_2\end{aligned}$$

By the mean value theorem we have

$$A_1 = n^{-1/2}\sum_{j=1}^{n}\rho(Z_j, \theta_0)w(t'\Phi(X_j)) + \hat{b}(\dot{\theta}, t)n^{1/2}\left(\hat{\theta} - \theta_0\right),$$

where $\dot{\theta}$ lies on the line joining $\hat{\theta}$ and θ_0, $\dot{\theta}$, $\hat{\theta} \in \Delta$, an open convex neighborhood of θ_0, with $\dot{\theta} \xrightarrow{p} \theta_0$. By Assumption 4 and the fact that $E[w^2(t'\Phi(X))]$ is finite, the dominance condition holds by Cauchy-Schwarz inequality

$$\begin{aligned}E\left[\sup_{\theta\in\Delta}\left\|w\left(t'\Phi(X)\right)\frac{\partial\rho(Z_j, \theta)}{\partial\theta'}\right\|\right] &= E\left[\left|\left|w\left(t'\Phi(X)\right)\right|\right|\sup_{\theta\in\Delta}\left\|\frac{\partial\rho(Z, \theta)}{\partial\theta'}\right\|\right] \\ &\leq \left[E\left|\left|w\left(t'\Phi(X)\right)\right|\right|^2\right]^{1/2}\left[E\sup_{\theta\in\Delta}\left\|\frac{\partial\rho(Z, \theta)}{\partial\theta'}\right\|^2\right]^{1/2} \\ &< \infty.\end{aligned}$$

So we have weakly uniformly convergence of

$$p\lim\sup_{\theta\in\Delta}\left\|\frac{1}{n}\sum_{j=1}^{n} w\left(t'\Phi(X)\right)\frac{\partial\rho(Z_j, \theta_0)}{\partial\theta'} - E\left[w\left(t'\Phi(X)\right)\frac{\partial\rho(Z, \theta)}{\partial\theta'}\right]\right\| = 0,$$

Following the Lemma 1 in Wang (2015), then $\hat{b}(\dot{\theta}, t)-$ $\hat{b}(\theta_0, t) = o_p(1)$. So

$$A_1 = n^{-1/2}\sum_{j=1}^{n}\rho(Z_j, \theta_0)w(t'\Phi(X_j)) + \hat{b}(\theta_0, t)n^{1/2}\left(\hat{\theta} - \theta_0\right) + o_p(1).$$

Similarly

$$E\left[\sup_{\theta\in\Delta}\left\|q^K(X)\frac{\partial\rho(Z,\theta)}{\partial\theta'}\right\|\right] = E\left[\left|\left|q^K(X)\right|\right|\sup_{\theta\in\Delta}\left\|\frac{\partial\rho(Z,\theta)}{\partial\theta'}\right\|\right]$$
$$\leq \left[E\left|\left|q^K(X)\right|\right|^2\right]^{1/2}\left[E\sup_{\theta\in\Delta}\left\|\frac{\partial\rho(Z,\theta)}{\partial\theta'}\right\|^2\right]^{1/2}$$
$$< \infty.$$

Then

$$p\lim\sup_{\theta\in\Delta}\left\|\frac{1}{n}\sum_{j=1}^{n}q^K(X)\frac{\partial\rho(Z_j,\theta)}{\partial\theta'} - E\left[q^K(X)\frac{\partial\rho(Z,\theta)}{\partial\theta'}\right]\right\| = 0,$$

By similar argument, we have

$$n^{-1/2}\sum_{j=1}^{n}\rho(Z_j,\hat{\theta})q^K(X_j) = n^{-1/2}\sum_{j=1}^{n}\rho(Z,\theta_0)q^K(X_j) + \hat{\Lambda}(\theta_0,K)n^{1/2}\left(\hat{\theta}-\theta_0\right) + o_p(1),$$

Similarly

$$E\left[\sup_{\theta\in\Delta}\left\|q^K(X)q^K(X)'[\rho(Z,\theta)]^2\right\|\right] = E\left[\left|\left|q^K(X)q^K(X)'\right|\right|\sup_{\theta\in\Delta}\left\|[\rho(Z,\theta)]^2\right\|\right]$$
$$\leq \left[E\left|\left|q^K(X)q^K(X)'\right|\right|^2\right]^{1/2}\left[E\sup_{\theta\in\Delta}[\rho(Z,\theta)]^4\right]^{1/2}$$
$$< \infty.$$

Then by similar argument, $\hat{\Omega}\left(\hat{\theta},K\right) - \hat{\Omega}(\theta_0,K) = o_p(1)$. Also by Assumptions 4 and 5, for any $K > p$, $\Omega(\theta_0,K)$ is positive definite. By continuous mapping theorem, for $K > p$, $\hat{\Lambda}\left(\hat{\theta},K\right)'\hat{\Omega}\left(\hat{\theta},K\right)^{-1}\hat{\Lambda}\left(\hat{\theta},K\right) - \hat{\Lambda}(\theta_0,K)'\hat{\Omega}(\theta_0,K)^{-1}\hat{\Lambda}(\theta_0,K) = o_p(1)$, $\hat{\Lambda}\left(\hat{\theta},K\right)'\hat{\Omega}\left(\hat{\theta},K\right)^{-1} - \hat{\Lambda}(\theta_0,K)'\hat{\Omega}(\theta_0,K)^{-1} = o_p(1)$. Then

$$\begin{aligned}
A_2 &= \hat{b}(\theta_0,t)\hat{A}(\theta_0,K)\hat{\Lambda}(\theta_0,K)'\hat{\Omega}(\theta_0,K)^{-1} \\
&\quad \times n^{-1/2}\left(\sum_{j=1}^{n}q^K(X_j)\rho(Z_j,\theta_0) + \hat{\Lambda}(\theta_0,K)n^{1/2}\left(\hat{\theta}-\theta_0\right) + o_p(1)\right) + o_p(1) \\
&= \hat{b}(\theta_0,t)\hat{A}(\theta_0,K)\hat{\Lambda}(\theta_0,K)'\hat{\Omega}(\theta_0,K)^{-1}n^{-1/2}\sum_{j=1}^{n}q^K(X_j)\rho(Z_j,\theta_0) \\
&\quad + \hat{b}(\theta_0,t)n^{1/2}\left(\hat{\theta}-\theta_0\right) + o_p(1)
\end{aligned}$$

So we prove that (8) holds.

To prove (9), we rewrite $\tilde{M}(\hat{\theta}, t, K)$ as

$$\tilde{M}(\hat{\theta}, t, K) = [1, -\hat{b}((\theta_0, t)\hat{A}(\theta_0, K)\hat{\Lambda}(\theta_0, K)'\hat{\Omega}(\theta_0, K)^{-1}]n^{-1/2}\sum_{j=1}^{n}\rho(Z_j, \theta_0)\left(w(t'\Phi(X_j)), q^K(X_j)'\right)' + o_p(1). \tag{15}$$

By Lindberg-Feller central limit theory and Slutsky theorem, we have

$$\tilde{M}(\hat{\theta}, t, K) \xrightarrow{d} N\left[0, s^2(\theta_0, t, K)\right]$$

□

Proof of Lemma 2. Denote $\bar{\Omega}(\theta_0, K) = \frac{1}{n}\sum_{i=1}^{n} E\left(\rho(Z_i, \theta)^2 | X_i\right) q^K(X_i) q^K(X_i)'$. By applying Lemma A.3 in Donald et al. (2003), $\hat{\Lambda}(\theta_0, K)'\bar{\Omega}(\theta_0, K)^{-1}\hat{\Lambda}(\theta_0, K) \xrightarrow{p} A_*(\theta_0)^{-1}$. By following the proof of Theorem 5.4 in Donald et al. (2003), we can get $\hat{\Lambda}(\theta_0, K)'\bar{\Omega}(\theta_0, K)^{-1}\hat{\Lambda}(\theta_0, K) - A(\theta_0, K)^{-1} \xrightarrow{p} 0$.So $A(\theta_0, K) \xrightarrow{p} A_*(\theta_0)$.

