University of
St Andrews

School of Economics & Finance Discussion Papers

A Consistent Nonparametric Bootstrap Test of Exogeneity

Jinhyun Lee

School of Economics & Finance
Discussion Paper Series
issn 0962-4031
http://ideas.repec.org/s/san/wpecon.html
info: econ@st-andrews.ac.uk

School of Economics and Finance Discussion Paper No. 1316

JEL Classification:
Keywords:

A Consistent Nonparametric Bootstrap Test of Exogeneity

Jinhyun Lee*
University of St Andrews

September 2013

Abstract

This paper proposes a novel way of testing exogeneity of an explanatory variable without any parametric assumptions in the presence of a "conditional" instrumental variable. A testable implication is derived that if an explanatory variable is endogenous, the conditional distribution of the outcome given the endogenous variable is not independent of its instrumental variable(s). The test rejects the null hypothesis with probability one if the explanatory variable is endogenous and it detects alternatives converging to the null at a rate $n^{-1/2}$. We propose a consistent nonparametric bootstrap test to implement this testable implication. We show that the proposed bootstrap test can be asymptotically justified in the sense that it produces asymptotically correct size under the null of exogeneity, and it has unit power asymptotically. Our nonparametric test can be applied to the cases in which the outcome is generated by an additively non-separable structural relation or in which the outcome is discrete, which has not been studied in the literature.

1 Introduction

In many economic examples measuring causality between economic variables is of major interest. Difficulties in measuring causal effects using data arise because of

*This paper is an extention of one of the chapters of my Ph D thesis. I am grateful to Andrew Chesher for his advice and discussion. Special thanks go to Adam Rosen for valuable comments and suggestions. I also thank Russell Davidson, Joel Horowitz, Sokbae Simon Lee, Arthur Lewbel, Werner Ploberger, and Yoon Jae Whang and other participants in 2009 Netherlands Econometric Study Group (NESG), 2009 Econometric Study Group(ESG) annual conference in Bristol, and the 2010 Econometric Society World Congress in Shanghai, China, 2011 RES conference in Royal Holloway in London, and SIRE Econometric Workshop in St Andrews. The support from SIRE for 2011 RES conference presentation is gratefully appreciated. All errors are my own.

the possible presence of *endogeneity or selection problem*. Endogenous variables in *economic* models are defined as variables that are determined by the model. On the other hand, the distinction between endogenous and exogenous variables in *econometric*s follows the tradition of the classical simultaneous equations models.[1] Such distinction would be required if data anaysis is *structural,* in other words, if individuals' socio-economic variables (decisions) are analyzed based on a theoretical economic model.

In this paper I refer to this attempt of analyzing data based on an economic model as a "structural analysis". This paper adopts Hurwicz's (1950) framework called "structure". Suppose we are interested in the impact of a variable (Y) chosen by individuals on their outcome (W) of interest, and suppose the economic process of W can be described by the following relation[2]

$$W = h(Y, X, U), \tag{SR}$$

where X is a vector of characteristics that are exogenously given to individuals such as age, gender, and race, and U is unobserved individual characteristics. Even among the individuals with the same observed characterisics we observe a distribution of the outcome due to the unobserved elements, U. Causal effects of a variable indicate the effects of the variable only, separated from other possible influences.

The conditional distribution of the outcome, $F_{W|YX}$, is determined by the *interaction*, indicated by the following Hurwicz Relation (HR), between the distribution of the unobserved elements, $F_{U|YX}$ and the structural relation, $h(\cdot, \cdot, \cdot)$

$$\begin{aligned} F_{W|YX}(w|y,x) &= \Pr[W \leq w | Y = y, X = x] \\ &= \Pr[h(Y,X,U) \leq w | Y = y, X = x] \\ \underbrace{\qquad\qquad}_{\text{"}Data\text{"}} &= \underbrace{\int_{\{u:h(y,x,u)\leq w\}} dF_{U|YX}(u|y,x)}_{Hurwicz\ Structure} \end{aligned} \tag{HR}$$

[1] Endogenous variables are defined in the classical analysis of simultaneous equations system as those that are determined by the system and are *not* independent of the unobserved elements in the system. Following Koopmans (1949, p133), endogenous variables are "observed variables which are not known, or assumed to be statistically *not* independent of the latent variables, and whose occurrence in one or more equations of the set of equations is necessary on grounds of "theory"".

[2] The structural relation may be derived from some optimization processes such as demand/supply functions. We are agnostic about this. If there is not a well-defined economic theory behind them, then the structural relations can be simply understood as how the outcome and the choice are determined by other relevant (both observable and unobservable) variables. The structural relation delivers the information on "contingent" plans of choice or outcome when different values of X and U are given.

The two components, $h(\cdot,\cdot,\cdot)$ and $F_{U|YX}$, are called the Hurwicz (1950) structure. The particular structural feature of interest in this paper is the distribution of the unobserved variables which contains the information of endogeneity of any explanatory variables in the economic system.[3] To deal with this hidden information we derive a "refutable" expression of $F_{U|YX}$. It is shown that if an explanatory variable, Y, is exogenous, then the distribution function of the outcome is independent of an IV,Z, conditional on the explanatory variable, that is $W \perp Z|Y$. Notably, only three variables (W, Y, Z) are used in constructing the nonparametric test statistic although there can exist other variables, X, appear in the structural relation, h.

In many economic examples, irrespective of whether it is structural or not, measuring causality between economic variables is of major interest. Difficulties in measuring causal effects using data arise because of the possible presence of *endogeneity or selection problem.* Following the tradition this paper defines endogeneity of an observed explanatory variable as that which is not independent of the unobserved elements in an equation to be estimated using data. Whether an explanatory variable is endogenous or not matters if the causal effect of that variable is the focus of data anaysis. Depending on whether there exists endogeneity or not identification and estimation strategies are determined. So far the nonparametric structural literature has more focused on how to identify and make inferences allowing for endogeneity. However, not only do identification and inference procedures under endogeneity involve more steps, but also allowing for endogeneity when the variable is actually exogenous may result in *efficiency loss.*[4] Therefore, evidence regarding the presence of *endogeneity* would be informative in directing identification and inference methods. If one can "statistically" be sure of exogeneity of an explanatory variable, it could guarantee simpler and more precise inference procedures.

Based on this testable implication we propose a nonparametric bootstrap test of exogeneity. The test statistic is shown to be asymptotically Gaussian under the null hypothesis of exogeneity. The test rejects the null hypothesis with probability one if the explanatory variable is endogenous and it detects alternatives converging to the null at a rate $n^{-1/2}$. However, the asymptotic distribution depends on the unknown features of the underlying distribution, therefore, we propose to use the bootstrap to find out the critical values. It is shown that this bootstrap test can be asymptotically justified in the sense that it produces asymptotically correct size under the null of exogeneity, and it has unit power asymptotically.

[3] The same information on the joint distribution of the unobservable and explanatory variables is contained in the conditional distribution of the unobservable variable given the explanatory variable when the marginal distribution of the explanatory variable is known. Thus, we focus on the conditional distribution.

[4] This fact is well known in the OLS and 2SLS context. This is also true in the quantile-based control function approach (QCFA) in Chesher (2003).

1.1 Related Studies and Contributions

The Durbin (1954)-Wu (1973)-Hausman (1978) type tests have been widely used to test endogeneity which are based on the difference between two estimators, one of which is more efficient under the null. There have been several tests for endogeneity when the outcome is censored or discrete. Grogger (1990) considered Durbin-Wu-Hausman type tests for endogeneity for binary and count data by comparing the nonlinear least squares estimator with a maximum likelihood estimator. Under a normality assumption and triangular structure Chesher (1985) considers testing whether the covariance between the unobserved variables is zero using the score test. Blundell and Smith (1986) considers a test of exogeneity under the simultaneous equations model for censored outcome. Rivers and Vuong (1988), and Vella (1993) considered tests of endogeneity in models with binary or censored outcomes. However, these tests are based on either linearity assumptions on functional form or Normality assumptions on the error terms. Thus the results are open to specification errors.

The first nonparametric approach is by Blundell and Horowitz (2007). They developed a test of exogeneity under the nonparametric functional form that is identified by mean independence of instrumental variables. However, this test would not be applicable to discrete or censored outcomes which are generated by additively nonseparable structural relations. The importance of using nonseparable structural functions have been emphasized recently. For example, in order to allow for heterogeneity in causal effects among observationally identical individuals, the structural function needs to be nonseparable, or to allow for discrete outcomes, use of additively nonseparable structural function is required. (see Angrist (2001) and Hahn and Ridder (2011)).

Our test should be distinguished from omitted variable/significance tests of a variable. The tests for omitted variables proposed in the nonparametric[5] setup so far (see Gozalo (1993), Fan and Li (1996), Lavergne and Vuong (2000), Delgado and Gonzalez-Manteiga (2001), and Ait-Sahalia, Bickel, and Stoker (2001)) have not considered the possibility of endogeneity. Our results imply that if an IV is detected as an omitted variable we should conclude that this is because there exists an endogeneity problem. Since it is assumed that the exclusion restriction of the IV is satisfied (exogeneity of an IV is satisfied), the IV should not be considered as an omitted variable although the omitted variable test concludes IV is significant. This shows that blindly applying an omitted variable test would lead to a misleading conclusion. "Economic" arguments would guide econometricians about which variables to include[6]. Therefore, our test would be appropriate to a case in which endogeneity

[5]See also other studies based on moment equality condition, Bierens (1982,1990), Lewbel (1995), Chen and Fan (1999), and Ait-Sahalia, Bickel, and Stoker (2001), for example.

[6]I suppose this is why in economics the structural approach should be preferred over "path analysis". What econometricians are interested in is often the sensitivity of one specific variable to

of an explanatory variable is likely to be present and when a well-defined IV exists and exclusion restriction is not controversial as in regression discontinuity designs, or natural experiments,.

