A Test Of Independence In Econometric Models (DRAFT)

Ivan Medovikov *

April 8, 2011

1 Introduction

Independence assumption is central to econometric modeling and many other statistical applications. We propose a new version of a non-parametric, margin-free test of statistical independence between vectors of continuous random variables of arbitrary dimension based on the Cramer-von-Mises type functionals of the empirical copula process. Our test is closely related to the family of tests proposed by Kojadinovic and Holmes (2009), Quessy (2010) and Genest and Remillard (2004), which are based on the decomposition of the empirical copula process into independent sub-processes using the formula of Möbius. In order to test for independence within a set of d random variables, the decomposition-based tests require calculating $2^d - d - 1$ integral-type test statistics from the data. While the limiting properties of the decomposition-based tests are easier to track, such regularity can be at the expense of test power. In this work we propose a test using a single integral-type test statistic calculated on the entire empirical copula of the data. Furthermore, the key difference of our test is the use of a weighted functional norm in the construction of the test statistic.

To my knowledge, the issue of the weighting function has not been explored in the context of independence testing. The weighting function establishes a connection between the classical tests of independence such as that of Blum, Kiefer, and Rosenblatt (1961) and the more recent copula-based tests. In particular, we show that the test-statistic of Blum, Kiefer, and Rosenblatt (1961) based on the empirical density function, and the copula-based statistics of Deheuvels (1979), Genest and Remillard (2004), Kojadinovic and Holmes (2009) can all be expressed as special cases of the weighted test considered here. The two tests arise from the weighted copula test by choosing specific respective weighting functions.

We obtain a closed-form expression for our weighted integral-type test statistic for a general class of integrable additively and product-separable weighting functions and investigate the asymptotics of the test statistic under the null of independence. When at least one of the variables has the dimension greater than two, the test statistic is no longer distribution-free. We find thatthe weighted version of the test has higher power than earlier copula-based tests for a range of alternatives. We also consider a version of the test where independent sampling is replaced by pseudo-observations, such as regression residuals. We propose a way to use the test to verify some of the key regularity assumptions in econometric models, such as homoscedasticity, lack of endogeneity, and model mis-specification.

The test can be extended to many econometric applications, including heteroscedasticity, serial correlation and model mis-specification testing. Other than its power, the key advantage of the test is the ability to detect departures from independence of any form.

The rest of this paper is organized as follows. Section 1.2 reviews the existing non-parametric copula tests of independence. Section 2 presents the new test and the associated theory. Section 3 provides the simulation study designed to assess the power of the test. Some of the applications of the new test together with possible extensions are discussed in Section 4.

*Department of Economics, University of Western Ontario. Web: http://publish.uwo.ca/~imedovik, E-mail: imedovik@uwo.ca

1.1 Empirical Copula Process

Let $(X_1, ..., X_p) = \mathbf{X}$ be continuous real-valued random variables of dimensions $d_1, ..., d_p$ respectively. In many statistical applications it is often important to test for independence among the X_p's. It is well understood that the dependence structure of $\mathbf{X}$ is best described by the so-called Copula function $C(u)$. Let F be the distribution function of $\mathbf{X}$, and F_j, $j \in S = \{1, ..., p\}$, be marginal distribution function of X_j. Copula $C(u)$ establishes a connection between the joint distribution function of random vector $\mathbf{X}$ and the marginal distributions of its components. In particular, Sklar (1959) has shown that $F(\mathbf{X}) = C(F_1, ..., F_p)$. A copula is therefore itself a distribution function of uniform random variables which take values in the unit cube.

In what follows, we adopt the notation and setup of Kojadinovic and Holmes (2009). Given a random sample $\{(X_{i1}, ..., X_{id})\}_{i=1}^n$ from F, where $d = d_1 + d_2 + ...d_p$, an estimate of copula $C(u)$ can be obtained from pseudo-observations such as sample ranks using the *empirical copula estimator* $C_n(u)$ in a fairly straightforward way. First studied by Deheuvels (1979), the estimator is commonly defined as:

$$C_n(u) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathbf{1}(\hat{U}_{ij} \leq u_j) \tag{1}$$

Where $\hat{U}_{ij} = \frac{1}{n} \sum_{l=1}^{n} \mathbf{1}(X_{lj} \leq X_{ij})$, for $j \in \{1, ..., d\}$, which amounts to working with the ranks $R_{ij} = n\hat{U}_{ij}$. The scaled and weighted version of (1) gives rise to the associated *empirical copula process* $\mathbb{C}_n$, which we use to establish the limiting properties of $C_n(u)$:

$$\mathbb{C}_n = \sqrt{n}\,[C_n(u) - C(u)] \tag{2}$$

van der Vaart and Wellner (1996) establish weak convergence of (2) in $\ell^\infty([a, b]^2)$ for $0 < a < b < 1$, and Fermanian, Radulovic, and Wegkamp (2004) show weak convergence in $\ell^\infty([0, 1]^2)$. Theorem 1 of Kojadinovic and Holmes (2009), which we restate here without proof, summarizes the results of van der Vaart and Wellner (1996) and Tsukahara (2005) and shows weak convergence of (2) in $\ell^\infty([0, 1]^p)$. It is a version of Donsker-Skorohod-Kolmogorov theorem for $\ell^\infty([0, 1]^p)$, where $\ell^\infty([0, 1]^p)$ is understood to be the space of all bounded real-valued functions on $[0, 1]^d$ equipped with uniform metric.

Theorem 1. *Suppose that $C(u)$ has continuous partial derivatives. Then, the empirical copula process*

$$\mathbb{C}_n = \sqrt{n}\,[C_n(u) - C(u)]$$

converges weakly in $\ell^\infty([0, 1]^p)$ to a tight centered Gaussian process

$$g(u) = \mathcal{B}(u) - \sum_{u=1}^{d} \partial_i C(u) \mathcal{B}(1, ..., 1, u_i, 1, ..., 1),\ u \in [0, 1]^d$$

where $\mathcal{B}(u)$ is a multivariate tied-down Brownian bridge with covariance function $E[\mathcal{B}(u)\mathcal{B}(u')] = C(u \wedge u') - C(u)C(u')$.

1.2 Existing Copula-Based Tests For Independence

In this section we review the currently available non-parametric tests for independence based on the empirical copula process. Let $b_j = \sum_{j=1}^{j} d_k$, $\forall j \in S$, which a convention that $b_0 = 0$. Given a vector $u \in [0, 1]^d$ and a set $B \subset S$, define vector $u^{\{B\}} \in [0, 1]^d$ as:

$$u_i^{\{B\}} = \begin{cases} u_i & \text{if } i \in \cup_{j \in B}\{b_{j-1}+1, ..., b_j\}, \\ 1 & \text{if otherwise.} \end{cases}$$

Then, for any $k \in S$, denote the marginal copula of X_k as $C(u^{\{k\}})$, $u \in [0,1]^d$. Mutual independence between variables $X_1, ..., X_p$ occurs when $C(u) = \prod_{k=1}^{p} C(u^{\{k\}})$.

It seems natural to base tests of independence on the Kolmogorov-Smirnov or Cramer-von-Mises-type functionals of the copula distance function:

$$D(C)(u) = \left[C(u) - \prod_{k=1}^{p} C(u^{\{k\}}) \right] \tag{3}$$

where under the H_0 of independence, $D(C)(u) = 0$, $\forall u \in [0,1]^d$, and the test statistic

$$I = n \int_{[0,1]^d} D(C)(u)^2 du \tag{4}$$

is zero. The associated *independence empirical copula process* is

$$\sqrt{n} \left[C_n(u) - \prod_{j=1}^{p} C_n(u^{\{k\}}) \right] \tag{5}$$

and the empirical counterpart of (4) based on (5) is given by $I_n = n \int_{[0,1]^d} D(C_n)(u)^2 du$.