By applying Lemma A.4 in Donald et al. (2003),

$$\hat{\Lambda}(\theta_0, K)'\bar{\Omega}(\theta_0, K)^{-1} n^{-1/2}\sum_{j=1}^{n} q^K(X_j)\rho(Z_j, \theta_0) \xrightarrow{p} n^{-1/2}\sum_{j=1}^{n} D(\theta_0, X_j)\sigma^{-2}(\theta_0, X_j)\rho(Z_j, \theta_0).$$

Also following the proof of Theorem 5.4 in Denald et al. (2003), we can get

$$\hat{\Lambda}(\theta_0, K)'\bar{\Omega}(\theta_0, K)^{-1} n^{-1/2}\sum_{j=1}^{n} q^K(X_j)\rho(Z_j, \theta_0) - \hat{\Lambda}(\theta_0, K)'\hat{\Omega}(\theta_0, K)^{-1} n^{-1/2}\sum_{j=1}^{n} q^K(X_j)\rho(Z_j, \theta_0) \xrightarrow{p} 0.$$

Then

$$\hat{\Lambda}(\theta_0, K)'\hat{\Omega}(\theta_0, K)^{-1} n^{-1/2}\sum_{j=1}^{n} q^K(X_j)\rho(Z_j, \theta_0) - n^{-1/2}\sum_{j=1}^{n} D(\theta_0, X_j)\sigma^{-2}(\theta_0, X_j)\rho(Z_j, \theta_0) \xrightarrow{p} 0.$$

□

Proof of Theorem 2. Based on Lemma 2, it is easy to get

$$\tilde{M}(\hat{\theta}, t, K) = n^{-1/2}\sum_{j=1}^{n}\rho(Z_j, \theta_0)\phi_*(\theta_0, X_j, t) + o_p(1).$$

By Lindberg-Feller central limit theory

$$n^{-1/2}\sum_{j=1}^{n} D\left(\theta_0, X_j\right)\sigma^{-2}\left(\theta_0, X_j\right)\rho(Z_j,\theta_0) \xrightarrow{d} N\left(0, A_*(\theta_0)^{-1}\right).$$

Then we can get

$$s_*^2\left(\theta_0, t\right) = E\left[\rho(Z,\theta_0)^2\phi_*(\theta_0, X, t)^2\right]$$

□

Proof of Theorem 3. We need to show that the finite distributions of the process $\tilde{M}(\hat{\theta}, t, K)$ converge to normal distributions, and that $\tilde{M}(\hat{\theta}, t, K)$ is tight as $K \to \infty$ and $\xi(K)^2K/n \to 0$. The asymptotic normality of the finite distributions of $\tilde{M}(\hat{\theta}, t, K)$ follows easily from the Liapunov-type version in Bierens (1994, Theorem 6.1.7) of McLeish's (1974) martingale difference central limit theorem. The tightness of $\tilde{M}(\hat{\theta}, t, K)$ follows from Lemma A.1 in Bierens and Ploberger (1997). Denote $\Upsilon = \sup_{t\in\mathcal{T}} \left\|\partial w\left(t'\Phi\left(X\right)\right)/\partial t'\right\|$. Firstly note $E\left[\Upsilon\right] \leq \left\{E\left[\Upsilon^4\right]\right\}^{1/4} < \infty$, and $w(t_1'\Phi(X))$ is differentiable, so for any $t_1, t_2 \in \mathcal{T}$, $|w(t_1'\Phi(X)) - w(t_2'\Phi(X))| \leq \Upsilon\left|t_1 - t_2\right|$. Secondly, $E\left[\rho(Z,\theta_0)^2\Upsilon^2\right] \leq E^{1/2}\left[\rho(Z,\theta_0)^4\right]E^{1/2}\left[\Upsilon^4\right] < \infty$. By continuous mapping theorem, we get the asymptotic distribution of $\widehat{ICM}\left(\hat{\theta}, K\right)$. □

Proof of Theorem 5. Since $e_i^*(\hat{\theta}) = e_i(\hat{\theta})V_i$ for $1 \leq i \leq n$, (12), it is easy to get

$$\begin{aligned}\tilde{M}^*(\hat{\theta}, t, K) = n^{-1/2}&\sum_{j=1}^{n}\rho(Z_j,\theta_0)V_i[w(t'\Phi(X_j)) - \hat{b}(\theta_0,t)\hat{A}\left(\theta_0, K\right)\hat{\Lambda}\left(\theta_0, K\right)'\hat{\Omega}\left(\theta_0, K\right)^{-1}q^K\left(X_j\right)]\\ &-\frac{1}{n}\sum_{j=1}^{n}g\left(X_j\right)[w(t'\Phi(X_j)) - \hat{b}(\theta_0,t)\hat{A}\left(\theta_0, K\right)\hat{\Lambda}\left(\theta_0, K\right)'\hat{\Omega}\left(\theta_0, K\right)^{-1}q^K\left(X_j\right)] + o_p(1),\end{aligned}$$

when $K \to \infty$ and $\xi(K)^2K/n \to 0$. Then the conclusion follows similar to the proof of Theorem 4. □

Proof of Lemma 3. We still assume that X is bounded itself, so that we may choose $\Phi(X) = X$. We set $\exp(t_1'X) = 1$. Note that we can always normalize $q^K(X)$ into $q^K(X) = (1, \exp((t_2 - t_1)'X), \cdots, \exp((t_K - t_1)'X))'$. For $K = 1, 2, \cdots$, since the probability density function of X is bounded away from zero, then the second moment of $\varpi_K(X)$ defined in the proof of Lemma 1 is larger than zero, that is $\left[E\varpi_K(X)^2\right] > 0$ almost surely. So for $K = 2, 3, \cdots$,

$$\psi_K(X) = \varpi_K(X)/\left[E\varpi_K(X)^2\right]^{1/2}.$$

For any K, define $\tilde{q}^K(X) = (\psi_1(X), \cdots, \psi_K(X))'$. When $t_{jK} \neq t_{iK}$ for $j, i = 1, \cdots, K$, $\tilde{q}^K(X)$ is linear transformation of $q^K(X)$: $q^K(X) = B\tilde{q}^K(X)$, where B is a nonsingular lower triangular matrix. So $\tilde{q}^K(X) = B^{-1}q^K(X)$. Since $(\psi_1(X), \cdots, \psi_K(X))'$ is an orthonormal set, so $E\left(\tilde{q}^K(X)\,\tilde{q}^K(X)'\right) = I_K$, which means that the condition of nonsingularity is satisfied.