There have been several studies on testing conditional independence : see Gozalo and Linton (1997), Su and White (2007,2008) and Song (2009). These existing conditional independence tests could also be used for testing exogeneity. We propose an alternative computationally intensive, but easy to implement, bootstrap test. In contrast with Gozalo and Linton (1997) which use the empirical distribution of the conditional distribution, we use conditional moment conditions to construct the test statistic, which is a multiple of indicator functions and bounded functions. Su and White (2007, 2008) can be applied to testing exogeneity with discrete outcomes. Song (2009) also proposed a conditional Kolmogorov-Smirnov type test, Song (2009) used a Rosenblatt transformation to yield an asymptotically pivotal test statistic.

The major contribution of this study is to provide a way to test exogeneity without parametric assumptions that can be applied to discrete variables. In contrast with Blundell and Horowitz (2007)'s nonparametric test, the proposed test can be used to test when the outcome is generated by a nonseparable structural function. As Hahn and Ridder (2009) discuss, when the structural relation is additively nonseparable, conditional moment restrictions do not have any identifying power of the average structural functions (ASF) defined by Blundell and Powell (2003). To the best of our knowledge this paper is the first study that considers a test of endogeneity that can be applied to nonseparable structural structural functions. Moreover, the test involves only three variables in constructing the test statistic which allows us not to concern about the typical curse of dimensionality problem in nonparametric analysis.

The proposed test is a new consistent nonparametric bootstrap test for conditional independence. This test can also be used as significance/omitted variable tests if exogeneity of all the variables is confirmed in some sense when the outcome of interest is discrete or censored. Delgado and Gonzalez-Manteiga's (2001) siginificance test can be modified to test conditional independence as is mentioned in their paper. The statistic I derive is simpler than Delgado and Gonzalez-Manteiga (2001) or Song (2009) due to use of the symmetry of the "kernel" function of a U-statistic. The proposed bootstrap strategy is distinct from Delgado and Gonzalez-Manteiga (2001). In contrast with Delgado and Gonzalez-Manteiga (2001) the results hold under weaker assumptions and do not require differentiability, thus, our test can be applied to a wider range of examples including non-differentiabl functions.

another specific variable, rather than what "factors" affect the outcome of interest. We do consider other factors that affect the outcome to measure the partial effect of the variable of interest better, not to find out what other factors are relevant.

1.2 Organization of the paper

Section 2 derives a testable implication of exogeneity of a regressor under the general nonparametric Hurwicz's (1950) structure. Section 3 discusses the construction of the test statistic based on the testable implication derived in Section2. Section 4 reports the asymptotic results of the test, and Section 5 proposes a bootstrap procedure to obtain critical values. The bootstrap procedure is justified in Section 5. Section 6 illustrate the proposed idea by simulation and by using Vietnam-era veteran data. Section 7 concludes.

2 Endogeneity and "Testability" of Exogeneity using Conditional Independence

This section provides a framework of discussion and derives a testable implication of "exogeneity" of a regressor based on the framework. Since exogeneity/endogeneity of a explanatory variable is defined by independence/dependence between the explanatory variable and other unobserved elements which are supposed to determinants of the outcome of interest. Which variables need to be considered to be determinants of the outcome may be guided by economic arguments. Economic models usually generate relations between variable by certain optimization processes. A relationship between variables generated by an economic model is called structural relation and following Hurwicz (1950) a structure is formally defined as a tuple of structural relations and the distribution of the unobserved elements as follows.

2.1 Local Identification of the Endogeneity Bias under Differentiability

The variables W and Y are discrete, continuous, or mixed discrete continuous random variable. The variable $X = \{X_k\}_{k=1}^{K}$ is a vector of exogeneous covariates. A vector of latent variates, U is jointly continuously distributed. We assume that the observed outcome, W, is determined by a structural relation, (SR), introduced in Section 1.Y is endogenous if it is not exogenous. An additively nonseparable structural function requires full independence for identification of the structural function, thus, we define exogeneity using full independence[7].

[7] The definition of endogeneity is related with the identification strategy. Whether the structural relation is assumed to be additively separable or not influences what type of restrictions are required to identify the causal effects. For example, with nonparametric structural function with additively separable error, existence of IVs that are mean independent of the regressors will be enough for identification (Newey and Powell (2003), and Newey, Powell, and Vella (1999)), whereas, when we allow for additively nonseparable errors, full independence of IV is required (Matzkin(2003), Chesher (2003), Imbens and Newey (2009), Chernozhukov and Hansen (2005), Chesher (2010) etc.)

Definition 1 Exogeneity of Y : Y is called an exogenous variable if $Y \perp U$.

Suppose that the structural function, h, is differentiable and that we are interested in the causal effects of a continuous endogenous variable, Y, defined by the partial derivative of h, $\nabla_y h$. Suppose also that U and Y are related by $U = g(Y)$. Specifying a structural relation reveals different routes of change caused by Y. This can be seen by differentiating (SR) :

$$\underbrace{\Delta_y W}_{\text{Observed change in } W \text{ due to } Y} = \nabla_y h + \underbrace{\nabla_u h \cdot \nabla_y g}_{\text{Indirect effect through } U}.$$

When Y is not independent of the unobservable variable, the observed change in W due to the change in Y, indicated by $\nabla_y W$, could be caused by two sources - the direct effect of Y on $h(\cdot,\cdot,\cdot)$ and the indirect effect of Y on $h(\cdot,\cdot,\cdot)$ through the effect of U on h. If one could identify the indirect effect, then $\nabla_y h$ can be identified by subtracting the indirect effect from the observed change in W. Note that since the indirect effect would be zero if there is no endogeneity, the indirect effect is called endogeneity bias.

Chesher (2003) derived an identifying relation for the causal effect, measured by the partial derivative, $\nabla_y h$, under a triangular structure, with $Y = t(Z, V)$,

$$\begin{aligned} \underbrace{\nabla_u h \cdot \nabla_y g}_{\text{Endogeneity Bias}} &= -\frac{\nabla_z Q_{W|YXZ}(\tau_U|y,x,z)}{\nabla_z Q_{Y|XZ}(\tau_V|x,z)}, \\ \text{if} \quad \nabla_z Q_{Y|XZ}(\tau_V|x,z) &\neq 0, \end{aligned} \tag{Bias}$$

by imposing restrictions that h is monotonic in scalar U and there exists an IV, Z, which is excluded in the structural relation, h, but is a determinant of the endogenous variable Y. Under the triangular structure, $U = g(Y) = g(t(Z, V))$. Note that if Y were exogeneous, that is, $\nabla_y g = 0$, the endogeneity bias would be zero. Thus, if $\nabla_z Q_{W|YXZ}(\tau_U|y,x,z) = 0$, exogeneity of Y locally at $Y = y$ and $X = x$, would be confirmed. If $\nabla_z Q_{W|YXZ}(\tau_U|y,x,z) = 0$, for all y, x, z in the support and for all $\tau_U \in (0,1)$, then exogeneity of Y, could be confirmed globally. However, this argument requires *differentiability* of $Q_{W|YXZ}(\tau_U|y,x,z)$, which does not hold with discrete variables, for example.

This testable implication is derived under *differentiability* of $h(\cdot,\cdot,\cdot)$. In the next subsection, a more general *refutable* implication of exogeneity is derived without relying on differentiability nor on other restrictions.

2.2 Exogeneity and Conditional Independence

We assume that there exist at least one "*conditional* instrumental variable".

Restriction C-IV (Existence of "conditional" IV) : There exists a variable Z such that (i) $U \perp Z \mid Y$, and (ii) $Y = \theta(Z, \Delta)$, where Δ is a vector of relevant variables including both observable and unobservable variables.

Discussion on Restriction C-IV

- Condition (i) implies that once the value of Y is known, the distribution of U needs to be independent of Z. For example, in Angrist (1990), if the distribution of the unobserved earnings potential (U) were independent of the draft lottery number, for each group of veterans and non-veterans, then the condition would hold.
- Note that the usual IV assumption is $U \perp Z$. As is well known (see Dawid (1979), for example) there is no relation between conditional independence and independence. The same is true for the relation between control function assumption and the single-equation IV assumption : neither implies the other. Moreover, the fact that an IV is used under the single-equation IV model does not mean that the same IV cannot be used in the control function approach since whether the IV satisfies conditional independence or marginal independence is not directly testable as U is not observable.
- This condition involves an unobserved element, thus, whether this condition is satisfied or not is *not* directly testable, while economically *arguable*.
- The second condition (ii) is the usual exclusion restriction for an IV.

We adopt the definition of conditional independence by Dawid (1979).

Definition 2 Conditional independence (Dawid (1979)) : X and Z are independent conditional on Y if $F_{X|YZ}(x|y,z) = H(x,y)$, for all x, y, z, for some function H.

We first report a "refutable" implication when Y is exogenous.

Theorem 1 We assume (SR) and (HR) introduced in Section 1. Under Definition 1 and Restriction C-IV, if Y is exogenous, then the distribution of W is independent of Z conditional on Y.

Proof. For some functions H, and H_Δ where Δ is defined previously in Restriction C-IV,

$$\begin{aligned}
F_{W|YXZ}(w|y,z) &= \Pr[W \leq w|Y=y, Z=z] \\
&= \Pr[h(Y,X,U) \leq w|Y=y, Z=z] \\
(*) \quad &= \int_{\{u:h(y,x,u)\leq w\}} dF_{U|YZ}(u|y,z) \\
&= \int_{\{u:h(y,x,u)\leq w\}} dF_{U|Y}(u|y) \\
&= \left\{ \begin{array}{ll} \int_{\{u:h(y,x,u)\leq w\}} dF_U(u) & \text{if} \quad U \perp Y \\ \int_{\{u:h(y,x,u)\leq w\}} dF_{U|YZ}(u|\theta(z,\Delta),z) & \text{o.w} \end{array} \right\} \\
(**) \quad &= \left\{ \begin{array}{ll} H(w,y) & \text{if} \quad U \perp Y \\ H_\Delta(w,y,z;\Delta) & \text{o.w} \end{array} \right\},
\end{aligned}$$

where the second equality follows from Restriction (SR) and the fourth equality is due to Restriction C-IV. Thus, we conclude that if $U \perp Y$ (Y is exogeneous), then $F_{W|YXZ}(w|y,x,z) = H(w,y,x)$ (that is, W is conditionally independent of Z given Y). ■

Notice that X, a vector of exogenous variables, does not appear in the final form. This fact facilitates the use of nonparametric methods significantly. More discussion regarding this testable implications follows in the next subsection.