While test statistics similar to (4) can give rise to a series of powerful copula-based tests, they are not the focus of the associated literature. Deheuvels (1979) shows that under mutual independence of $X_1, ..., X_p$, the empirical process (2) can be decomposed into $2^d - d - 1$ sub-processes $\sqrt{n} M_A(C_n)$, $A \subseteq \{1, ..., d\}$, and $|A| > 1$ which converge to centered mutually independent Gaussian processes. Under mutual independence of $X_1, ..., X_p$, we have that $M_A(C)(u) = 0$, $\forall u \in [0,1]^d$ and $A \subseteq \{1, ..., d\}$. Kojadinovic and Holmes (2009), Genest and Remillard (2004), Quessy (2010) and Kojadinovic and Yan (2009) consider versions of the copula-based test of independence which are based on the $2^d - d - 1$ test-statistics of the form

$$M_{A,n} = n \int_{[0,1]^d} M_A(C_n)(u)^2 du \tag{6}$$

where the map $M_A(f)(x) : l^\infty([0,1]^d) \to l^\infty([0,1]^d)$ is given by

$$M_A(f)(x) = \sum_{B \subseteq A} (-1)^{|A|-|B|} f(x^B) \prod_{k \in A \setminus B} f(x^{\{k\}}),\ x \in [0,1]^d, \text{ and } P_s = \{B \subseteq S : |B| > 1\}. \tag{7}$$

Rejection occurs when at least one of the $M_{A,n}$'s, or some functional of $M_{A,n}$'s exceeds a critical value. Note that if any of the X_k is of size $d_k \geq 2$, the statistic M_A, n is a functional of $C_n(u^{\{k\}})$ even if $X_1, ..., X_p$ are mutually independent, and hence the test is not distribution-free. While the critical values can be tabulated for the "univariate" version of the test, in the vector case they need to be determined via bootstrap, conditional on the observed sample.

2 Proposed Test Statistic

In this work we propose a new test statistic which is based on an alternative version of the independence empirical copula process. In particular, we use a weighting function $w(u) : [0,1]^d \to \Re$ to alter the scaling of the process $C_n(u)$. To our knowledge, the issue of a weighting function in the context of non-parametric testing was not addressed in the literature. Section 2.1 reviews the weighting function in more detail. We consider the following empirical process:

$$\mathbb{W}_n = \sqrt{n}\left[\left(C_n(u) - \prod_{j=1}^{p} C_n(u)\right)\sqrt{w(u)}\right] = \sqrt{n}\left[\nu(C_n)(u) - \nu(\prod_{j=1}^{p} C_n)(u)\right] \tag{8}$$

where $\nu(f)(u)$ is a map from $l^\infty([0,1]^d)$ to $l^\infty([0,1]^d)$ given by

$$\nu(f)(x) = f(x)\sqrt{w(x)}, \text{ where } w(x) \in l^\infty([0,1]^d) \tag{9}$$

In order to study the asymptotic properties of (8), consider the map $\phi : l^\infty([0,1]^d) \to l^\infty([0,1]^d)$ given by:

$$\phi(f)(x) = \left(f(x) - \prod_{j=1}^{p} f(x^{\{k\}})\right)\sqrt{w(x)} \tag{10}$$

Lemma 1. *The map ϕ is Hadamard-differentiable tangentially to $l^\infty([0,1]^d)$ and its derivative at $f \in l^\infty([0,1]^d)$ is:*

$$\phi'_f(a)(x) = \left(a(x) - \sum_{i=1}^{p} a(x^{\{i\}}) \prod_{j=1, j\neq i}^{p} f(x^{\{j\}})\right)\sqrt{w(x)} \tag{11}$$

Proof. Let t_n, $n = 1, 2...$ be a sequence of reals converging to zero, and let a_n, $n = 1, 2...$ be a sequence of maps in $l^\infty([0,1]^d)$ converging to $a \in l^\infty([0,1]^d)$ such that $f + t_n a_n \in l^\infty([0,1]^d)$. Then,

$$\begin{aligned}
&\frac{\phi(f + t_n a_n)(x) - \phi(f)(x)}{t_n} = \\
= &\lim_{n\to\infty} \frac{\left\{f(x) + t_n a_n(x) - \prod_{j=1}^{p}\left(f(x) + t_n a_n(x^{\{j\}})\right) - f(x) + \prod_{j=1}^{p} f(x^{\{j\}})\right\}\sqrt{w(x)}}{t_n} \\
= &\; a(x)\sqrt{w(x)} \\
- &\lim_{n\to\infty}\left\{\left(\prod_{k=1}^{p} f(x^{\{k\}}) + t_n \sum_{k=1}^{p} a_n(x^{\{k\}}) \prod_{j=1, j\neq k} f(x^{\{j\}})\right.\right. \\
+ &\left.\left.\sum_{k=2}^{p-1} t_n^k \left(\sum_{l=1}^{p} f(x^{\{k\}}) \prod_{i=1, i\neq l} a_n(x^{\{i\}})\right) + t_n^p \prod_{k=1}^{p} a_n(x^{\{k\}}) - \prod_{k=1}^{p} f(x^{\{k\}})\right)/t_n\right\}\sqrt{w(x)} \\
= &\left(a(x) - \sum_{k=1}^{p} a_n(x^{\{k\}}) \prod_{j=1, j\neq k} f(x^{\{k\}})\right)\sqrt{w(x)}.
\end{aligned}$$

□

Theorem 2. *If C has continuous partial derivatives, when, when $\phi(C)(u) = 0$, $u \in [0,1]^d$, i.e. when $X_1, ..., X_p$ are mutually independent, the empirical process $\sqrt{n}\phi(C_n)(u)$, $u \in [0,1]^d$ converges weakly in $l^\infty([0,1]^d)$ to the tight centered Gaussian process*

$$\phi'_C(g)(u) = \left(g(u) - \sum_{j=1}^{p} g(u^{\{j\}}) \prod_{i=1, i\neq j}^{p} C(u^{\{i\}}) \right) \sqrt{w(u)}$$

with covariance function

$$\begin{aligned} E[\phi'_C(u)\phi'_C(v)] = \Bigg(& C(u \wedge v) - C(u)C(v) \\ & - 2\sum_{i=1}^{p} \left[C(u^{\{i\}} \wedge v^{\{i\}}) - C(u^{\{i\}})C(v^{\{i\}}) \right] \prod_{j=1, j\neq i}^{p} C(u^{\{i=j\}}) \\ & + \sum_{k=1}^{p}\sum_{l=1}^{p} \left(C(u^{\{k\}} \wedge v^{\{l\}}) - C(u^{\{k\}})C(v^{\{l\}}) \right) \prod_{i=1, i\neq k}^{p} C(u^{\{i\}}) \prod_{j=1, j\neq l}^{p} C(v^{\{j\}}) \Bigg) \sqrt{w(u)w(v)}. \end{aligned}$$

Proof. By applying the functional delta method as in van der Vaart and Wellner (1996) (Theorem 3.9.4) with Hadamard-differentiable map ϕ to the process (2) with weak limit $g(u)$ given in Theorem 1 we have that $\sqrt{n}[\phi(C_n)(u) - \phi(C)(u)] \rightsquigarrow \phi'_C(g)(u)$. The limiting covariance function is:

$$\begin{aligned} & E[\phi'_C(u)\phi'_C(v)] = \\ & \quad E\left[\left(g(u) - \sum_{j=1}^{p} g(u^{\{j\}}) \prod_{i=1, i\neq j}^{p} C(u^{\{i\}}) \right) \left(g(v) - \sum_{j=1}^{p} g(v^{\{j\}}) \prod_{i=1, i\neq j}^{p} C(v^{\{i\}}) \right) \sqrt{w(u)w(v)} \right] \\ & = \Bigg(E[g(u)g(v)] - \sum_{i=1}^{p} E[g(u)g(v^{\{i\}})] \prod_{j=1, j\neq i}^{p} C(v^{\{i\}}) - \sum_{i=1}^{p} E[g(v)g(u^{\{i\}})] \prod_{j=1, j\neq i}^{p} C(u^{\{i\}}) \\ & + \sum_{k=1}^{p}\sum_{l=1}^{p} E[g(u^{\{k\}})g(v^{\{l\}})] \prod_{i=1, i\neq k}^{p} C(u^{\{i\}}) \prod_{j=1, j\neq l}^{p} C(v^{\{j\}}) \Bigg) \sqrt{w(u)w(v)}. \end{aligned}$$

Using the result from Theorem 1 that $E[\mathcal{B}(u)\mathcal{B}(u')] = C(u \wedge u') - C(u)C(u')$ and our assumption of independence we get the desired expression for the covariance function.