Note that $||\tilde{q}^K(X)|| = [\sum_{j=1}^K \psi_j(X)^2]^{1/2}$, $\varpi_K(X) = \sum_{j=1}^K \alpha_{K,j}\exp\left(t_j' X\right)$, and X is bounded, So we have

$$\begin{aligned}\sup_{X\in U}||\tilde{q}^K(X)|| &\leq C[\sum_{k=1}^K 1]^{1/2}\\ &\leq CK^{1/2}.\end{aligned}$$

To prove $s^2(\theta_0, t, K) > 0$, note that for any $t \in \Pi$, $t \neq t_j$ for $j = 1, \cdots, K$. Denote $q^{K+1}(X) = (\exp(t_1'\Phi(X)), \cdots, \exp(t_K'\Phi(X)), \exp(t'\Phi(X)))'$. Then we can obtain that $E\left(q^{K+1}(X)\,q^{K+1}(X)'\right)$ has smallest eigenvalue bounded away from zero based on Lemma 2. Note that $E[(Y - f(X, \theta_0))|X]^2 > 0$, then $E\left((Y_j - f(X_j, \theta_0))^2\,q^{K+1}(X)\,q^{K+1}(X)'\right)$ is positive definite. From (15) in the proof of Theorem 1 it is easy to obtain that $s^2(\theta_0, t, K) > 0$. □

Proof of Theorem 6. Under H_0, Define

$$\begin{aligned}z_n(\theta_0, t, K) &= n^{-1/2}\sum_{j=1}^n [Y_j - f(X_j, \theta_0)][\exp(t'\Phi(X_j))\\ &\quad -\hat{b}(\theta_0, t)\hat{A}(\theta_0, K)\,\hat{\Lambda}(\theta_0, K)'\,\hat{\Omega}(\theta_0, K)^{-1}\,q^K(X_j)]/\sqrt{(}s^2(\theta_0, t, K)).\end{aligned}$$

Following the Proof of Theorem 1, we have under H_0

$$p\lim_{n\to\infty}\sup_{t\in\Pi}|\tilde{W}(\hat{\theta}, t, K) - z_n^2(\theta_0, t, K)| = 0.$$

Following the Proof of Lemma 4 in Bierens (1990), we can obtain under H_0, $z_n(\theta_0, t, K)$ is tight. Then We allow $K \to \infty$, $\frac{K^2}{n} \to 0$, the following result holds

$$p\lim_{n\to\infty}\sup_{t\in\Pi}|z_n^2(\theta_0, t, K) - z_*^2(\theta_0, t)| = 0.$$

It is also easy to prove that for arbitrary $t_1, \cdots, t_m$ in Π, $(z_n(\theta_0, t_1, K), \cdots, z_n(\theta_0, t_m, K))'$ is asymptotically distributed as $(z_*(\theta_0, t_1), \cdots, z_*(\theta_0, t_m))'$. Then z_n converges weakly to z_*. Following the functional limit theory of Billingsley (1968 p. 47), we have the results. □

References

[1] Bierens H.J., 1982. Consistent model specification test. Journal of Econometrics 20, 105-134.

[2] Bierens H.J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443-1458.

[3] Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129-1152.

[4] Billingsley, P., 1968. Convergence of Probability Measures. New York: Wiley.

[5] Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods (2nd Edition). Springer.

[6] Chamberlain, G., 1987. Asymptotic efficiency in estimation with conditional moment restrictions. Journal of Econometrics 34, 305–334.

[7] Chamberlain, G., 1992. Comment: sequential moment restrictions in panel data. Journal of Business and Economic Statistics 10, 20-26.

[8] Carrasco, M., Florens J.P., 2000. Generalization of GMM to a continuum of moment conditions. Econometric Theory 16, 797-834.

[9] De Jong, R.M., 1996. The Bierens test under data dependence. Journal of Econometrics 72, 1-32.

[10] Domínguez, M.A., Lobato,I.N., 2004. Consistent estimation of models defined by conditional moment restrictions. Econometrica 72, 1601-1615.

[11] Domínguez, M.A., Lobato,I.N., 2015. A Simple Omnibus Overidentification Specification Test for Time Series Econometric Models. Econometric Theory 31, 891-910.

[12] Donald, G., Imbens, W., Newey, W.K., 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. Journal of Econometrics 117, 55-93.

[13] Escanciano, J.C., 2006a. A consistent diagnostic test for regression models using projections. Econometric Theory 22, 1030-1051.

[14] Escanciano, J.C., 2006b. Goodness-of-fit statistics for linear and nonlinear time series models. Journal of the American Statistical Association 101, 531-541.

[15] Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64, 865-890.

[16] Gozalo, P.L., 1993. A consistent model specification test for nonparametric estimation of regression function models. Econometric Theory 9, 451-477.

[17] Hahn, J., 1997. Efficient estimation of panel data models with sequential moment restrictions. Journal of Econometrics 79, 1-21.

[18] Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. The Annals of Statistics 21, 1926-1947.

[19] Hong, Y., White, H., 1995. Consistent specification testing via nonparametric series regression. Econometrica 63, 1133-1159.

[20] Mammen, E., 1993. Bootstrap and wild bootstrap for high-dimensional linear model. Annals of Statistics 21, 225–285.

[21] Mcleish, D. L., 1974. Dependent Central Limit Theorems and Invariance Principles. Annals of Probability, 2, 620-628.

[22] Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.

[23] Stinchcombe, M., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295-325.

[24] Stute, W., 1997. Nonparametric model checks for regression. The Annals of Statistics 25, 613-641.

[25] Stute, W., Thies, S., and Zhu, L. X., 1998. Model checks for regression: an innovation process approach. The Annals of Statistics 26, 1916-1934.

[26] Stute, W., Gonzalez-Manteiga, W., and Presedo-Quindimil, M., 1998. Boot-strap Approximations in Model Checks for Regression. Journal of the American Statistical Association, 93, 141–149.

[27] Wang, X., 2015. A general approach to conditional moment specification testing with projections. Econometric Reviews, Forthcoming

[28] White, H., 1994. Estimation, Inference and Specification Analysis. New York: Cambridge University Press.

[29] Wu, C. F. J., 1986. Jacknife, Bootstrap and Other Resampling Methods in Regression Analysis (with discussion). The Annals of Statistics, 14, 1261–1350.

[30] Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics 75, 263-289.