2.2.1 Discussion

1. Although it is clear this holds under a triangular system, it also holds without assuming triangularity.

   (a) Single equation - IV models use $F_{W|Z}$, while control function approaches use the information from $F_{W|YZ}$. As is noted the fact that an IV is used under the single-equation IV model does not mean that the same IV can not be used in the control function approach. The same IV often is used in the two distinct identification and estimation strategy.

   (b) Simultaneity : We allow for bi-directional simultaneity in the sense that Δ can include W. Although we specify the structural relations as (SR) the test does not involve the estimation of the structural relation.

2. Vector unobservables : Note that the unobserved variable can be a vector. However, with multi-dimensional unobserved heterogeneity identification of the distribution of the unobserved variables is not achievable. (*) shows that it is impossible to identify $\{h, F_{U|Y}\}$ separately without further restrictions, but the refutable implication can still be derived.

3. This result holds as long as Restriction C-IV holds. Weak instruments would have an impact on step (**). If the instrument is weak, there would not be much difference in $H(w, y)$ and $H_\Delta(w, y, z; \Delta)$, resulting in failure to reject the null when the alternative is true.

4. Identifiability vs. testability : When W is discretely observed, testing exogeneity via the above proposition is *not* confirmable because the variation in U would not be fully observed in W in this case. That is, the fact that W is independent of Z conditional on Y does not imply U and Y are independent. However, if W is not independent of Z conditional on Y, then we can reject the hypothesis that U and Y are independent.

5. Identification of the shape of the distribution of the unobservables : when triangularity and scalar unobserved variable assumptions are additionally imposed one can establish identification of the shape of $F_{UY|XZ}$. In this case, if the outcome shows continuous variation, the null hypothesis of exogeneity of Y, that is, the hypothesis that U and Y are independent, is refutable, as well as confirmable.

3 The Test Statistic

The testable implication that if an explanatory variable is exogenous, the outcome is independent of an IV for the explanatory variable conditional on the explanatory variable can be implemented by using the Cramer-von Mises test or the Kolmogorov-Smirnov test.

As other exogenous explanatory variables, X, can be added as additional conditioning variables without changing the results from this on X is omitted. Let $(\mathbf{\Psi}, \mathcal{S}, P)$ be the probability space of the random vector $\Psi = (W, Y, Z)$, and let $\mathbf{\Psi} \subseteq R^{dz+2}$ be the support of Ψ, where d_z is the dimension of Z. W and Y are scalar random variables. For simplicity we assume that $d_z = 1$. Let $\psi = (w, y, z)$ be the realization of Ψ. By Theorem 1 when there exists an IV, Z, satisfying Restriction C-IV, if Y is not endogenous,

$$W \perp Z \mid Y \text{ for all } \psi \in \mathbf{\Psi}.$$

The null hypothesis to be tested is

$$H_0 : P[F_{W|YZ}(w|Y, Z) - F_{W|Y}(w|Y) = 0] = 1, \forall \psi \ \in \mathbf{\Psi} \tag{1}$$

and the alternative can be written by the negation of the null as

$$H_1 : P[F_{W|YZ}(w|Y,Z) - F_{W|Y}(w|Y) = 0)] < 1, \text{ for some } \psi \ \in \mathbf{\Psi}. \tag{2}$$

This then can be equivalently written, using indicator functions as the following conditional moment

$$\begin{aligned} F_{W|YZ}(w|Y,Z) - F_{W|Y}(w|Y) &= E[\mathbf{1}(W \leq w)|Y,Z] - E[\mathbf{1}(W \leq w)|Y] \\ &= 0 \end{aligned} \tag{3}$$

By iterated expectation, we can rewrite it as

$$E\{\mathbf{1}(W \leq w) - E[\mathbf{1}(W \leq w|Y]|Y,Z\} = 0. \tag{4}$$

For some bounded nonnegative function $g(Y)$ (4) is equivalent to

$$g(Y)E\{\mathbf{1}(W \leq w) - E[\mathbf{1}(W \leq w)|Y]|Y,Z\} = 0.$$

As $g(Y)$ is nonstochastic conditinal on Y and Z, we rewrite this as

$$E[g(Y)\{\mathbf{1}(W \leq w) - E[\mathbf{1}(W \leq w)|Y]|Y,Z\}] = 0. \tag{4-1}$$

This conditional moment condition can be equivalently re-expressed as an *unconditional* moment condition by using the indicator functions of the conditioning variables, Y and Z[8]

$$E[g(Y)\mathbf{1}(Y \leq y)\mathbf{1}(Z \leq z)\{\mathbf{1}(W \leq w) - E(\mathbf{1}(W \leq w|Y)\}]. \tag{5}$$

Thus, the null hypothesis now is stated in terms of unconditional moments using the fact that $E[\mathbf{1}(W \leq w)|Y] = F(w|Y)$

$$\begin{aligned} H_0 \ &: \ P[E[g(Y)\mathbf{1}(Y \leq y)\mathbf{1}(Z \leq z)\{\mathbf{1}(W \leq w) - F(w|Y)\} = 0], \\ \forall\psi \ &\equiv \ (w,y,z) \ \in \mathbf{\Psi}. \end{aligned} \tag{6}$$

[8] We use an unconditional moment condition of the form $E(gh) = 0$ to conclude $E(g|X) = 0$, where g is an integrable function of a vector of random variables and h is an integrable function of a random variable, X. (X can be a *vector*). Bierens (1990) considered exponential function for this purpose and Bierens and Ploberger (1997) showed a general class of real valued weight functions, h. The function $h(\cdot)$ is called a "test function" in Stinchcombe and White (1998) and they showed that a class of non-polynomial functions, called *totally revealing*, can be used for this purpose. Indicator functions for the possibly vector, X, are *comprehensively revealing* which is always *totally revealing*, can be used as a test function.

The following arguments apply to a continuous W. The test we consider is based on the sample analog of (6) using a kernel estimator for $F(w|Y)$. Let the sample analog, $T_n(w, y, z)$, be

$$\begin{aligned}
T_n(w,y,z) &= \frac{1}{n}\sum_{i=1}^{n}[\widehat{g}(Y_i)\mathbf{1}(Y_i \le y)\mathbf{1}(Z_i \le z)\{\mathbf{1}(W_i \le w) - \widehat{F}(w|Y_i)\} \quad (7)\\
&= \frac{1}{n^2 h}\sum_i\sum_j K_{ij}^h[\mathbf{1}(W_i \le w) - \mathbf{1}(W_j \le w)]\mathbf{1}(Y_i \le y)\mathbf{1}(Z_i \le z)\\
\textit{where} \quad \widehat{F}(w|Y_i) &= \sum_{j=1}^{n}\frac{\mathbf{1}(W_j \le w)K_{ij}^h}{\sum_{j=1}^{n}K_{ij}^h}\\
&= \frac{1}{\widehat{g}(Y_i)\cdot nh}\sum_{j=1}^{n}\mathbf{1}(W_j \le w)K_{ij}^h,\\
\textit{and} \quad \widehat{g}(Y_i) &= \frac{1}{nh}\sum_{j=1}^{n}K_{ij}^h,\\
\textit{where} \quad K_{ij}^h &\equiv K\left(\frac{Y_i - Y_j}{h}\right), \quad \text{and } h \text{ is bandwidth.}
\end{aligned}$$

We consider the Kolmogorov-Smirnov statistic of the form[9]

$$KS_n = \sup_{\psi \in \mathbf{\Psi}} |\sqrt{n}T_n(\psi)|. \quad \text{(KS)}$$

The test statistic is a functional of the random element, $\sqrt{n}T_n$. We now specify assumptions required in section 4 to establish asymptotic properties of the test statistic, KS_n.

1. **Assumption A1 (INDEPENDENT DATA)** : The observed sample consists of the n random variables $\{\Psi_i : i \le n\}$, where Ψ is defined previously. We assume that the data are generated by a distribution function P.

2. **Assumption A2 (REGULARITY for kernels,** $K(s)$**)** **(i)** *$K(s)$ is a bounded nonnegative function such that $\int K(s)ds = 1$ and $\int sK(s)ds = 0$. $K(s) = 0$*

[9] One can also use the Cramer-von Mises' statistic of the form

$$CM_n = \frac{1}{n}\sum(\sqrt{n}T_n(\psi_i))^2.$$

for all boundary points of the support of S, (ii) $K(s)$ *is a symmetric function* $K(s) = K(-s)$, for *all* s, (iii) $sup_s|K(s)|+\int |K(s)|ds < \infty$, and (iv) $\int K(\frac{Y_i - s_2}{h})P(ds_2) = O_p(1)$, and $\int\int K(\frac{u-v}{h})P(du)P(dv) = O_p(1)$ and (v) $nh \to \infty$ as $n \to \infty$ for some finite nonnegative bandwidth h.

Comments on the assumptions

1. A2 (i) and (ii) are used to find out the symmetric "kernel" of the U-statistic, which simplifies the asymptiotic results.
2. A2 (iii) and boundedness of the kernel function, $K(\cdot)$ are used in showing that the class of the functions that characterize the empirical process is Euclidean.

In the next section we first establish the asymptotic distribution of the random element $\sqrt{n}T_n$, then we discuss the asymptotic null distribution of the statistic, KS_n, by adopting the continuous mapping theorem to $\sqrt{n}T_n$.To derive the asymptotic distribution of $\sqrt{n}T_n$, which is a U-statistic, we transform this to obtain an empirical process using the Hoeffding (1948) decomposition (also described in Serfling (1980)).