□

Note that for any $w(u) \in l^\infty([0,1]^d)$ the covariance function is bounded. The weighted test statistic based on $\mathbb{W}_n$ takes the following form:

$$I^w = n \int_{[0,1]^d} \phi^2(C)(u)du = n \int_{[0,1]^d} \left[C(u) - \prod_{j=1}^{p} C(u^{\{j\}}) \right]^2 w(u)du \tag{12}$$

and it's asymptotic distribution under $H0$ of independence is non-degenerate. The sample analog of (12) is given by:

$$I_n^w = n \int_{[0,1]^d} \phi^2(C_n)(u)du = n \int_{[0,1]^d} \left[C_n(u) - \prod_{k=1}^{p} C_n(u^{\{k\}}) \right]^2 w(u)du \tag{13}$$

Corollary 1. *Suppose that C has continuous partial derivatives. Then under the null of independence between $X_1, ..., X_p$, I_n^w converges in distribution to*

$$\int_{[0,1]^d} \left[\phi'(g)(u)\right]^2 w(u)du$$

Proof. Similar to Corollary 9 of Kojadinovic and Holmes (2009), convergence is established as a direct consequence of Theorem 2 and the continuous mapping theorem. □

Clearly, $I^W = I$ if we choose $w(u) = 1$, $\forall u \in [0,1]^d$, in which case our test statistic coincides with that briefly considered in Section 3.1 of Kojadinovic and Holmes (2009). Another interesting choice of weights is $w'(u) = [f(F_1^{-1}(u_1) \times ... \times f(F_p^{-1}(u_p))]^{-1}$. Consider a classical test for independence proposed by Hoeffding (1948) and Blum, Kiefer, and Rosenblatt (1961) (HBKR) based on the distance between the joint and the product of the marginal distribution functions of $\mathbf{X}$. Assume that $d_j = 1$, $\forall j \in S$, that is, none of the X_j's is of dimension greater than one, in which case $d = p$. Then the HBKR test statistic is given by:

$$B = \int_{\Omega_1 \times ... \times \Omega_p} \left[F(X_1, ..., X_p) - F_1(X_1)F_2(X_2)...F_p(X_p)\right]^2 d\mathbf{X} \tag{14}$$

where Ω_j is the support of $F_j(X_j)$, and perform the change of variable by letting $u_j = F_j(X_j)$. We then have that $X_j = F_j^{-1}(u_j)$ and $dX_j = [f_j(F_j^{-1}(u_j)]^{-1}du_j$, where $f_j(x_j)$ is the p.d.f. of X_j, and

$$\begin{aligned} B &= \int_{\Omega_1 \times ... \times \Omega_p} \left[F(F_1^{-1}(u^{\{1\}}), ..., F_p^{-1}(u^{\{p\}})) - F_1(X_1)F_2(X_2)...F_p(X_p)\right]^2 d\mathbf{X} && (15) \\ &= \int_{\Omega_1 \times ... \times \Omega_p} \left[C(u_1, ..., u_p) - \prod_{j=1}^{p} u_j\right]^2 \prod_{j=1}^{p}[f_j(F_j^{-1}(u_j))]^{-1}du_1...du_p && (16) \\ &= \int_{\Omega_1 \times ... \times \Omega_p} \left[C(u) - \prod_{j=1}^{p} u_j\right]^2 w'(u)du && (17) \\ &= I^w && (18) \end{aligned}$$

when $w(u) = w'(u)$. At least in such "univariate" case, we can see that the classic HBKR test for independence can be though of as a weighted copula-based test with a special choice of weighting function $w'(u)$.

It seems that the only difference between the family of classic HBKR-type tests based on the distribution function, and the newer tests based on the copula function, is the choice of a weighting function.

2.1 The weighting function

Weighted metrics and the associated weighted empirical processes received much attention in the statistical and econometric literature, as they often arise in the context of parameter estimation (weighted least squares estimators, for example) and smoothing. However, to our knowledge, weighted metrics were not considered in the context of empirical copula processes, and in particular, in the context of independence testing.

While our results are valid for a broad class of bounded weighting functions $w(u) : [0,1]^d \to [0, \infty)$, we chose the following weighting function to formulate our version of the a weighted test for independence:

$$w(u) = \sum_{k=1}^{d} (u_k - u_k^2) \tag{19}$$

.

The function $w(u)$ weights each observation in the sample proportionally to the maximum variability in the geometry of the underlying copula permitted at that sample point. Since all copula functions are constrained by the upper and lower Frechet-Hoeffding bounds (Nelsen (2006)), the maximum variability in the geometry of the copula at each point is the distance between the bounds[1]. Such distance can be viewed as proxy for the maximum amount of information that each sample point can carry about dependence.

For example, the distance is maximized at the median of the joint distribution, and is zero at the boundary of copula support. Therefore, the smallest and largest sample points, carry no useful information about the dependence structure, while sample points around the median are ex ante most informative. The function $w(u)$ places higher weights to sample points around the median of the distribution, and suppresses the tails, as such, giving more weight to potentially more informative observations.

2.1.1 Closed-Form Expression

In what follows we provide a closed form expression for the integral test statistic in (13) for an arbitrary additively-separable and integrable weighting function $W(u)$. A similar derivation for an arbitrary integrable and product-separable weighting function $H(u) = \prod_{j=1}^{d} h_j(u_j)$ is available in the appendix, which can accommodate functions similar to $w'(u)$.

Theorem 3. *Let $H(u) : [0,1]^d \to [0, \infty)$ be an integrable and additively separable weighting function which can be expressed in the form $W(u) = \sum_{j=1}^{d} w_j(u_j)$. Then,*

$$
\begin{aligned}
I_n^w &= \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{d}\tilde{w}_k(\hat{u}_{ik} \vee \hat{u}_{lk}, 1) - 2\frac{1}{n^p}\sum_{i=1}^{n}\prod_{k=1}^{p}\sum_{l=1}^{n}\sum_{m=b_{k-1}+1}^{b_k}\tilde{w}(\hat{u}_{im} \vee \hat{u}_{lm}, 1) \\
&+ \frac{1}{n^{2p-1}}\prod_{k=1}^{p}\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{m=b_{k-1}+1}^{b_k}\tilde{w}(\hat{u}_{im} \vee \hat{u}_{lm}, 1).
\end{aligned}
$$

where $\tilde{w}_k(a, b)$ is the definite integral of $w_j(u_j)$ from a to b, and $\vee$ is the max() *operator.*

Proof.

$$
I_n^w = n\int_{[0,1]^d}\left(C_n(u)^2 W(u) - 2C_n(u)\prod_{k=1}^{p} C_n(u^{\{k\}})W(u) + \left[\prod_{k=1}^{p} C_n(u^{\{k\}})\right]^2 W(u)\right) du
$$

.