Figure 1: Size of testing at 5% level, Sample size 200, $u_j = e_j$

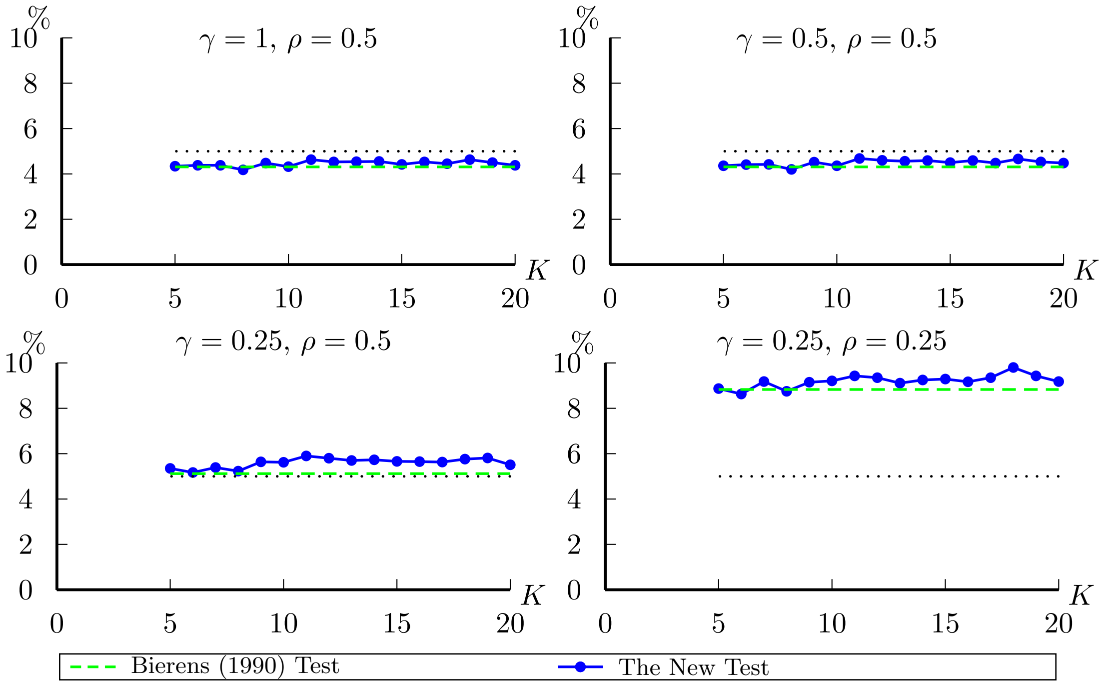

Figure 2: Size of testing at 5% level, Sample size 400, $u_j = e_j$

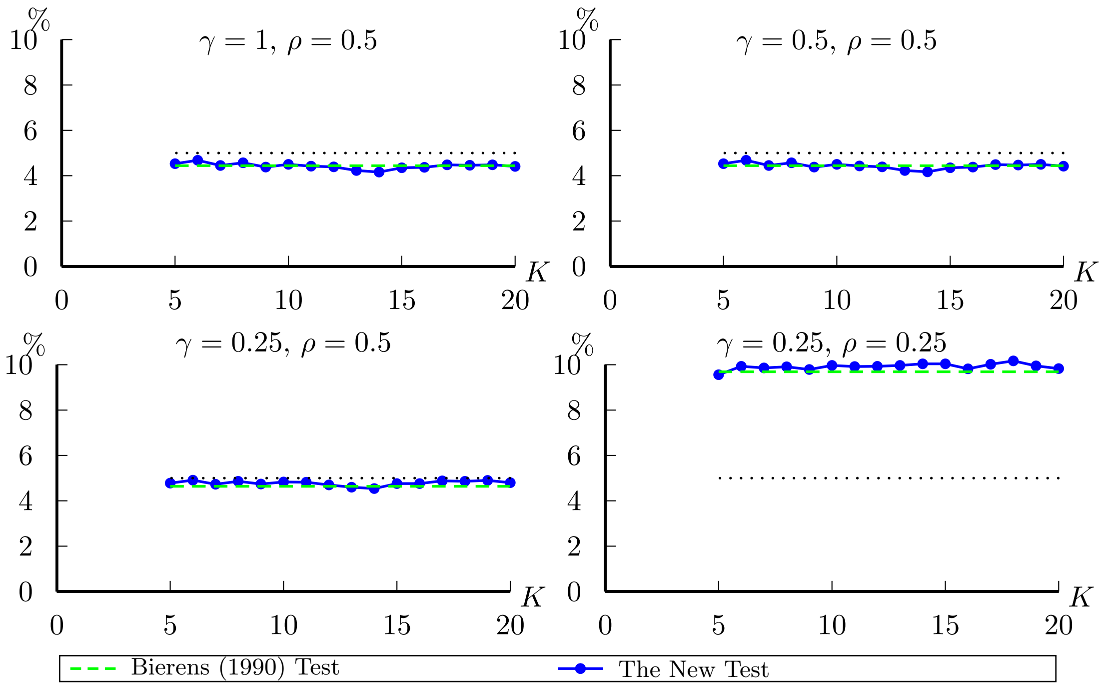

Figure 3: Size of testing at 5% level, Sample size 200, $u_j = \left(0.1 + 0.5x_{1j}^2\right)^{1/2} e_j$

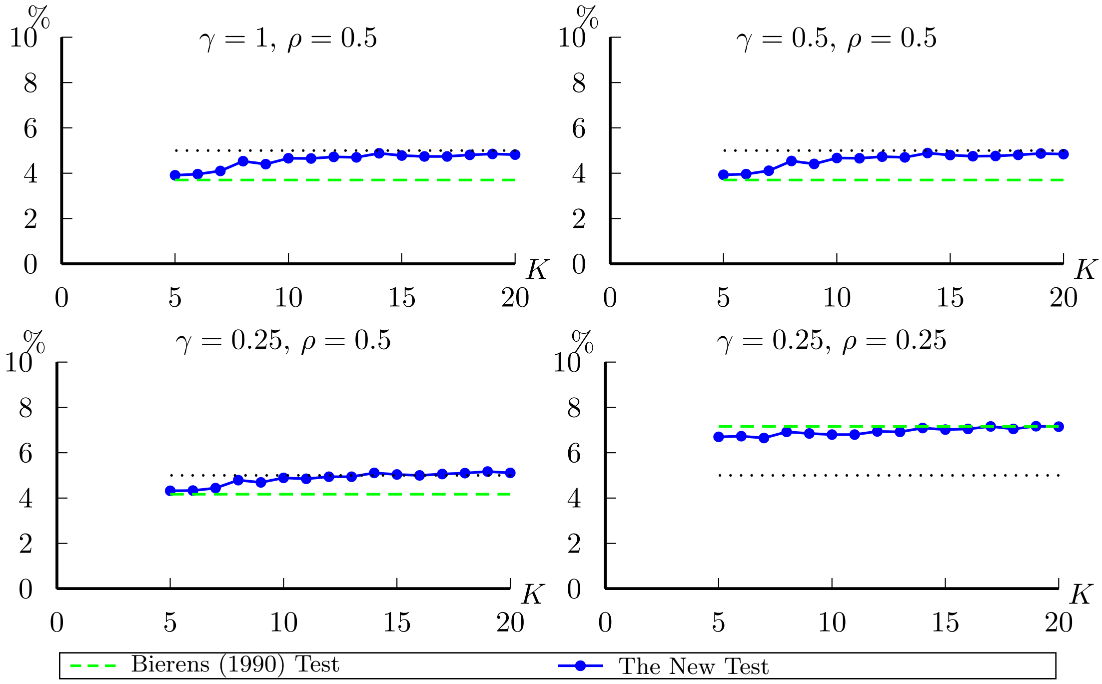

Figure 4: Size of testing at 5% level, Sample size 400, $u_j = \left(0.1 + 0.5x_{1j}^2\right)^{1/2} e_j$