Symmetry of $K(\cdot)$ and $\int sK(s)ds = 0$ imply that $K(0) = 0$. Rearranging (7) by using the symmetry of $K(\cdot)$ and $K(0) = 0$, we have

$$\begin{aligned} T_n(\psi) &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n-1} q_\psi^h(\Psi_i, \Psi_j), \qquad (8) \\ \text{where} \quad q_\psi^h(\Psi_i, \Psi_j) &= K_{ij}^h[\mathbf{1}_{W_i}^w - \mathbf{1}_{W_j}^w][\mathbf{1}_{Y_i}^y\mathbf{1}_{Z_i}^z - \mathbf{1}_{Y_j}^y\mathbf{1}_{Z_j}^z], \\ \textit{with} \quad K_{ij}^h &\equiv K(\frac{Y_i - Y_j}{h}), \mathbf{1}_B^b \equiv \mathbf{1}(B \leq b), \\ \Psi &= (W, Y, Z), \psi = (w, y, z). \end{aligned}$$

Although the random element $T_n(\psi)$ is a sum of functions of i.i.d data $\{\Psi_i\}$, the summands are not independent due to the double summation structure, thus a direct application of CLT or LLN is not possible. We use the theory from U-statistic of order 2 to approximate this into a sum of independent elements. Delgado and Gonzalez-Manteiga (2001) consider similar idea in their significance test, but in their proofs they do not use the same symmetric "kernel[10]" as we do. The symmetric form found here simplifies the proof significantly. Also the bootstrap method that depends on (8) also would lead to different asymptotic properties from Delgado and Gonzalez-Manteiga (2001).

[10]The symmetric function q is called "kernel" following Serfling (1980).

Let $\Psi_1, \Psi_2, ..$ be independent observations taken from a distribution P on a set $\mathbf{\Psi}$, and $\mathcal{H}_K^h$ be a class of real-valued symmetric functions on $\mathbf{\Psi}\otimes\mathbf{\Psi}$. We consider the U-process of order 2 for the U-statistic $T_n(\psi)$ of the following

$$
\begin{aligned}
\{T_n(\psi) &: q \in \mathcal{H}_K^h\}, \qquad (9)\\
\textit{where } \mathcal{H}_K^h &= \{q : q_\psi^h(\Psi_i, \Psi_j) = K_{ij}^h[\mathbf{1}_{W_i}^w - \mathbf{1}_{W_j}^w][\mathbf{1}_{Y_i}^y \mathbf{1}_{Z_i}^z - \mathbf{1}_{Y_j}^y \mathbf{1}_{Z_j}^z]\},\\
\text{where } K_{ij}^h &\equiv K(\frac{Y_i - Y_j}{h}), \mathbf{1}_B^b \equiv \mathbf{1}(B \le b), \psi \equiv (w, y, z) \in \mathbf{\Psi}, \, h \in R.
\end{aligned}
$$

The test statistic is constructed by fixing h and $K(\cdot)$ and how to choose the optimal h and $K(\cdot)$ is not considered in this paper. For given h and $K(\cdot)$, $\mathcal{H}_K^h$ is indexed by $\psi \in \mathbf{\Psi}$.

As is discussed in Serfling (1980), the asymptotic distribution of this U-statistic, $T_n(\psi)$ for given $\psi \in \mathbf{\Psi}$, can be found by the usual CLT once we approximate this into a form of sum of independent observation. That is, the asymptotic results are for each point $\psi \in \mathbf{\Psi}$, nevertheless, for our purposes this is not enough because as the observation varies, $q_\psi^h(\Psi_i, \Psi_j)$ will vary, thus, we need to consider the whole process of the U-statistic, $T_n(\psi)$ indexed by $\mathcal{H}_K^h$.We view now $T_n(\psi)$ as a U-process indexed by a class of functions, $q \in \mathcal{H}_K^h$.

With the form of U-process in (8) in hand, we apply the U-process theory (see Nolan and Pollard (1987,1988), Sherman (1994), Stute(1986), for example). We start from the decomposition of the U-statistic into an empirical process and a P-degenerate U-process.[11]

P_n is the empirical measure of a sample of random elements $\Psi_1, \Psi_2, ..., \Psi_n$, putting mass $\frac{1}{n}$ at each observation ans we denote $P_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{\Psi_i}$,where δ_{Ψ_i} is the Dirac measures at the observation. Using the Hoeffding decomposition of a U-statistic, we have

$$
\begin{aligned}
T_n(\psi) &= \widehat{T}_n(\psi) + R_n(\psi), \quad \psi \in \mathbf{\Psi} \qquad (10)\\
&\quad \textit{where}\\
\widehat{T}_n(\psi) &= E_{\Psi_i}E_{\Psi_j}[q_\psi^h(\Psi_i, \Psi_j)] + 2P_n\left\{E_{\Psi_j}[q_\psi^h(\Psi_i, \Psi_j)|\Psi_i] - E_{\Psi_i}E_{\Psi_j}[q_\psi^h(\Psi_i, \Psi_j)]\right\}\\
R_n(\psi) &= T_n(\psi) - \widehat{T}_n(\psi).
\end{aligned}
$$

That is, $T_n(\psi)$ can be decomposed into $\widehat{T}_n(\psi)$, the projection, and the remainder, $R_n(\psi)$. It turns out that only the projection part affects the asymptotic distribution

[11] A conditional mean zero U-process is called a P-degenerate U-process.

of $\sqrt{n}T_n$ under the null hypothesis. Note that if the null is true, $E[T_n(\psi)] = 0$, which implies $E[q^h_\psi(\Psi_i, \Psi_j)] = 0$. Thus, the null distribution of $\sqrt{n}T_n$ is determined by $\sqrt{n}\widehat{T}_n(\psi) = 2P_nE[q^h_\psi(\Psi_i, \Psi_j)|\Psi_i]$, if the remainder term, $R_n(\psi)$ is asymptotically uniformly negligible, as will be shown later. Therefore, $\sqrt{n}\widehat{T}_n(\psi) = 2P_nE[q^h_\psi(\Psi_i, \Psi_j)|\Psi_i]$ is the form of the empirial process that we will work on.

To apply the empirical theory results, we need to find the form of $P_n \left\{E_{\Psi_j}[q^h_\psi(\Psi_i, \Psi_j)|\Psi_i]\right\}$. We integrate out Ψ_j conditional on Ψ_i resulting in the following :

$$\begin{aligned} & P_n(E[q^h_\psi(\Psi_i, \Psi_j)|\Psi_i]) \qquad (11)\\ = & \frac{1}{n}\sum_{i=1}^{n}\int K(\frac{s_2 - Y_i}{h})[\mathbf{1}^w_{W_i} - \mathbf{1}^w_{W_j}][\mathbf{1}^y_{Y_i}\mathbf{1}^z_{Z_i} - \mathbf{1}^y_{Y_j}\mathbf{1}^z_{Z_j}]P(ds),\\ = & P_nr_1 + P_nr_2 + P_nr_3 + P_nr_4, \ \ where \end{aligned}$$

$$\begin{aligned} P_nr_1(\psi) &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(W_i \le w)\mathbf{1}(Y_i \le y)\mathbf{1}(Z_i \le z)\int K(\frac{Y_i - s_2}{h})P(ds)\\ P_nr_2(\psi) &= -\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(W_i \le w)\mathbf{1}(Z_i \le z)\int K(\frac{Y_i - s_2}{h})\mathbf{1}(s_2 \le y)P(ds)\\ P_nr_3(\psi) &= -\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(W_i \le w)\int K(\frac{Y_i - s_2}{h})\mathbf{1}(s_2 \le y)\mathbf{1}(s_3 \le z)P(ds) \qquad (12)\\ P_nr_4(\psi) &= \frac{1}{n}\sum_{i=1}^{n}\int K(\frac{Y_i - s_2}{h})\mathbf{1}(s_1 \le w)\mathbf{1}(s_2 \le y)\mathbf{1}(s_3 \le z)P(ds),\\ where \ \ s &= (s_1, s_2, s_3) \text{ and } \psi = (w, y, z) \in \mathbf{\Psi}. \end{aligned}$$

As under the null, $E_{\Psi_j}[E[r_1(\psi) + r_2(\psi) + r_3(\psi) + r_4(\psi)|\Psi_i]] = 0$, the empirical process of concern is $\sqrt{n}\widehat{T}_n(\psi) = \sqrt{n}(P_n - P)(r)$, where $r(\psi) = r_1(\psi) + r_2(\psi) + r_3(\psi) + r_4(\psi)$ and Pr is defined as $Pr \equiv \int rdP$ for measure P. Define $\mathcal{F} = \{r : r(\psi) = r_1(\psi)+r_2(\psi)+r_3(\psi)+r_4(\psi), \psi \in \mathbf{\Psi}\}$. The asymptotic properties are therefore, characterized by the properties of $\mathcal{F}$. In the next section we report results on the asymptotic properties of the proposed test-statistic.

4 Asymptotic properties

4.1 Asymptotic null distribution of KS_n

Using the U-process and empirical process theory we show the asymptotic null distribution of the test statistics proposed in Section 3. $R_n(\psi)$ is a P-degenerate U-process.

We first show that the remainder term is uniformly asymptotically negligible. All proofs appear in Appendix.

Define $\mathcal{R} = \{R_n : R_n(\psi) = T_n(\psi) - \widehat{T}_n(\psi), \psi \in \mathbf{\Psi}\}$, where $T_n(\psi)$, $\widehat{T}_n(\psi)$ are defined in Section 3. We can show that R_n is P-degenerate U-statistic and that $\mathcal{R}$ is Euclidean by Lemma 6 in Sherman (1994), which allows us to apply their results to R_n to characterize the behavior of R_n. The results in Sherman (1994) is used to show that the remainder term is uniformly asymptotically negligible.

Theorem 2 Under Assumption A1 and A2,

$$\underset{\psi\in\mathbf{\Psi}}{sup}|R_n(\psi)| = O_p(\frac{1}{n}).$$

Proof. See Appendix. ■

Theorem 2 shows that $\sqrt{n}R_n(\psi) = o_p(1)$. Note that under Assumption A1 and A2, $\mathcal{H}_K^h$ is a class with the envelope, $H \equiv \sup_{q\in\mathcal{H}_K^h} |q_\psi^h(\Psi_i, \Psi_j)| = |K(\cdot)| < \infty$, and $\sup_{q\in\mathcal{H}_K^h} |Pq| < \infty$ and $\sup_{q_1,q_2\in\mathcal{H}_K^h} |Pq_1q_2| < \infty$, where Pq is defined previously. Likewise, under Assumptions A1 and A2, $\mathcal{F}$ is the class with the envelope, $F \equiv \sup_{r\in\mathcal{F}} |r| = 4|\int K(\frac{Y_i - s_2}{h})P(ds_2)| < \infty$ and $\sup_{r\in\mathcal{F}}|Pr| < \infty$ and $\sup_{r_1,r_2\in\mathcal{F}}|Pr_1r_2| < \infty$. Let $l^\infty(\mathcal{F})$ denote the Banach space of real bounded functions on $\mathcal{F}$ with the supremum norm, $||Qq||_\mathcal{F} = \sup_{f\in\mathcal{F}}|Qf|$.