Integrating the first term we have

$$
\begin{aligned}
\int_{[0,1]^d} C_n(u)^2 W(u) du &= \int_{[0,1]^d}\left(\frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\mathbf{1}(\hat{u}_{ij} \le u_j)\frac{1}{n}\sum_{j=1}^{n}\prod_{m=1}^{d}\mathbf{1}(\hat{u}_{lm} \le u_m)\right) W(u)du \\
&= \int_{[0,1]^d}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\prod_{j=1}^{d}\mathbf{1}(\hat{u}_{ij} \le u_j)\mathbf{1}(\hat{u}_{lj} \le u_j)\right) W(u)du \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\int_{\hat{u}_{i1}\vee\hat{u}_{li}}^{1} \ldots \int_{\hat{u}_{id}\vee\hat{u}_{ld}}^{1} W(u)du_1, \ldots, du_d \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{d}\int_{\hat{u}_{ik}\vee\hat{u}_{lk}}^{1} w_d(u_k)du_k = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{d}\tilde{w}_k(\hat{u}_{ik} \vee \hat{u}_{lk}, 1).
\end{aligned}
$$

[1] The exact distance is given by $w'(u) = min(u_1, u_2, ..., u_d) - max(u_1 + u_2 + ... + u_d - d + 1, 0)$, however $w(u)$ assigns similar relative weights, but is easier to integrate.

For the second term we get

$$
\begin{aligned}
&\int_{[0,1]^d} C_n(u) \prod_{k=1}^{p} C_n(u^{\{k\}}) W(u) du \\
= &\int_{[0,1]^d} \frac{1}{n} \left(\sum_{i=1}^{n} \prod_{k=1}^{d} \mathbf{1}(\hat{u}_{ik} \leq u_k) \right) \left(\prod_{k=1}^{p} \frac{1}{n} \sum_{l=1}^{n} \prod_{j=1}^{d} \mathbf{1}(u_{lj} \leq u_j^{\{k\}}) \right) W(u) du \\
= &\frac{1}{n^{p+1}} \sum_{i=1}^{n} \prod_{k=1}^{p} \sum_{l=1}^{n} \sum_{m=b_{k-1}+1}^{b_k} \tilde{w}_m(\hat{u}_{im} \vee \hat{u}_{lm}, 1).
\end{aligned}
$$

Lastly, integrating the third term we have

$$
\begin{aligned}
&\int_{[0,1]^d} \prod_{k=1}^{p} C_n(u^{\{k\}})^2 W(u) du = \int_{[0,1]^d} \prod_{k=1}^{p} \left(\frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \mathbf{1}(\hat{u}_{ij} \leq u_j^{\{k\}}) \right) W(u) du \\
= &\frac{1}{n^{2p}} \prod_{k=1}^{p} \sum_{i=1}^{n} \sum_{l=1}^{n} \sum_{m=b_{k-1}+1}^{b_k} \tilde{w}_m(\hat{u}_{im} \vee \hat{u}_{lm}, 1).
\end{aligned}
$$

□

3 Power Study

To gage the power of the proposed weighted test, we construct empirical power curves for the test statistic in Theorem (3) and compare that to the power of the test proposed by Kojadinovic and Holmes (2009) performed on the same sample. We find that for broad range of alternatives (several bivariate distributions with different margins constructed using one-parameter Archimedian copulas in Gumbel, Clayton and Frank families), the weighted test possesses superior power to the test of Kojadinovic and Holmes (2009). This gain in power is not trivial.

The performance of the test may be further improved, however, by a different choice of the weighting function. At this point we only wish to demonstrate the fact that using a weighting function can improve the power of the copula-based test for independence significantly for a fairly broad range of alternatives.

Figure 1 shows power curves for Kojadinovic and Holmes (2009) unweighted test statistic (KH), our proposed test (T1), and a standard F-test of $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ in a polynomial regression model $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \epsilon$, which we also use to test the null-hypothesis of statistical independence between X and Y. To obtain the power of the three tests, we generate 10,000 bivariate random samples of size $n = 25$ each with standard normal margins and Gumbel copula for each value of the dependence parameter k. The power is the percent of rejections in 10,000 samples. Note that for dependence parameter $k = 1$, Gumbel copula reduces to independence, and the empirical power is just the Type I error.

Our proposed weighted test possesses higher power than both the unweighted test of Kojadinovic and Holmes (2009) and a standard F-test on the coefficients from polynomial regression.

Figure 2 shows the similar power curves now based on 10,000 samples of size $n = 50$, instead of $n = 25$.

We verify that we get a good control of Type I error (Table 1).

4 Applications and extensions

The test has many applications and warrants several interesting extensions.

	Type I Error
Unweighted test	0.0992
Weighted test	0.0990

Table 1: Empirical Type I error rates based on 100 samples (for test size 0.1)

4.1 Using the test with pseudo-observations

[The following section is tentative, and mostly a conjecture. Some known issues include:

- Proof of Lemma 2 is by reference to the outside source and is possibly incomplete
- Lemma 3 is without proof
- The map $\psi(H)(u, v)$ and subsequent Lemma 4 are defined for maps from $\mathbb{R}^2$, but are invoked with maps from $\mathbb{R}^d$ in Theorem 4
- Weak convergence in Lemmas 2 and 3 is shown in $D(T)$, $T \subset \mathbb{R}$ and $D(\bar{\mathbb{R}})$ respectively, but the Lemmas are invoked to show convergence in $l^\infty([0, 1]^d)$

]

In this section we adapt the test to the case where independent sample is replaced by a set of estimated quantities (pseudo-observations). This allows the test to be used in situations where some of the data are not observed directly, but inferred using an estimation procedure.

Regression residuals are one example of pseudo-observations which we consider in more detail in Subsection 4.1.1. Other examples of possible applications of this version of the test include using the estimates of realized volatility to test for a form of financial contagion, or evaluating the performance of out of sample forecasts.

4.1.1 A copula-based test using regression residuals

Consider a classical linear regression model $Y = X\beta + \epsilon$, $X, Y, \epsilon \in \mathbb{R}$, and let $Z = (Y, X)$. The lack of dependence between the regressors and model errors is a common regularity assumption imposed on the model, which, when fails, can lead to biased and inefficient estimates. A copula-based test for independence between regressors and model residuals can therefore be used to identify the presence of endogenous variables and serve as a test for omitted variables, model mis-specification, or certain forms of heteroscedasticity.

With $\epsilon = H(Z) = Y - X\beta$, let the pseudo-observations be given by model errors $\hat{\epsilon} = H_n(Z) = Y - X\hat{\beta}_n$, where $\hat{\beta}_n$ is some estimate of β. We are interested to test the $H_0 : X||\epsilon$ using the sample $(X_i, \hat{\epsilon}_i)$, $i = 1..n$, where $||$ is used to denote independence. Let K_n be the empirical distribution function of $\hat{\epsilon}_i$'s. The next lemma establishes conditions for weak convergence of the empirical process $\mathbb{K}_n(t) = \sqrt{n}\,\{K_n(t) - K(t)\}$, where K is the distribution function of ϵ.

Lemma 2. *Suppose that the law of ϵ admits density $k(t)$ which is continuous on the support $T \subset \mathbb{R}$, and suppose that $E[|X|] < \infty$. Then, if $\sqrt{(}\hat{\beta}_n - \beta) \to \beta$, then $\mathbb{K}_n(t)$ converges weakly in $D(T)$ to continuous stochastic process $\mathbb{K}(t)$ given by:*

$$\mathbb{K}(t) = B(K(t)) - \{E[X]\beta\}\, k(t) \tag{20}$$

where $B(t)$ is a Brownian bridge.