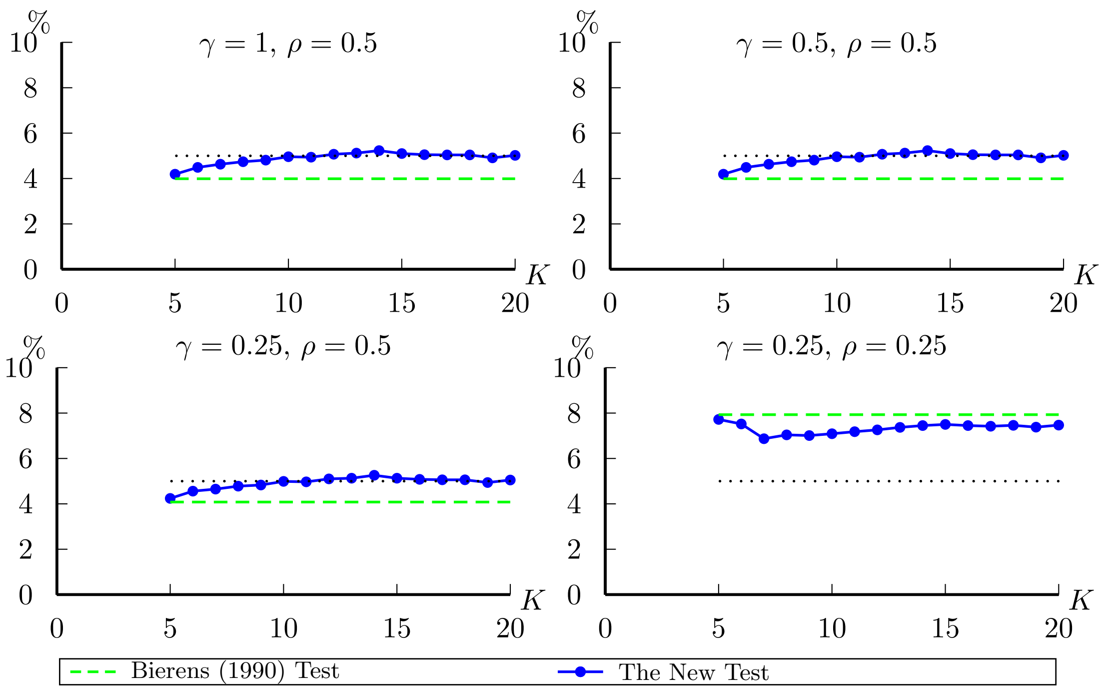

Figure 5: Power of testing at 5% level, DGP 1.1

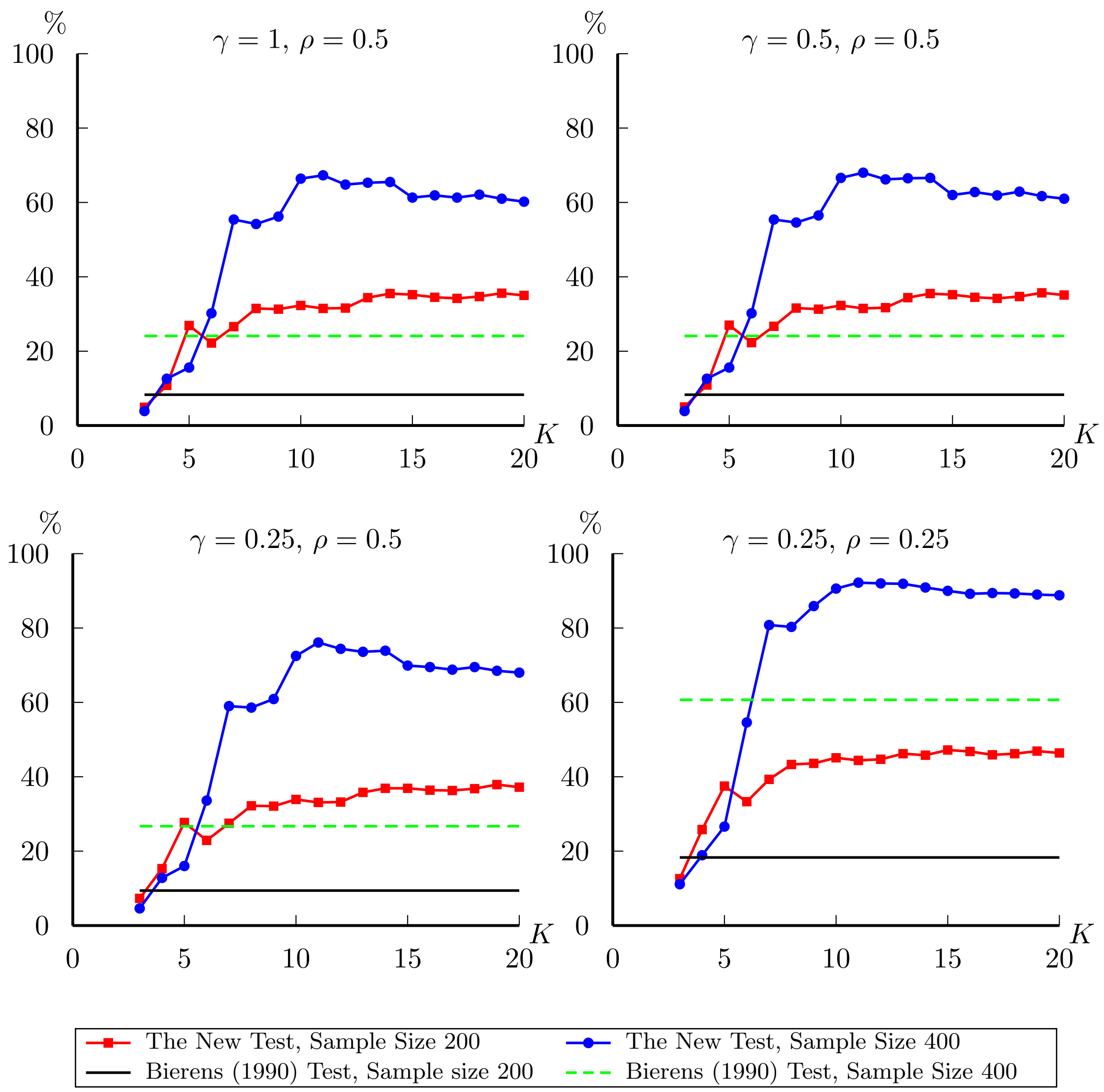

Figure 6: Power of testing at 5% level, DGP 1.2