Since the remainder term is uniformly asymptotically negligible, the asymptotic distribution of $\sqrt{n}T_n$ is determined by the projection part, $(\sqrt{n}\widehat{T}_n(\psi))$,by the decomposition in (10) as $E[q_\psi^h(\Psi_i, \Psi_j)] = 0$. The projection under the null then is,

$$\widehat{T}_n(\psi) = 2P_nE[q_\psi^h(\Psi_i, \Psi_j)|\Psi_i] = 2P_n[f_1(\Psi_i) + f_2(\Psi_i) + f_3(\Psi_i) + f_4(\Psi_i)],$$

with $Pr = 0$ for $r \in \mathcal{F}$.

We use Andrews (1994)'s proposition in establishing Theorem 3. The main difficulty is to show stochastic equicontinuity of $\mathcal{F}$. The sufficient conditions in Andrews (1994) are used. We use the pseudometric

$$\rho(\tau_1, \tau_2) = \sup_{N\geq 1}\{\frac{1}{N}\sum E[\widehat{T}_n(\tau_1) - \widehat{T}_n(\tau_2)]^2\}^{1/2}.$$

The stochastic equicontinuity results hold for this pseudometric.

Theorem 3 Under Assumption A1 and A2,

$$\mathcal{F} = \{r : r = r_1(\psi) + r_2(\psi) + r_3(\psi) + r_4(\psi), \psi \in \mathbf{\Psi}\} \; is \;\; P - Donsker,$$
$$\text{where } r_1(\psi), r_2(\psi), r_3(\psi), r_4(\psi) \text{ defined in (12).}$$

Proof. See Appendix. ■

Theorem 3 states that $\sqrt{n}T_n(\psi)$ weakly converges to a tight Borel measurable element in $l^\infty(\mathcal{F})$. The nature of the tight limit process follows from consideration of its marginal distributions. The marginal distributions $P_n r$ converge if and only if r is square integrable in which case the multivariate central limit theorem yields that for any finite set $r_1, r_2, ..., r_n$ of functions

$$(P_n r_1, ..., P_n r_n) \text{ weakly converges to } N_k(0, \Sigma)$$

where Σ is the $k \times k$ matrix with the (i, j)th element $P(r_i - Pr_i)(r_j - Pr_j)$. Since convergence in $l^\infty(\mathcal{F})$ implies marginal convergence, it follows that the limit process must be a zero-mean Gaussian process withcovariance function

$$P(r_1 - Pr_1)(r_2 - Pr_2), \quad \text{for } r_1, r_2 \in \mathcal{F}$$

Theorem 4 Under H_0, if Assumption A1 and A2 hold,

$$\sqrt{n}T_n(\psi) \text{ weakly converges to } G_p \text{ in } l^\infty(\mathcal{F}),$$

where G_p is sample continuous Gaussian, and the underlying distribution P, with zero mean and covariance given by

$$E(G_p r_1 G_p r_2) = P(r_1 - Pr_1)(r_2 - Pr_2), \quad \text{for } r_1, r_2 \in \mathcal{F}.$$

Note that $E(G_p r_1 G_p r_2)$ is finite since $\sup_{r \in \mathcal{F}} |Pr| < \infty$ and $\sup_{r_1, r_2 \in \mathcal{F}} |P\, r_1 r_2| < \infty$. Since $KS_n = \sup_{\psi \in \mathbf{\Psi}} |\sqrt{n}T_n(\psi)|$ is equal to the norm in $l^\infty(\mathcal{F})$, KS_n is continuous. Then we have the following results by applying the continuous mapping theorem to $\sqrt{n}T_n(\psi)$.

Corollary 1 Under H_0, if Assumption A1 and A2 hold,

$$KS_n \text{ weakly converges to } ||G_p r||_{\mathcal{F}}.$$

Notice that the asymptotic null distribution of the test statistic depends on the underlying features of the true distribution of the data, P, which are unknown. Thus, we propose to use the bootstrap to find out the critical values. This will be discussed in the next section. We first show in the next subsection that the test statistic is consistent, that is, when the null is false, the proposed test can detect it with probability one when the sample size is large.

4.2 Consistency and Local Power

4.2.1 Consistency

We show the asymptotic unit power of the test. We specify the alternative hypothesis in assumption A3.

Assumption A3 (ALTERNATIVE HYPOTHESIS) The alternative hypothesis is

$$H_1 : P[F_{W|YZ}(w|Y,Z) - F_{W|Y}(w|Y) = 0)] < 1, \; for \; some \; \psi \in \mathbf{\Psi}.$$

Theorem 5 Under Assumption A1-A3, for all sequences of random variables $\{c_n : n \geq 1\}, with \; c_n = O_p(1)$, we have $\lim_{n\to\infty} P(KS_n > c_n) = 1$.

Under the alternative, we have one more term to consider when we derive the asymptotic distribution of $\sqrt{n}T_n(\psi)$. Under the null, $E[T_n(\psi)] = 0$ in the decomposition (4), thus, the asymptotic null distribution can be derived without considering this term. Under the alternative, $E[T_n(\psi)]$ needs to be added. Note that

$$\begin{aligned} E[q_\psi^h(\Psi_i, \Psi_j)] &= E_{\Psi_i} E_{\Psi_j}[q_\psi^h(\Psi_i, \Psi_j)|\Psi_i] \\ &= E_{\Psi_i}[r_1(\Psi_i) + r_2(\Psi_i) + r_3(\Psi_i) + r_4(\Psi_i)] \end{aligned}$$

where $r_i, i = 1, 2, 3, 4$, is defined in (6). Then we have

$$\begin{aligned} Er_1(\Psi) &= \int \mathbf{1}(t_1 \leq w)\mathbf{1}(t_2 \leq y)\mathbf{1}(t_3 \leq z)\int K(\frac{t_2 - s_2}{h})P(ds)P(dt) \\ Er_2(\Psi) &= -\int \mathbf{1}(t_1 \leq w)\mathbf{1}(t_3 \leq z)\int K(\frac{t_2 - s_2}{h})\mathbf{1}(s_2 \leq y)P(ds)P(dt) \\ Er_3(\Psi) &= -\int \mathbf{1}(t_1 \leq w)\int K(\frac{t_2 - s_2}{h})\mathbf{1}(s_2 \leq y)\mathbf{1}(s_3 \leq z)P(ds)P(dt) \\ Er_4(\Psi) &= \int\int K(\frac{t_2 - s_2}{h})\mathbf{1}(s_1 \leq w)\mathbf{1}(s_2 \leq y)\mathbf{1}(s_3 \leq z)P(ds)P(dt), \\ \text{where } s &= (s_1, s_{2,} s_3) \text{ and } t = (t_1, t_{2,} t_3). \end{aligned}$$

From (5), under the alternative, we need consider the random element

$$\sqrt{n}\widehat{T}_n(\psi) = \sqrt{n}E[q_\psi^h(\Psi_i, \Psi_j)] + 2\sqrt{n}P_n\left\{E_{\Psi_j}[q_\psi^h(\Psi_i, \Psi_j)|\Psi_i] - E[q_\psi^h(\Psi_i, \Psi_j)]\right\}. \quad (13)$$

As under Assumption A2 (v) $E[q_\psi^h(\Psi_i, \Psi_j)] = O(1)$, when the null is false $\sqrt{n}\widehat{T}_n(\psi)$ tends to ∞ as $n \to \infty$ because the second term will converge to a Gaussian, with $E[q_\psi^h(\Psi_i, \Psi_j)]$ and $E(B_p q_1 B_p q_2)$ bounded. This guarantees the consistency of the test since the critical values are fixed, whereas the test-statistic, KS_n, tends to infinity as n increases. Thus, the null hypothesis is rejected with probability one.

4.2.2 Local Power

To examine the local power of the test we propose to consider $E[q_{\psi}^{h}(\Psi_i,\Psi_j)] = \frac{d(\psi)}{\sqrt{n}}$, $d(\psi) \neq 0$ and $d(\psi) < \infty$. That is, the alternatives converge to the null at a rate, $\frac{1}{\sqrt{n}}$. Then (13) becomes

$$\sqrt{n}\widehat{T}_n(\psi) = \sqrt{n}\frac{d(\psi)}{\sqrt{n}} + 2\sqrt{n}P_n\{E_{\Psi_j}[q_{\psi}^{h}(\Psi_i,\Psi_j)|\Psi_i] - \frac{d(\psi)}{\sqrt{n}}\} \tag{14}$$

Since the first term is bounded (because $E[q_{\psi}^{h}(\Psi_i,\Psi_j)] = O(1)$) and the second term is square integrable empirical process which weakly converges to a Gaussian process, G_P. Therefore, $\sqrt{n}\widehat{T}_n(\psi)$ weakly converges to $O(1) + B_P$.

Assumption A4 $E[q(\psi)] = \dfrac{d(\psi)}{\sqrt{n}}$, with $d(\psi) \neq 0$ and $d(\psi) < \infty$.

Theorem 6 Under Assumption A1, A2, and A4, for $c_\alpha = \inf\{q : F_{||G_pf||_{\mathcal{F}}}(q) \geq 1-\alpha\}$,

$$F_{KS_n}(c_\alpha) \leq 1 - \alpha + o(1).$$

Theorem 6 states that as the alternative approaches to the null at a rate of $n^{-1/2}$, and the test can still detect the alternative.

5 Bootstrap Critical Values

5.1 Nonparametric Bootstrap Critical Values

Corollary 1 provides the asymptotic null distribution of the proposed Kolmogorov-Smirnov statistic, KS_n. However, since the asymptotic null distribution depends on the unknown underlying distribution, P, the asymptotic distribution is not pivotal. A bootstrap resampling procedure is proposed to obtain critical values.