Proof. A complete proof is given in Theorem 2.1 and Hypothesis I and II of Ghoudi and Remillard (1998). □

Lemma 3. *Suppose that the conditions of Lemma 2 are satisfied, and let* $\tilde{K}_n(x,e) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le x, \hat{\epsilon}_i \le e)$ *be the empirical joint distribution function of* (X,ϵ)*. Then, the empirical process* $\tilde{\mathbb{K}}_n = \sqrt{n}\left\{\tilde{K}_n(x,e) - \tilde{K}(x,e)\right\}$, *where* $\tilde{K}(x,e)$ *is the joint distribution function of (X,e), converges weakly in* $D(\bar{\mathbb{R}})$ *to a continuous stochastic process* $d(z,e)$ *given by...*

Proof. ... □

Let the map $\psi(H)(u,v)$ from bivariate distribution functions H on $\mathbb{R}^2$ to bivariate distribution functions on $[0,1]^2$, be defined as follows:

$$\psi(H)(u,v) = H(H^{-1}(u,\infty), H^{-1}(\infty,v))$$

The following lemma (van der Vaart and Wellner (1996), Lemma 3.9.28), which we restate here without proof, establishes Hadamard-differentiability of $\psi(H)(u,v)$ at H tangentially to $C(\mathbb{R}^2)$, where $C(T)$ is the space of all continuous functions from T to $\mathbb{R}$.

Lemma 4. *Suppose that for any* p, q, *s.t.* $0 < p < q < 1$, H *is a distribution function on* $\mathbb{R}^2$ *with marginal distribution functions* F *and* G *that are continuously differentiable on the intervals* $\left[F^{-1}(p) - \delta, F^{-1}(q) + \delta\right]$ *and* $\left[G^{-1}(p) - \delta, G^{-1}(q) + \delta\right]$ *with positive derivatives* f *and* g*, for some* $\delta > 0$*. Furthermore, assume that* $\partial H/\partial x$ *and* $\partial G/\partial y$ *exist and are continuous on the product of these intervals. Then, the map* $\psi(H)(u,v)$ *is Hadamard-differentiable tangentially to* $C(\mathbb{R}^2)$ *and its derivative is given by:*

$$\begin{aligned}
\psi'_H(a)(u,v) &= a(F^{-1}(u), G^{-1}(v)) \\
&- \frac{\partial H}{\partial x}\left(F^{-1}(u), G^{-1}(v)\right)\frac{a(F^{-1}(u),\infty)}{f(F^{-1}(u))} \\
&- \frac{\partial H}{\partial y}\left(F^{-1}(u), G^{-1}(v)\right)\frac{a(\infty, G^{-1}(v))}{g(G^{-1}(v))}.
\end{aligned}$$

Let $\tilde{C}(u,v)$, $u \in [0,1]^d$, $v \in \mathbb{R}$ denote the copula of (X,ϵ), and the empirical copula of (Z,ϵ) based on pseudo-observations $\hat{\epsilon}$ be given by:

$$\tilde{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(F_n^{-1}(X_i) \le u)\mathbf{1}(G_n^{-1}(\hat{\epsilon}_i) \le v) \tag{21}$$

The next theorem establishes the weak limit of the empirical process based on (21).

Theorem 4. *Suppose that* $X_1, ..., X_n$ *are i.i.d, and* $\hat{\epsilon} = Y - X\hat{\beta}$*. Further, suppose that the hypotheses of Lemmas 2, 3 and 4 are satisfied. Then, the empirical process* $\tilde{\mathbb{C}}_n = \sqrt{n}\left\{\tilde{C}_n(u,v) - \tilde{C}(u,v)\right\}$ *converges weakly in* $l^\infty([0,1]^2)$ *to continuous stochastic process* $\zeta(u,v) = \psi'_H(d)(u,v)$*, where* ψ'_H *is given by Lemma 4, and the process* $d(x,e)$ *is given in Lemma 3.*

Proof. Since by Lemma 3, $\sqrt{n}\left\{\tilde{K}_n(x,e) - \tilde{K}(x,e)\right\} \rightsquigarrow d(z,e)$, by applying the functional Delta-method (van der Vaart and Wellner (1996), Theorem 3.9.4) combined with the result from Lemma 4 we have that $\sqrt{n}\left\{\psi'_h(\tilde{K}_n)(x,e) - \psi'_H(\tilde{K})(x,e)\right\} = \sqrt{n}\left\{\tilde{K}_n(F_n^{-1}(u), G_n^{-1}(v)) - \tilde{K}(F^{-1}(u), G^{-1}(v))\right\} = \sqrt{n}\left\{\tilde{C}_n(u,v) - \tilde{C}(u,v)\right\} \rightsquigarrow \psi'_H(d)(u,v)$. □

When the dimension of X is greater than one, independence under H_0 is characterized by $\tilde{C}(u,v) = \tilde{C}_x(u)v$, where $\tilde{C}_x(u)$ is the copula of X. The next theorem establishes the limiting properties of the empirical process $\tilde{\mathbb{C}}_n^I$ given by

$$\tilde{\mathbb{C}}_n^I = \sqrt{n}\left\{\left(\tilde{C}_n(u,v) - \tilde{C}_{n,x}(u)v\right)\sqrt{w(u,v)}\right\} \tag{22}$$

under H_0, where $\tilde{C}_{n,x}(u)$ is the empirical copula of X.

Theorem 5. *Suppose that $X_1, ..., X_n$ are i.i.d, and $\hat{\epsilon} = Y - X\hat{\beta}$. Further, suppose that the hypotheses of Lemmas 2, 3 and 4 are satisfied, and that X is independent from ϵ. Then, empirical process $\tilde{\mathbb{C}}_n^I$ converges weakly in $l^\infty([0,1]^{d+1})$ to a continuous Gaussian process $\varsigma(u,v)$ with covariance function given by $E[\varsigma(u,v)\varsigma(u',v')] = ...$*

Proof. Apply $\phi(a)(x)$ to $\tilde{\mathbb{C}}_n$ to get

$$\begin{aligned} & \sqrt{n}\left\{\phi(\tilde{C}_n)(u,v) - \phi(\tilde{C})(u,v)\right\} = \\ = \ & \sqrt{n}\left\{\left(\tilde{C}_n(u,v) - \tilde{C}_{x,n}(u)v\right)\sqrt{w(u,v)} - \left(\tilde{C}(u,v) - \tilde{C}_x(u)v\right)\sqrt{w(u,v)}\right\} \\ = \ & \sqrt{n}\left\{\left(\tilde{C}_n(u,v) - \tilde{C}_{x,n}(u)v\right)\sqrt{w(u,v)}\right\} \end{aligned}$$

under H_0. Applying the Delta-method once more (van der Vaart and Wellner (1996), Theorem 3.9.4) we get that $\tilde{\mathbb{C}}_n^I \rightsquigarrow \phi'_{\tilde{C}_n}(\zeta)(u,v) = \varsigma(u,v)$, with covariance function given by $E[\varsigma(u,v)\varsigma(u',v')] =$
$E[\phi'_{\tilde{C}_n}(\zeta)(u,v)\phi'_{\tilde{C}_n}(\zeta)(u',v')]$...

□

Corollary 2. *If $\tilde{C}$ has continuous partial derivatives, the statistic given by*

$$\tilde{I}_n^w = n\int_{[0,1]^{d+1}}\left(\tilde{C}_n(u,v) - \tilde{C}_{n,x}(u)v\right)^2 w(u,v)dudv$$

converges in distribution to

$$\int_{[0,1]^{d+1}}\left(\phi'_{\tilde{C}_n}(\varsigma)(t)\right)^2 dt$$

Proof. The result follows by Theorem 5 and the continuous mapping theorem. □

Note that as with the i.i.d. case, if X is of dimension greater than one, the test is not distribution-free, and hence we use the permutation approach to obtain the approximate p-values. The closed form given in Theorem 3 can, however, be used to find the exact value of $\tilde{I}_n^w$.

4.1.2 Copula-based linear heteroscedasticity test: Monte-Carlo study

In what follows we run a simulation to compare the power of our test at detecting heteroscedasticity in model residuals and compare it to the power of the classic test of White (1980).