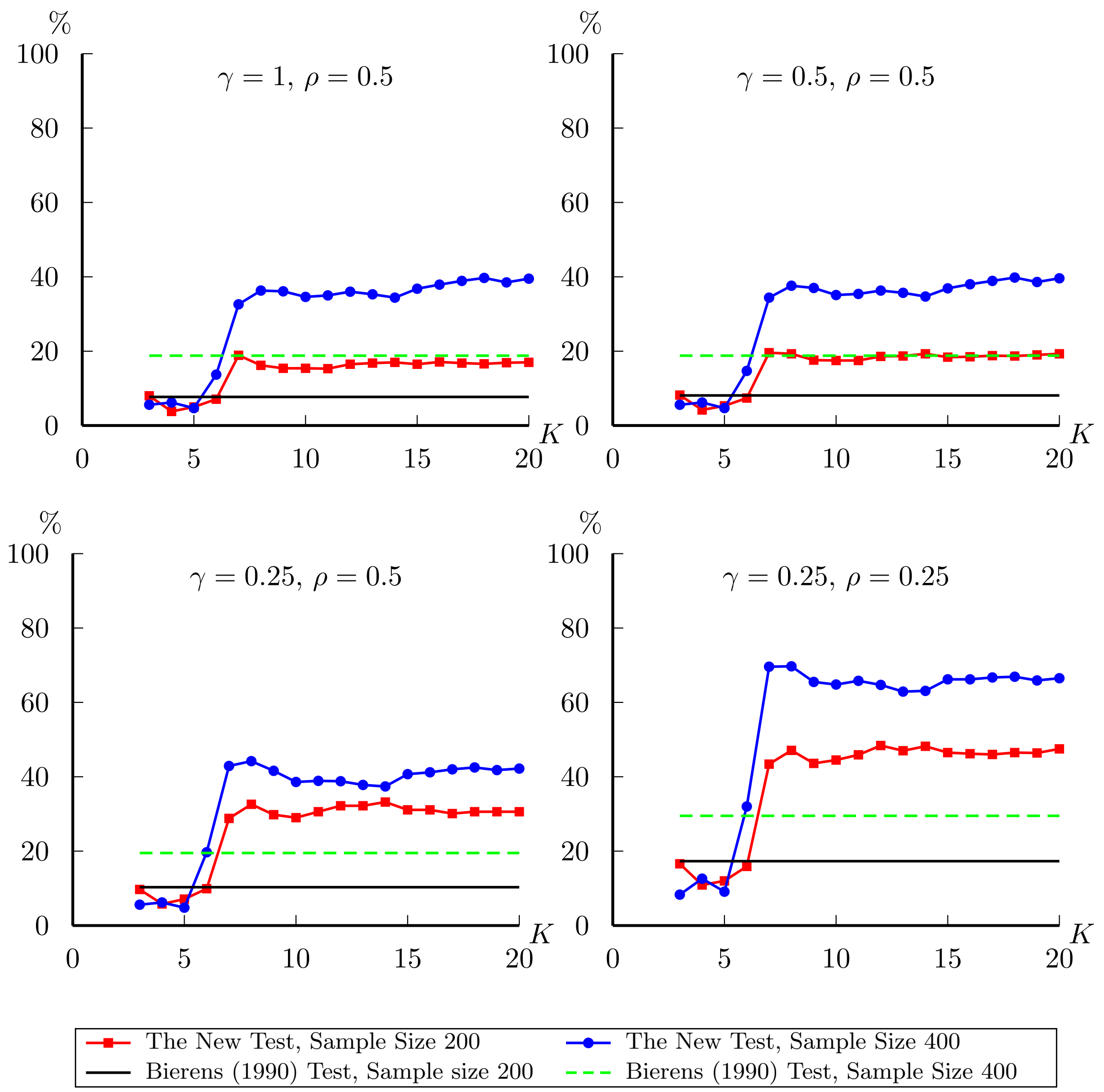

Figure 7: Power of testing at 5% level, DGP 2.1

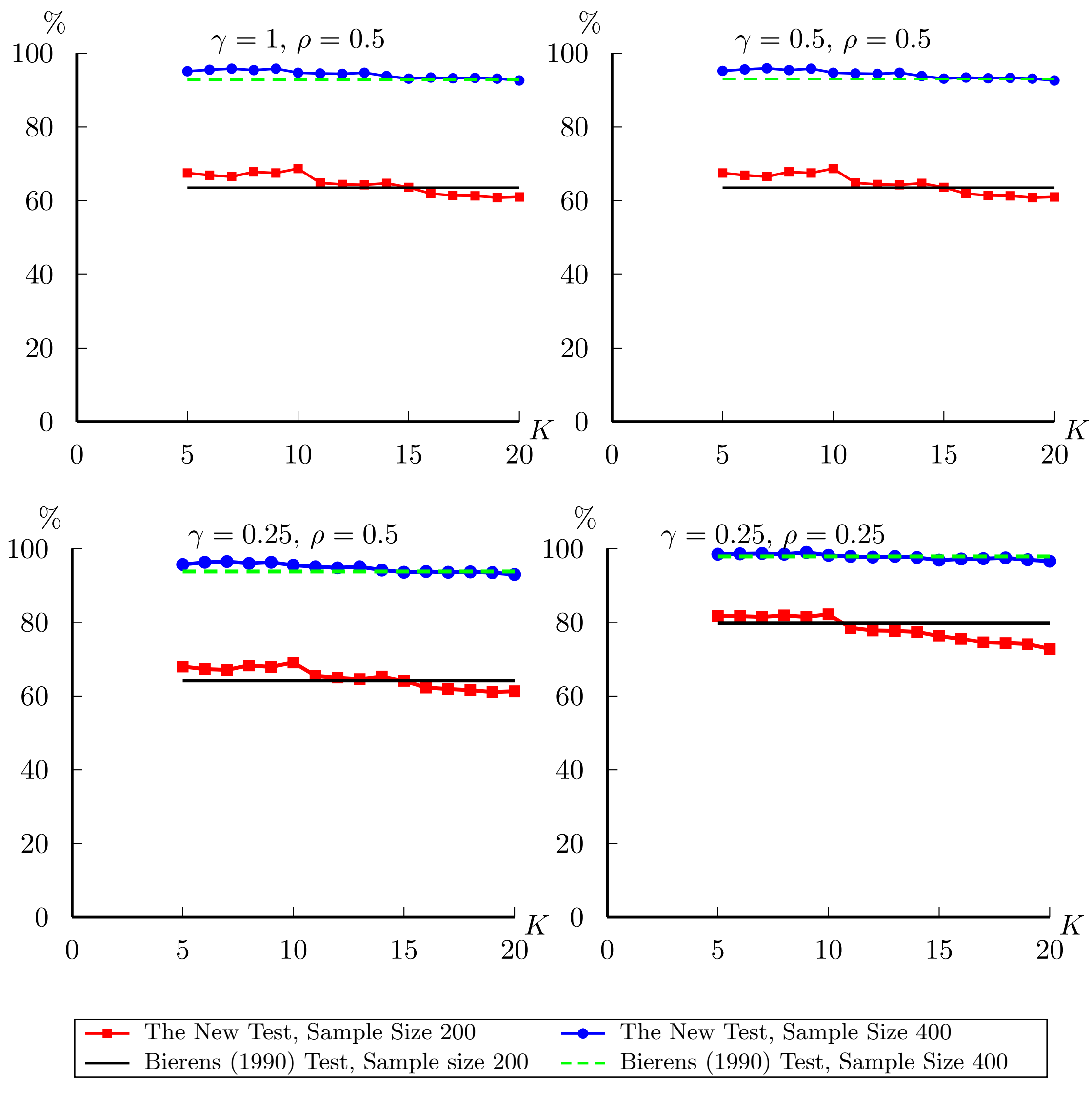

Figure 8: Power of testing at 5% level, DGP 2.2