Consider the sample of $(\Psi_1, \Psi_2, ..., \Psi_n)$. Let $(\Psi_1^*, \Psi_2^*, ..., \Psi_n^*)$, where $\Psi_i^* = (W_i^*, Y_i^*, Z_i^*)$ be the bootstrapped sample from $(\Psi_1, \Psi_2, ..., \Psi_n)$. The procedure is described by the following five steps :

1. **Step 1** : Compute the realized value of the test statistic, KS_n, using the sample $(\Psi_1, \Psi_2, ..., \Psi_n)$.

2. **Step 2** : Resample n observations, $(\Psi_1^*, \Psi_2^*, ..., \Psi_n^*)$, where $\Psi_i^* = (W_i^*, Y_i^*, Z_i^*)$, $i = 1, 2, ..., n$, from the original data with replacement.

3. **Step 3** : Repeat **Step 2** B_n times to obtain B_n- batch of bootstrapped samples. Let $\{\Psi_i^{*k}\}_{i=1}^{B_n} = \{(W_i^{*k}, Y_i^{*k}, Z_i^{*k})\}_{i=1}^{B_n}$ denote the $k^{th}-$bootstrapped sample. Define $T_n^{(*k)}(\psi)$ from the $k^{th}-$bootstrapped sample as follows :

$$\begin{aligned}
T_n^{(*k)}(\psi) &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n-1} q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k}), \text{ where} \\
q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k}) &= K_{ij}^h[\mathbf{1}_{W_i^{*k}}^w - \mathbf{1}_{W_j^{*k}}^w][\mathbf{1}_{Y_i^{*k}}^y \mathbf{1}_{Z_i^{*k}}^z - \mathbf{1}_{Y_j^{*k}}^y \mathbf{1}_{Z_j^{*k}}^z], \\
with\ \mathbf{1}_B^b &= \mathbf{1}(B \le b), \quad K_{ij}^h \equiv K(\frac{Y_i^{*k} - Y_j^{*k}}{h}), \\
\psi &= (w, y, z), \\
\text{for } k &= 1, 2, ..., B_n.
\end{aligned}$$

From the B_n- value obtained from the above, define the mean of the values of the bootstrapped statistic, $\overline{T}_{B_n}^{(*)}(\psi)$,

$$\overline{T}_{B_n}^{(*)}(\psi) = \frac{1}{B_n} \sum_{k=1}^{B_n} T_n^{(*k)}(\psi).$$

4. **Step 4** : To impose the null hypothesis, define the centered object, $T_n^{*k}(h)$

$$T_n^{*k}(\psi) = T_n^{(*k)}(\psi) - \overline{T}_{B_n}^{(*)}(\psi).$$

Define the test statistic for the $k^{th}-$bootstrapped sample as

$$KS_n^{*k} = \sup_{\psi \in \mathbf{\Psi}} |\sqrt{n} T_n^{*k}(\psi)|$$

to find the *bootstrapped* null distribution of KS_n.

5. **Step 5** : The bootstrap test rejects the null hypothesis at significance level α if KS_n exceeds the empirical $\alpha - th$ quantile of $\{KS_n^{*k}\}_{k=1}^{B_n}$.

Recall that P is the underlying true probability measure, and P_n is the empirical measure. Let P_n^* be the bootstrap empirical measure. Two distinct empirical processes, $\sqrt{n}(P_n^* - P_n)$ and the $\sqrt{n}(P_n - P)$ indexed by $\mathcal{H}_k^h$ need to be considered.

Note that the asymptotic behavior of the bootstrapped object, $T_n^{*k}(\psi)$, is determined by the *projected* part since the bootstrapped remainder term is asymptotically uniformly negligible. Thus, the behaviour of the bootstrapped empirical process, $\sqrt{n}(P_n^* - P_n)$ indexed by $\mathcal{H}_k^h$ needs to be considered. We apply the same procedure of decomposition of U-statistic and show that the set of projection terms is $P-$Donsker.

The centering procedure in Step 4 is necessary to impose the null hypothesis to the bootstrapped object. Without this procedure, if the data are generated under the alternative, then the distribution simulated from the bootstrapped data will not provide information regarding the null distribution. We prove that the centered object obtained in Step 4,$T_n^{*k}(\psi)$, is of the same form regardless of whether the null hypothesis is true or not. Then it will be shown that the asymptotic behaviour of $T_n^{*k}(\psi)$ is determined by the projection term only which is in the same class of functions. As we have shown that the class of function in Theorem 3 is P-Donsker, the two empirical processes, $\sqrt{n}(P_n^* - P_n)$ and $\sqrt{n}(P_n - P)$ converge to the same Gaussian process by applying the extended version of the Gine and Zinn (1990)'s bootstrap central limit theorem that allow for nonmeasurability. (See ven der Vaart and Wellner (1996), for example)

Theorem 7 Suppose $B_n = O(n)$. Under Assumption A1-A2, the procedure described in Step1 - Step 5 (i) provides correct asymptotic size, α, under the null, (ii) is consistent against any fixed alternatives.
Proof. See Appendix. ■

If the null hypothesis were true, the bootstrap procedure would result in (asymptotically) correct size of the test, because the bootstrap test statistic, KS_n^*, converges to the same limiting distribution of KS_n under the null. When the alternative were true, because KS_n goes to infinity, whereas, the bootstrap critical value is still finite, the bootstrap procedure would result in a consistent test.

6 Illustration

6.1 Illustration - Endogeneity, Conditional Independence, and the Impact of Weak IV

I illustrate that the idea can be informally used to test exogeneity by plotting the conditional distribution functions. I also illustrate the possible loss of power due to the use of weak instruments. In each part we generate W, Y, and Z by the following data generating processes :

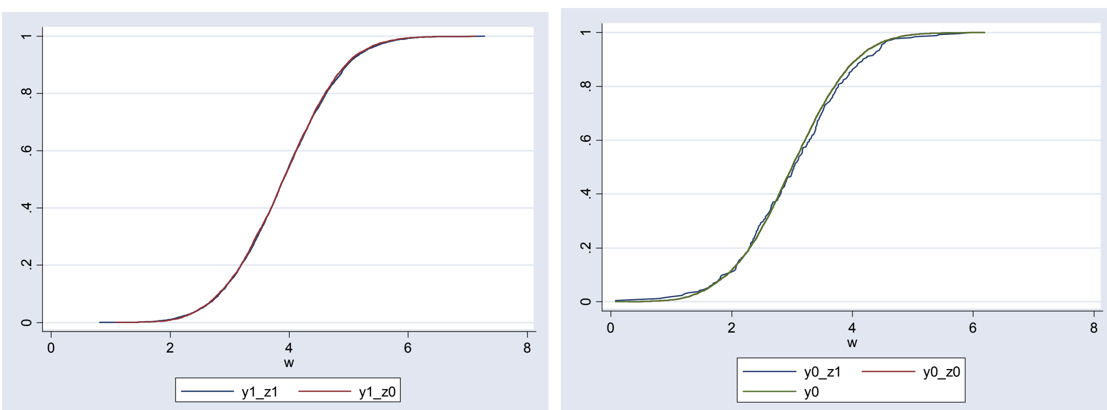
$$\begin{aligned} Z &\sim Poisson\ (\lambda),\ \lambda = 0.5 \\ Y &= 1(a_0 + a_1 Z + V \geq 0) \\ W &= b_0 + b_1 Y + U \\ \begin{pmatrix} U \\ V \end{pmatrix} | Z &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{VU} \\ \sigma_{VU} & \sigma_U^2 \end{pmatrix} \right) \end{aligned}$$

By varying a_1, we can control the "strength" of IV and by varying σ_{VU}, we control the degree of endogeneity. The distributions of W given Y and Z shown below are drawn using the data generated by the above processes. We draw the cumulative distribution functions, $F_{W|YZ}$, for $Y \in \{0,1\}$, and $Z \in \{0,1\}$ to examine the link between conditional independence and endogeneity, and how the link is affected by the strength of IV.

When $\sigma_{VU} = 0$, that is, when there is no endogeneity, the two conditional distributions for different values of Z are the same, while when $\sigma_{VU} \neq 0$, that is, when there is endogeneity, the two conditional distributions differ when the instrument is strong, but they do not show much difference when the instrument is weak.

1. Exogenous Y and conditional independence

I set $\sigma_{VU} = 0$. The two graphs show the distribution functions of W given Y and Z. The first panel shows whether $F_{W|YZ}$ is independent of Z once we condition on $Y = 1$. It shows that $F_{W|Y=1,Z=1} = F_{W|Y=1,Z=0}$[12]. The second panel is the distribution functions of W given Y and Z for $Y = 0$ for different values of $Z \in \{0,1\}$.

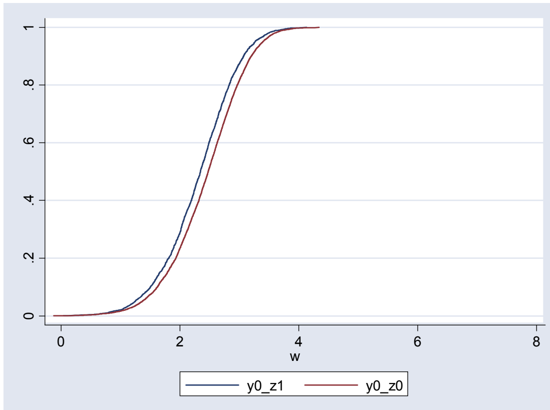


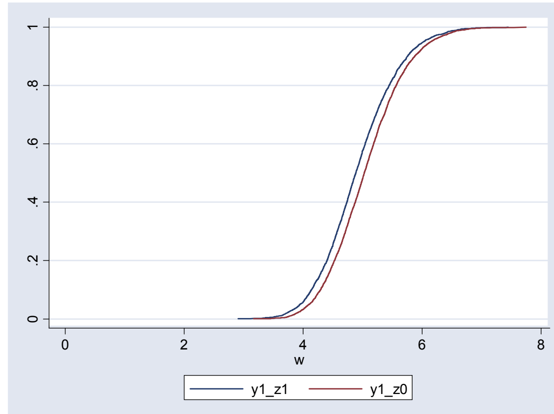
2. Weak IV, endogenous Y

2.1 $\sigma_{UV} = 0.7, and\ a_1 = 0.3$

I consider endogenous Y($\sigma_{UV} = 0.7$) and "relatively " weak IV ($a_1 = 0.3$). As long as Z is "relevant" the distribution of outcome seems to be affected by the values of Z.