To obtain test power, we generate $n = 10,000$ bivariate heteroscedastic samples (X, Y) of size $s = 25$, 50, 100, 200 300 each, such that $X_i \sim N(\mu_x, \sigma_x^2)$, $\epsilon_i \sim N(0, \sigma_i^2(X_i))$, and $Y_i = \alpha + \beta X_i + \epsilon_i$, where the variance of the residuals is not constant and varies according to the values of X_i: $\sigma_i^2(X_i) = \alpha_0 + \beta_0 X_i$. As in White (1980), we then test for $H_0 : \beta_0 = 0$ by first obtaining fitted residuals $\hat{\epsilon}_i$ from the regression of Y on X, and then by estimating the following equation: $\hat{\epsilon}_i^2 = \gamma_0 + \gamma_1 X_i + \gamma_2 X_i^2 + u_i$. The test statistic is then sR^2, which is $\chi^2(2)$ asymptotically. Alternatively, we use the proposed weighted copula test to check for dependence between $\hat{\epsilon}_i^2$ and X_i. We choose the following parameters for our simulation: $\{\mu_x = 5, \sigma_x^2 = 1, \alpha = 0.5, \beta = 3, \alpha_0 = 1, \beta_0 = 0.5\}$. Figure 3 shows the scatter plot of a generated sample of X and Y.

Figure 4 shows the power of each test for each of the sample sizes considered. For each sample size, test power is the percent of correct rejections in $10,000$ samples. Copula-based tests possess considerably higher power than the parametric test of White (1980). While both tests achieve full power for samples of size 400 and greater, power advantage of the copula-based test may be preserved in large samples if the nature of heteroscedasticity is less pronounced, or is non-linear.

4.1.3 Copula-based heteroscedasticity test: Empirical example

We use our test and that of White (1980) to detect heteroscedasticity in a dataset provided by Ramanathan (2002), which reports state-level percentages of females with high school education or higher, and respective median female earnings, for a total of 50 states. Figure 5 shows the scatter plot of education propensities against earnings in the sample. Schooling is positively related to earnings, and states with lower education propensities have smaller earning variances, perhaps due to more competitive female labor markets.

We find that the White (1980) is unable to reject the null of constant variance at 5% level of significance in this sample (p-value of 0.055). Our weighted copula-based test rejects homoscedasticity at below 1% level of significance in this sample (p-value of 0.008). Such drastic reduction in the probability of Type I error is consistent with our Monte-Carlo results which show higher power of Copula-based test, especially in smaller samples.

4.2 Financial contagion and cross-market linkages

4.3 Testing for serial dependence in time series

References

Blum, J., Kiefer, J., and Rosenblatt, M. "Distribution free tests of independence based on the sample distribution function." *Annals of Mathematical Statistics*, 32:485–498 (1961).

Deheuvels, Paul. "La fonction de dependance empirique et se propriers: Un test nonparametrique dindependance." *Academie Royale de Belgique. Bulletin de la Classe des Sciences*, 65:274–292 (1979).

Fermanian, Jean-David, Radulovic, Dragan, and Wegkamp, Marten. "Weak convergence of empirical copula processes." *Bernoulli*, 10(5):847–860 (2004).

Genest, Christian and Remillard, Bruno. "Tests of independence and randomness based on the empirical copula process." *Test (Sociedad de Estadistica e I vestigacidn Operativa)*, 13(2):335–369 (2004).

Ghoudi, Kilani and Remillard, Bruno. "Empirical processes based on pseudo-observations." In B Szyskowitz, editor, "Proceedings of the International Conference on Asymptotic Methods in Statistics," Elsevier, Amsterdam (1998).

Hoeffding, Wassily. "A nonparametric test of independence." *Annals of Mathematical Statistics*, 19:546–557 (1948).

Kojadinovic, Ivan and Holmes, Mark. "Tests of independence among continuous random vectors based on cramer-von mises functionals of the empirical copula process." *Journal of Multivariate Analysis*, 100:1137–1154 (2009).

Kojadinovic, Ivan and Yan, Jun. "Tests of serial independence for continuous multivariate time series based on a mobius decomposition of the independence empirical copula process." *Annals of the Institute of Statistical Mathematics*, 10 (2009).

Medovikov, Ivan. "Serial dependence function: Copula-based non-parametric analysis of non-linear time series." *Working paper* (2011).

Nelsen, Roger B. *An Introduction To Copulas*. Springer Science, 233 Springer Street, New York, USA (2006).

Quessy, Jean-Francois. "Applicationsandasymptoticpowerofmarginal-freetestsof stochasticvectorialindependence." *Journal ofStatisticalPlanningandInference*, 140:3058–3075 (2010).

Ramanathan, Ramu. *Introductory Econometrics With Applications*. Michael Roche, Mason, Ohio (2002).

Sklar, A. "Fonctions de repartition a n dimensions et leurs marges." *Publications de l'Institut de Statistique de l'Universite de Paris*, 8:229–231 (1959).

Tsukahara, Hideatsu. "Semiparametric estimation in copula models." *The Canadian Journal of Statistics*, 33(357-375) (2005).

van der Vaart, Aad and Wellner, Jon. *Weak Convergence and Empirical Processes*. Springer-Verlag, New York (1996).

White, Halbert. "A heteroscedasticity-consistent covariance matrix and a direct test for heteroscedasticity." *Econometrica*, 48:817–838 (1980).