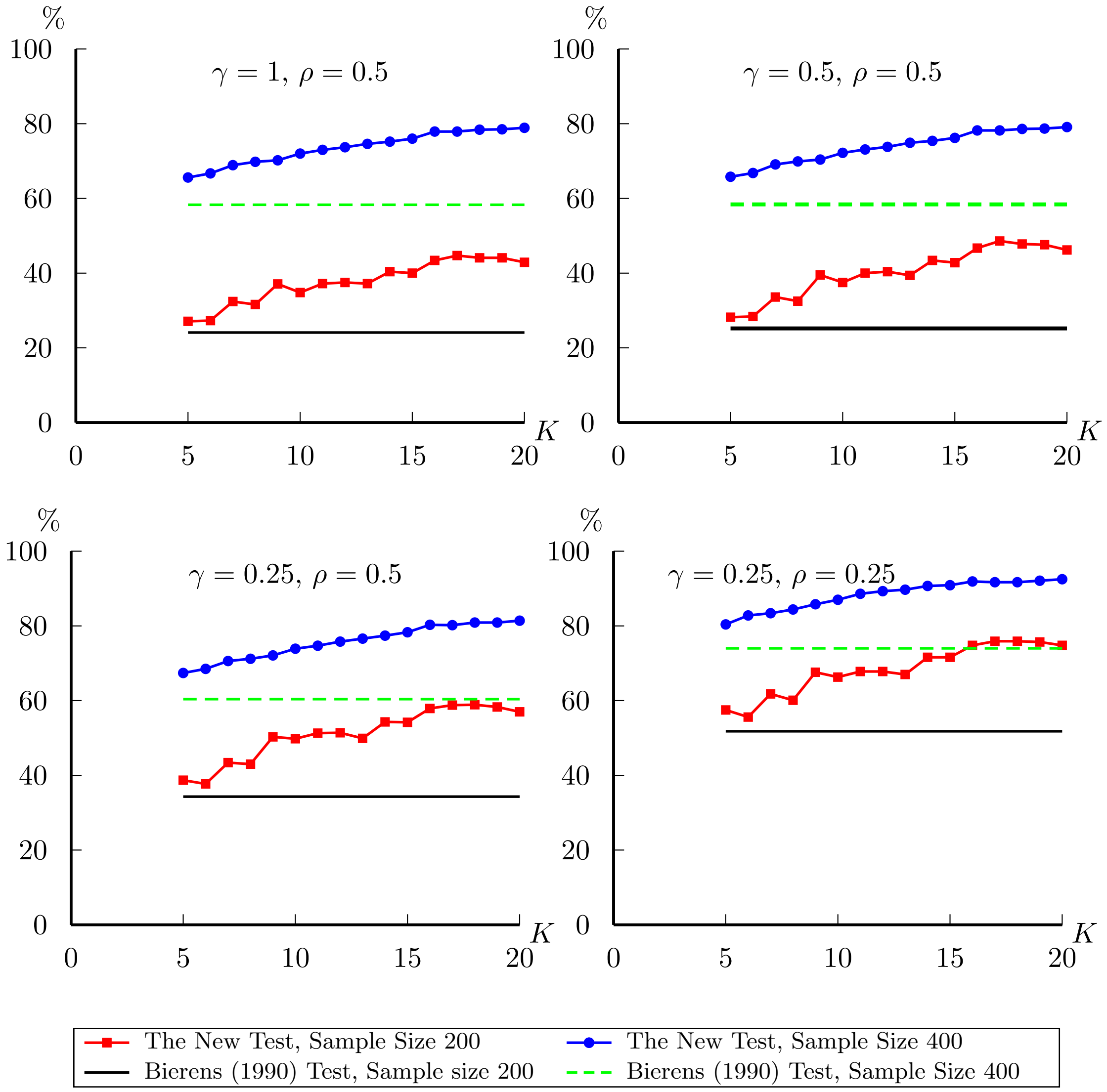

Figure 9: Power of testing at 5% level, DGP 3.1

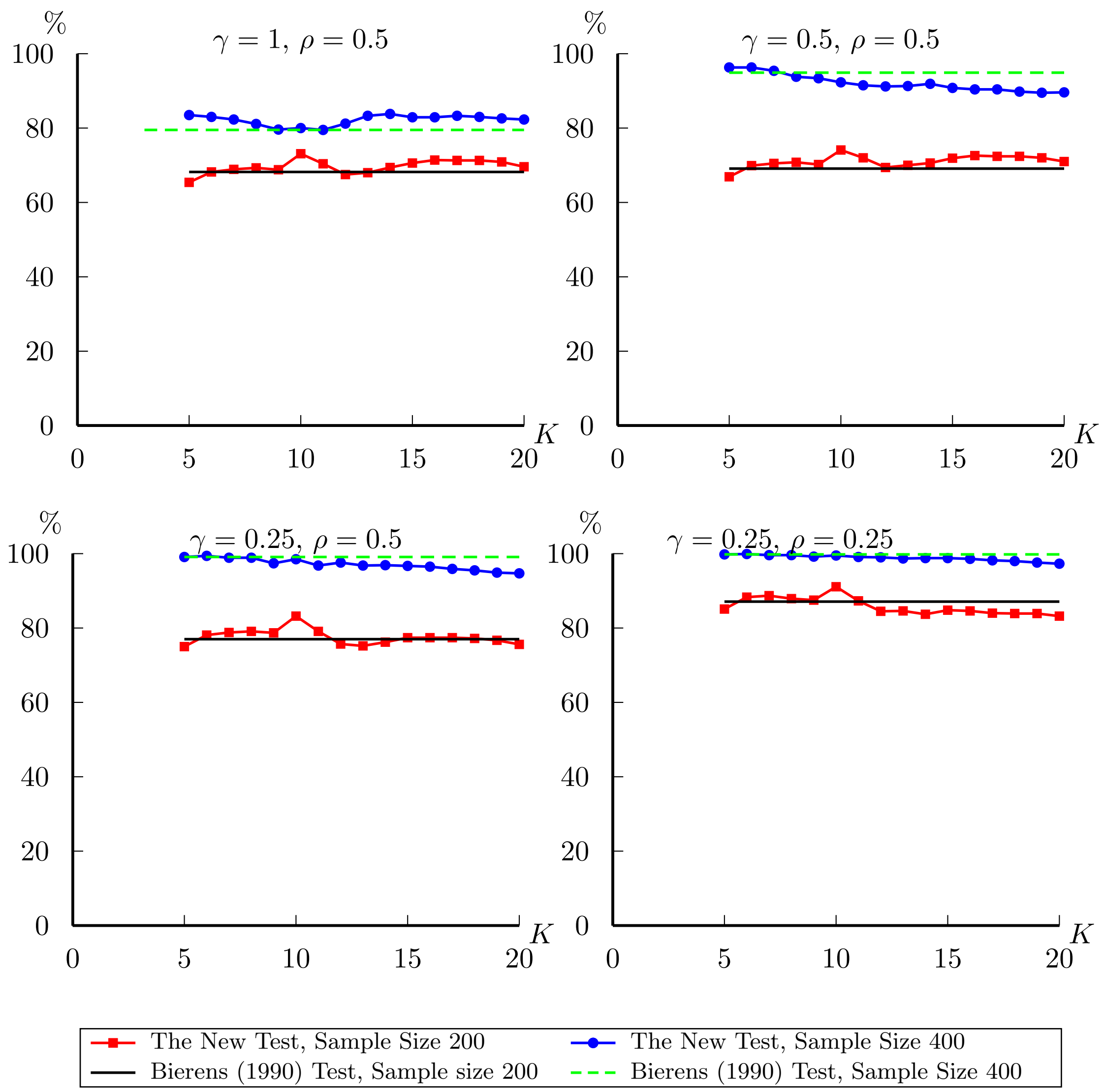

Figure 10: Power of testing at 5% level, DGP 3.2