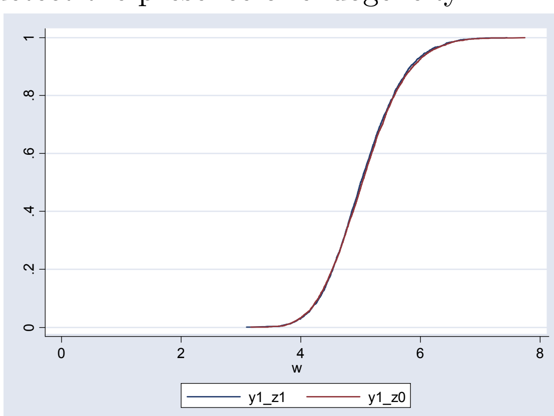
[12] Z is distributed by Poisson, but with mean $\lambda = 0.5$, there are a few observations for the values $Z = 2, 3, ...$

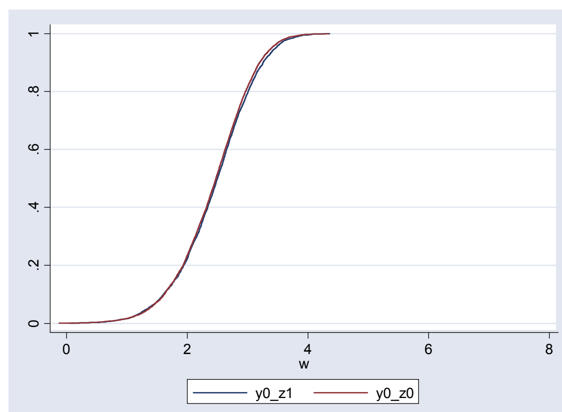




2.2 $\sigma_{UV} = 0.7$ and $a_1 = 0$

When Z is not "relevant", as we expected, the distribution of the outcome is not affected by the irrelevant IV conditioning on Y. Even though Y is endogenous, plotting $F_{W|Y=1Z=1}$ and $F_{W|Y=1Z=0}$ implies that $F_{W|YZ}$ is independent of Z. This shows the case in which testing exogeneity via testing conditional independence fails to detect the presence of endogeneity.

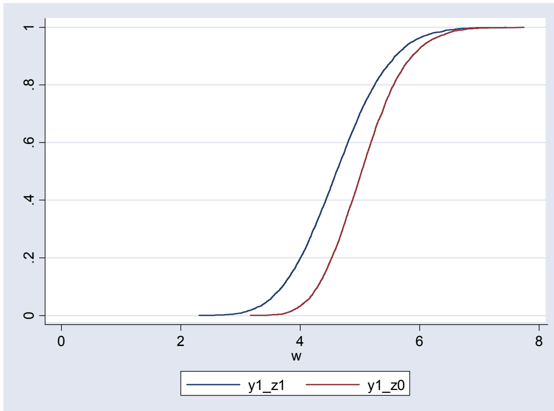


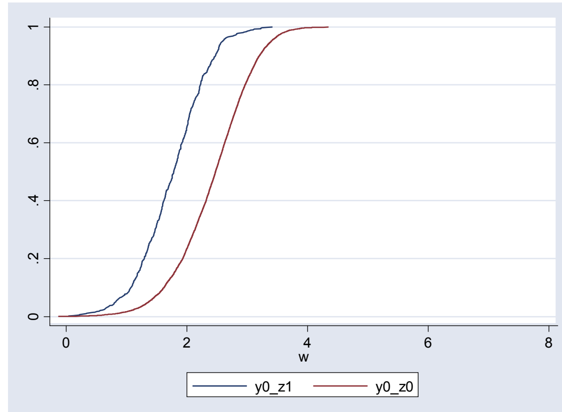


3. Strong IV, endogenous Y

$$\sigma_{UV} = 0.7, and\ a_1 = 1.3$$

Now consider a strong IV and endogenous Y. The conditional distribution is affected by both Y and Z even though Z is excluded from the outcome equation.





6.2 Test of Exogeneity of Veteran Status

I illustrate how the results can be used by examining the effects of the Vietnam-era veteran status on the civilian earnings using the data used in Abadie (2002)[13]. I use a sample of 11,637 white men, born in 1950-1953, from March Current Population Surveys of 1979 and 1981-1985. Annual earnings are used as an outcome, and the veteran status is the binary endogenous variable of concern.

Let W be annual labour earnings, Y be the veteran status, and Z be the binary variable determined by draft lottery. Age, race, and gender are controlled so that the subgroup considered is observationally homogenous. The unobserved variables U and V indicate scalar indices for "earnings potential" and "participation preference"/"aptitude for the army" each. There can be many factors that determine these indices, but we assume that these multi-dimensional elements affect the outcome only through a "scalar" index.

6.2.1 Selection on Unobservables - Endogeneity

Enrollment in military service during the Vietnam-era may have been determined by the factors which are associated with the unobserved earnings potential. This concern about selection on unobservables is caused by several aspects of decision processes both of the military and of those cohorts to be drafted. On the one hand, the military enlistment process selects soldiers on the basis of factors related to earnings potential. For example, the military prefer high school graduates and screens out those with low test scores, or poor health. As a consequence, men with very low earnings potential are unlikely to end up in the army. On the other hand, for some volunteers military service could be a better option because they expected that their careers in the civilian labour market would not be successful, while others with high

[13] The data are obtainable in Angrist Data Archive :
http://econ-www.mit.edu/faculty/angrist/data1/data

earnings potential probably found it worthwhile to escape the draft. This shows that the direction of selection could vary with where each individual is located in the distribution of the unobservable earnings potential.

6.2.2 Draft Lottery as an Instrument

As in Angrist (1990) the Vietnam era draft lottery is used as an instrument to identify the effects of veteran status on earnings. The lottery was conducted every year between 1970 and 1974. The lottery assigned numbers from 1 to 365 to dates of birth in the cohorts being drafted. Men with the lowest numbers were called to serve up to a ceiling[14] which was unknown in advance. We construct a binary IV based on the lottery number. It is assumed that this IV is not a determinant of earnings, and the unobserved scalar indices are independent of draft eligibility[15].

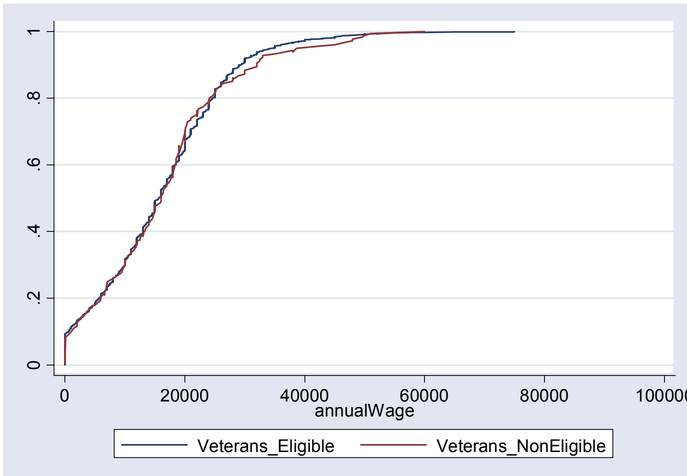
6.2.3 Illustration of the Testable Implication

The tesable implication can be used to investigate whether data show endogeneity both formally and informally. Informal testing involves plotting the conditional distribution of outcome given both endogenous variable and IV, Z. If there is no difference, then y is exogenous. However, the converse is not true unless Z is strongly correlated with Y. We show this in the investigation of the impact of the veteran status on the earnings using the draft-eligibility based on the lottery number as an IV.

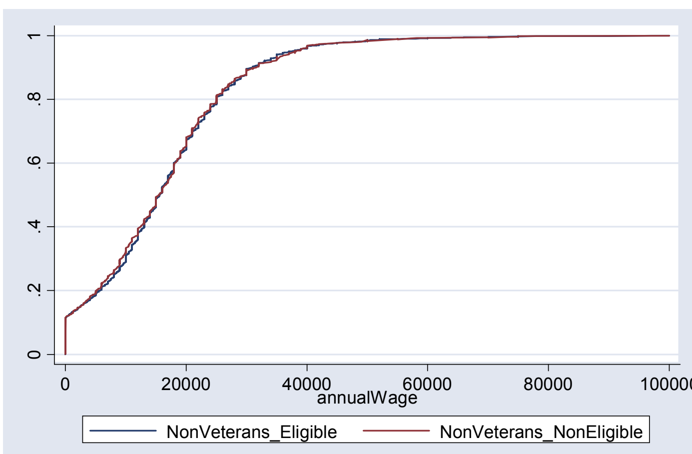
The distributions of annual earnings conditional on the values of Y and Z are drawn. <Figure 1> shows the conditional distribution of veterans for different eligibility criteria, and <Figure 2> shows that of non-veterans. As we can see the conditional distributions of the veterans seem to differ by different eligibility criteria. However, the conditional distributions of non-veterans do not seem to be different.

[14] See Angrist (1990) for more details.

[15] There has been some discussion on whether individuals' draft lottery numbers caused their behavior, e.g. some men could have volunteered in the hope of serving under better terms and gaining some control over the timing of their service once the lottery numberw were known. If those who change their behavior according to their draft lottery number show certain patterns in their unobserved factors, then the quantile invariance restriction may be violated.



Conditional distributions of annual earnings for veterans with different draft-eligibility



Conditional distributions of annual earnings for non-veterans with different draft-eligibility

7 Discussion and Conclusion

This paper proposes a nonparametric test of exogeneity. Since the asymptotic null distribution is not pivotal, we propose to use the bootstrap to find out the critical values by approximating the null distribution of the test statistic. We also show that the bootstrap test provides asymptotically correct size, and that the test is consistent.