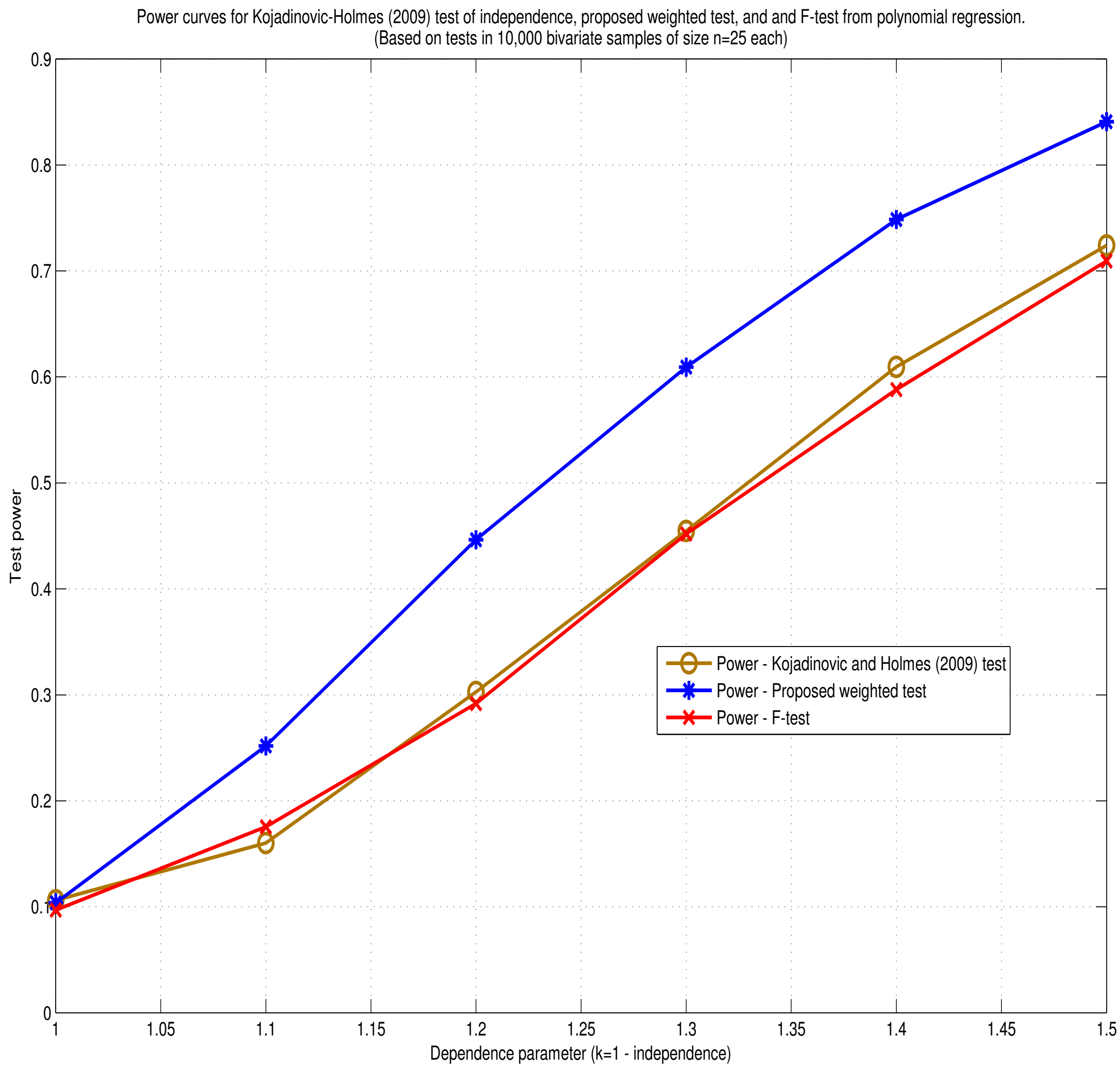

Figure 1: Power curves

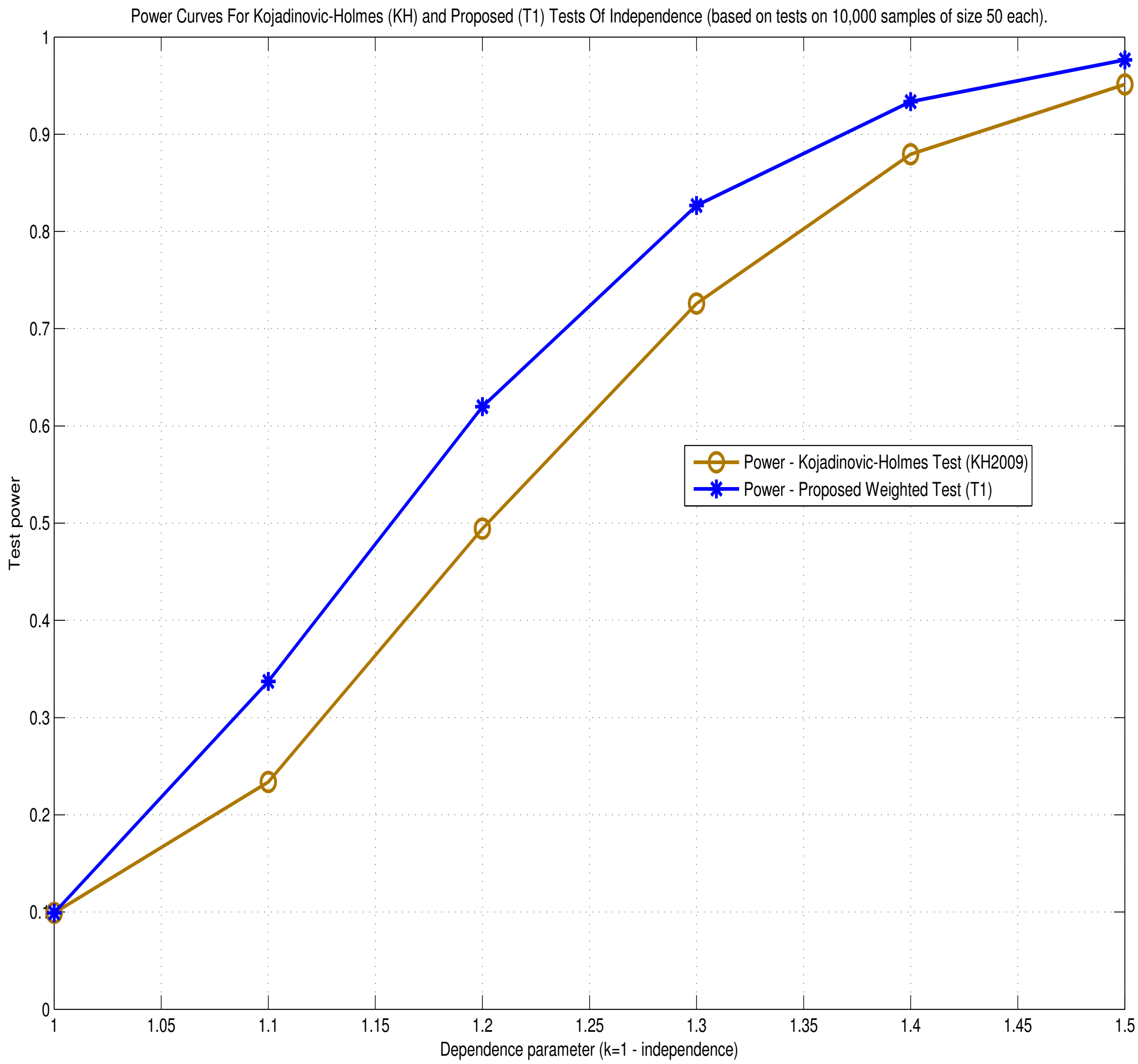

Figure 2: Power curves

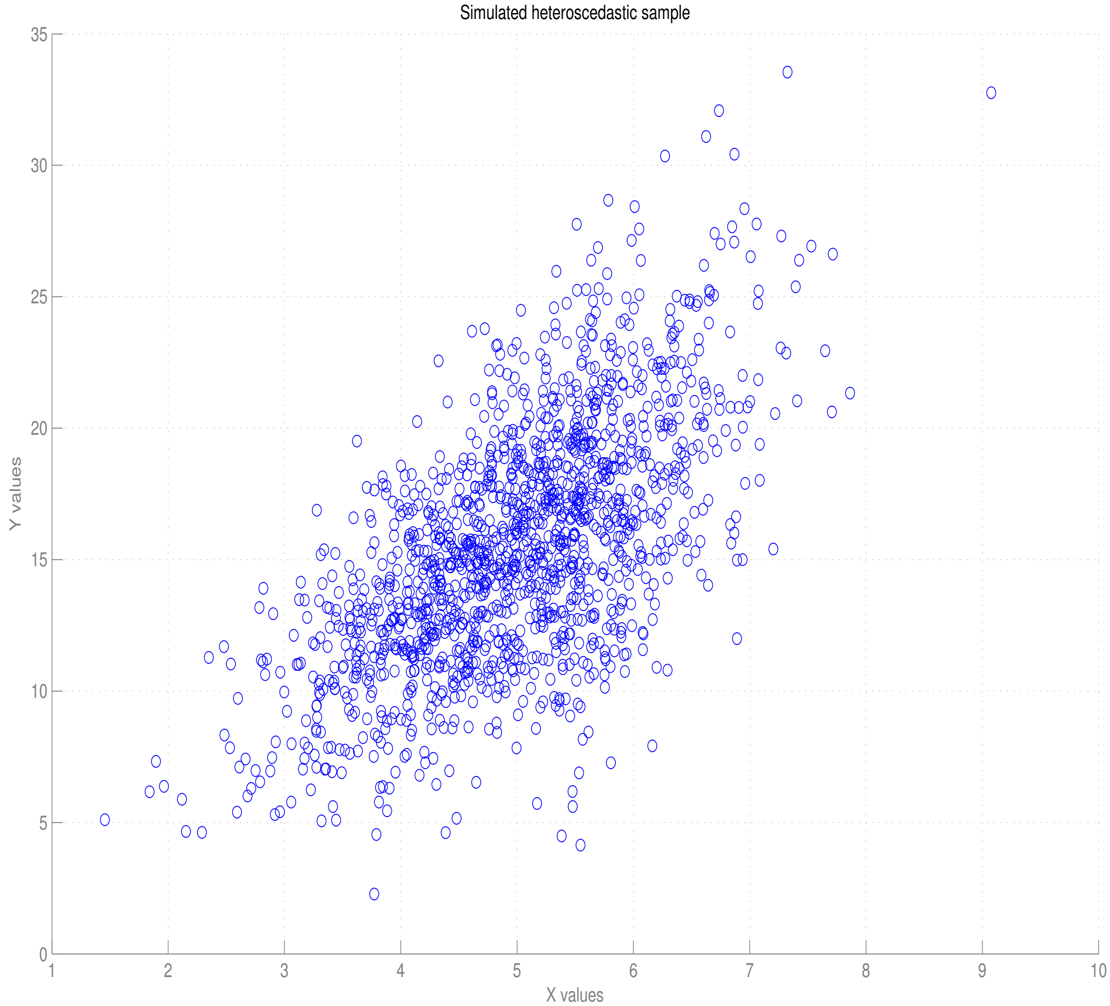