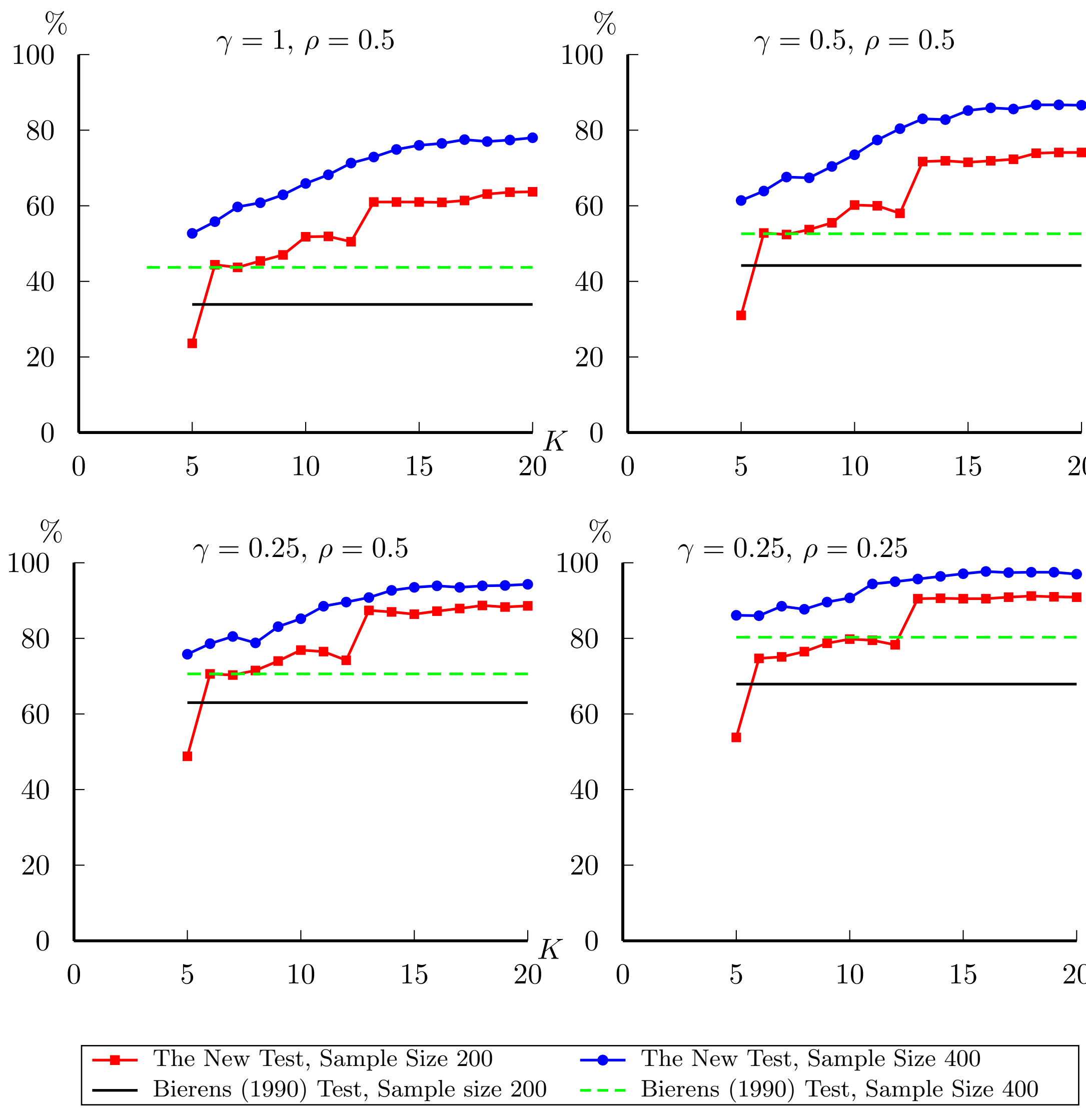

Figure 11: Power of testing at 5% level, DGP 4.1

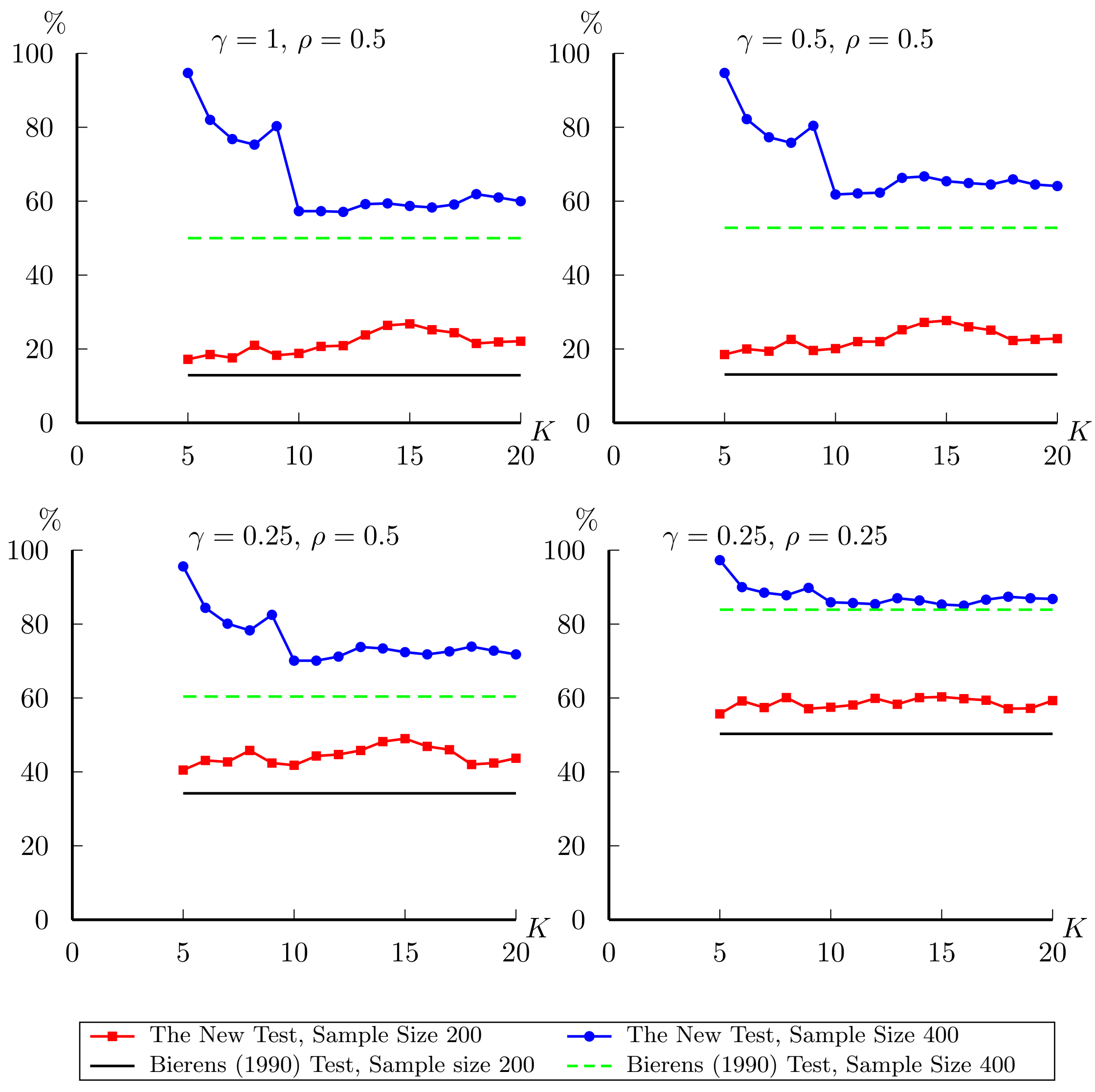

Figure 12: Power of testing at 5% level, DGP 4.2