Figure 3: Power curves

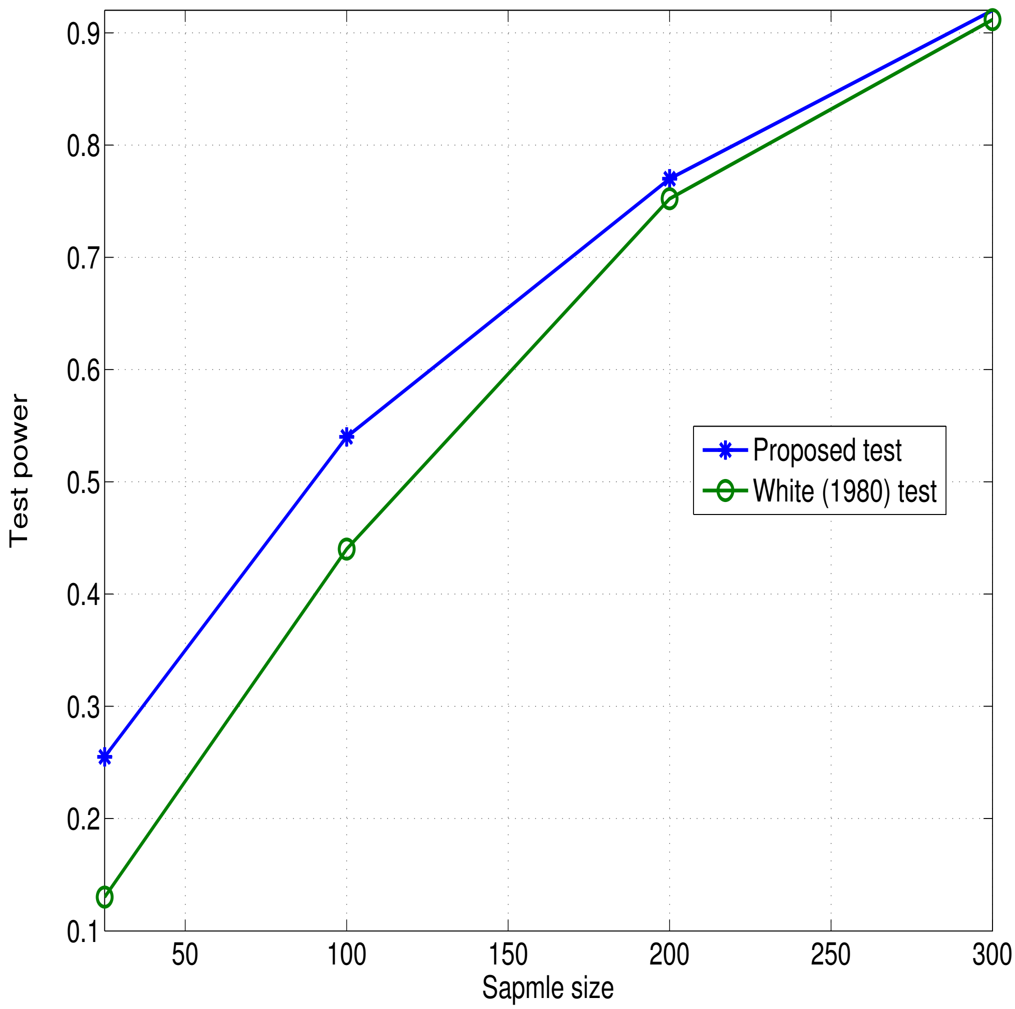

Figure 4: Power curves

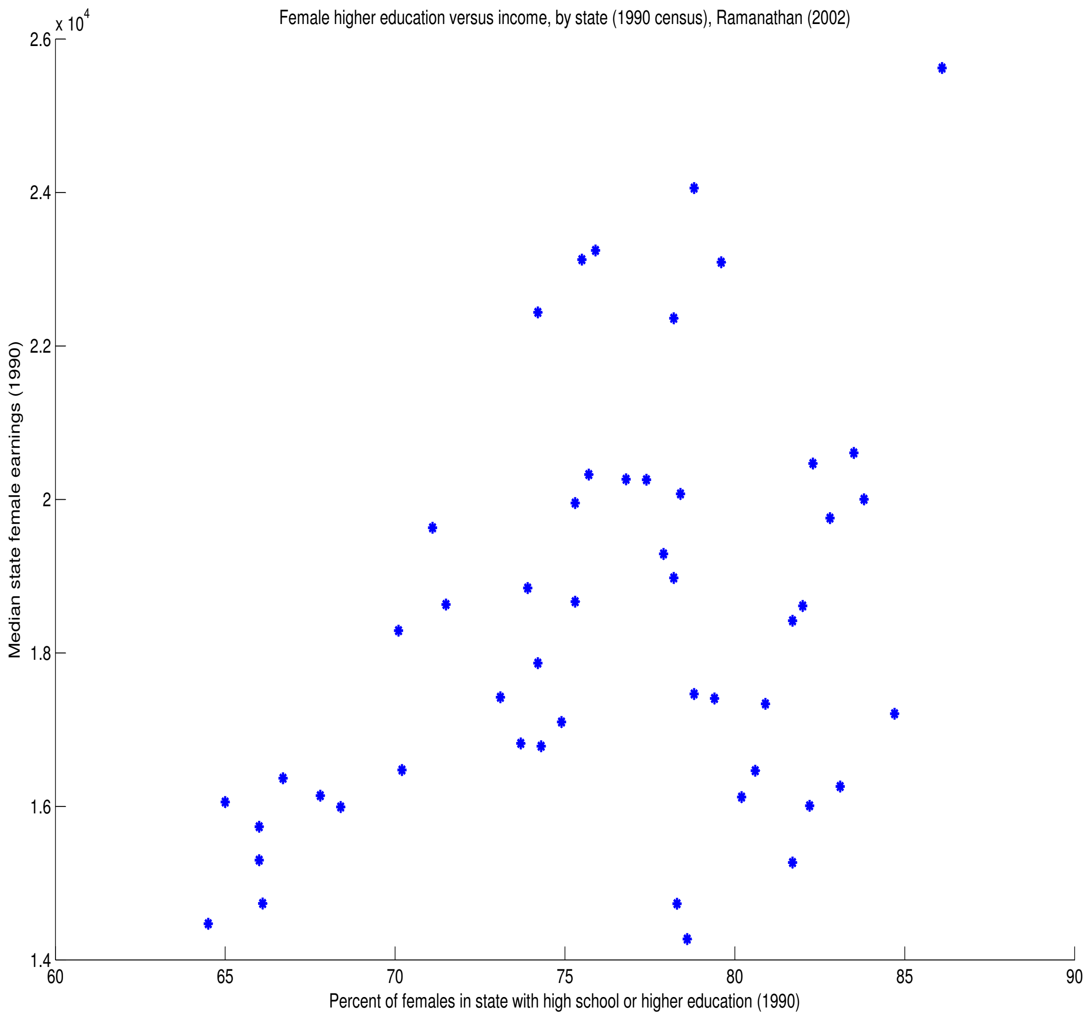

Figure 5: Power curves