Our test would be used for test of exogeneity in the context that there is no controversy on the exclusion of Z in the structural relation, for example, when natural experiments, randomized trials with noncompliance, or exogenous policy changes are used as instruments[16].

This test crucially relies on the existence of instrumental variables and the quality of instrumental variables. How the power is affected by the strength of IVs should be investigated.

Our test should be distinguished from omitted variables test. If some of the explanatory variables are endogenous, the omitted variable tests would misleadingly conclude that an IV is significant.

References

[1] Abadie, A. (2002), "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the America Statistcal Association*, 97. 284-292.

[2] Ait-Sahalia,Y., P. Bickel, TM Stoker (2001), "Goodness-of-fit tests for kernel regression with an application to option implied volatilities ," *Journal of Econometrics,* 105(2), 363-412.

[3] Andrews, D. (1994), "Empirical process methods in econometrics," *Handbook of Econometrics*, Vol IV, edited by R. Engle, and D. Mcfadden.

[4] Andrews, D. (1997), "A conditional Kolmogorov test," *Econometrica*, 65(5), 1097-1128.

[5] Angrist, J. (1990), "Lifetime earnings and the Vietnam era draft lottery : evidence from the social security administrative records," *American Economic Review*, 80, 313-336.

[6] Bierens, H.J.(1982), "Consistent model specification tests," *Journal of Econometrics,* 20, 105–134.

[7] Bierens, H.J.(1990), "A consistent conditional moment test of functional form," *Econometrica* 58, 1443–1458.

[8] Bierens, H. and W. Ploberger (1997), "Aymptotic theory of integrated conditional moment tests," Econometrica, 65(5), 1129-1151.

[9] Blundell, R. and R. Smith (1989), "Estimation in a class of simultaneous equation limited dependent variable models," *Review of Economic Studies*, 56, 37-58.

[16] Regression discontinuity design would be a good example.

[10] Blundell, R. and J. Horowitz (2007), "A nonparametric test of exogeneity," *Review of Economic Studies,* 74(4), 1035-1058.

[11] Breusch, T. (1986), "Hypothesis testing in unidentified models," *Review of Economic Studies*, 635-651.

[12] Chen, X. and Y. Fan (1999), "Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series," *Journal of Econometrics,* 91(2), 373-401.

[13] Chesher, A. (1985), "Score tests for zero covariances in recursive linear models for grouped or censored data," *Journal of Econometrics*, 28, 291-305.

[14] Chesher, A. (2003), "Identification in nonseparable models," *Econometrica*, 71, 1405-1441.

[15] Dawid, A.P. (1979) "Conditional independence in statistical theory," *Journal of Royal Statistical Society Series B* 41, 1-31.

[16] Delgado, M and W. Gonzalez Manteiga (2001), "Significance testing in nonseparametric regression based on the Bootstrap," *The Annals of Statistics*, 29(5), 1469-1507.

[17] Durbin, J. (1954), "Errors in variables," *R eview of the International Statistical Institute*, 22. 23-32.

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

[19] Gine, E. and J. Zinn, (1990), "Bootstrapping general empirical measures," *The Annals of Probability*, 18(2), 851-869.

[20] Grogger, D. (1990), "A simple test for exogeneity in probit, logit and Poisson regression models," *Economics letters*, 33, 329-332.

[21] Hahn, J. and G. Ridder (2011), "Conditional moment restrictions and triangular simulatenous equations," *Review of Economics and Statistics*, forthcoming.

[22] Hausman, J. (1978), "Specification tests in econometrics," *Econometrica*, 46. 1251-1272.

[23] Hoeffding, W. (1948), "A non-parametric test of Independence," *The annals of mathematical statistics*, 19(4), 546-557.

[24] Hurwicz, L. (1950) "Generalization of the concept of identification," in Statistical Inference in Dynamic Economic Models, Cowles Commision Monograph 10. Wiley, New York.

[25] Lavergn, P. and Vuong, Q. (1996), "Nonparametric selection of regressors : the nonnested case," *Econometrica*, 64. 207-219.

[26] Lavergn, P. and Vuong, Q. (2000), "Nonparametric significance testing," *Econometric Theory*, 16, 576-601.

[27] Lewbel, A. (1995), "Consistent nonparametric hypothesis tests with an application to Slutsky symmetry," *Journal of Econometrics,* 68(2)379-401

[28] Rivers, D. and Q. Vuong (1988), "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39. 347-366.

[29] Robinson, P. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931-954.

[30] Serfling, R. (1980), *Approximation theorems of Mathematical statistics*, Wiley Series of Probability and Mathematical Statistics.

[31] Sherman, R. (1994), "Maximal inequalities for degenerate U-processes with applications to optimization estimators," *The Annals of Statistics*, 22(1), 439-459.

[32] Smith, R. and R. Blundell, (1986), "An exogeneity test for a simultaneous tobit model," *Econometrica*, 54, 679-685.

[33] Song, K. (2007), "Testing Conditional Independence using Conditional Martingale Transforms," *Annals of Statistics*, 37, 4011-4045.

[34] Stute, W. (1986), "Conditional empirical processes," *The Annals of Statistics*, Vol. 14, No. ", 638-647.

[35] Su, L. and H. White (2007), "A Nonparametric Hellinger-metric nonparametric test for conditional independence," *Journal of Econometrics*,

[36] Su, L. and H. White (2008), "A consistent characteristic function-based test for conditional independence," *Journal of Econometrics*, 141(2), 807-834.

[37] van der Vaart,A. and J. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer Series in Statistics, Springer.

[38] Wu, D (1973), "Alternative tests of independence between stochastic regressors and disturbances," *Econometrica*, 41. 733-750.

Appendix - Proofs

Let the pdf/pmf of a random variable Z be $f(z)$ and $K(\cdot)$ be the kernel function used for smoothing. S_Q indicates the support of a eandom variable Q and C_1 and C_2 indicate arbitrary finite real values.

Lemma A1 Robinson (1988, lemma 2) Let $\sup_z f(z) < \infty,\ \sup_u |K(u)| + \int |K(u)|du < \infty$. Then

$$\begin{aligned} \sup_z E\left|K(\tfrac{Z-z}{h})\right| &= O(h^q), \\ \text{where } z &\in R^q. \end{aligned}$$

Lemma A2 Robinson (1988, lemma 3) Let $\sup_z f(z) < \infty,\ \sup_u |K(u)| + \int |K(u)|du < \infty$, $E|g(Z)| < \infty$. Then

$$\begin{aligned} E_{Z_1 Z_2}\left|g(Z_1)K(\tfrac{Z_1-Z_2}{h})\right| &= O(h^q), \\ \text{where } z &\in R^q. \end{aligned}$$

Lemma A3 $A_1 = \{l(x) : l(x) = \int K(\frac{s_2 - x}{h})P(ds_2) = E_X[K(\frac{X-x}{h})], x \in S_{Y_2}\}$ is stochastic equicontinuous .

Proof. By Lemma A1, there exists a constant C s.t. $l(x) = \int K(\frac{s_2 - x}{h})P(ds_2) = E_X[K(\frac{X-x}{h})] \leq C_1 h$ uniformly in x. Let the finite square integrable envelop function of the class A_1 be the constant $C_1 h$. Then A_1 is type II class as in Andrews (1994) since $l(x)$ is Lipschitz[17] in $x \in S_{Y_2}$.This is because there exists a constant C_2 s.t.

$$\begin{aligned} |l(x_1) - l(x_2)| &= \left|\int [K(\frac{s_2 - x_1}{h}) - K(\frac{s_2 - x_2}{h})]P(ds_2)\right| \leq C_2\left|x_1 - x_2\right|, \\ \text{for } C_2 &= \frac{C_1 h}{|x_1 - x_2|}. \end{aligned}$$

Then by Theorem 2 in Andrews(1994) the result follows. ■

Lemma A4 $A_2 = \{l(x) : l(x) = \int K(\frac{s_2 - x}{h})1(s_2 \leq y_2)P(ds_2) = E|1(s_2 \leq y_2)K(\frac{Z_1 - Z_2}{h})|\}$ is stochastic equicontinuous .

[17] A function $f(a)$ is Lipschitz in a if

$$|f(a_1) - f(a_2)| \leq B|a_1 - a_2|, \text{ for some real valued function } B(\cdot) \quad (1)$$

Proof. By Lemma A2 there exists a constant C s.t. $f(x) = \int K(\frac{s_2 - x}{h})P(ds_2) = E_X[K(\frac{X-x}{h})] \leq C_1 h$ uniformly in x. Let the finite square integrable envelop function of the class A_2 be the constant $C_1 h$. Then A_2 is type II class as in Andrews (1994) since $l(x)$ is Lipschitz in $x \in S_{Y_2}$.This is because there exists a constant C_2 s.t.

$$\begin{aligned} |l(x_1) - l(x_2)| &= \left| \int 1(s_2 \leq y_2)[K(\frac{s_2 - x_1}{h}) - K(\frac{s_2 - x_2}{h})]P(ds_2) \right| \leq C_2 \left| x_1 - x_2 \right|, \\ \text{for } C_2 &= \frac{C_1 h}{|x_1 - x_2|}. \end{aligned}$$

Then by Theorem 2 in Andrews (1994) the result follows. ■

The following is from Sherman (1994). Let $Z_1, ..., Z_n$ be independent observations from a distribution P on a set S. Let k be a positive integer and F a class of real-valued functions on $S^k = S \otimes ... \otimes S, k \geq 1$. For each $f \in F$, define

$$U_n^k f = (n)_k^{-1} \sum_{i_k} f(Z_1, ..., Z_{i_k})$$

where $(n)_k = n(n-1)\cdots(n-k+1)$, and $i_k = (i_1, ..., i_k)$ ranges over the $(n)_k$ ordered k-tuples of distinct integers from the set $\{1, 2, ..., n\}$. The collection of $\{U_n^k f;\, f \in F\}$ is called a U-process of order k and is said to be indexed by F.

Now we consider the decomposition of $U_n^k f$ into a sum of degenerate U-statistics. Let P denote the distribution on a set S. Fix $f \in F$. Suppose $P^k < \infty$, where P^k is the product measure $P \otimes ... \otimes P$. Then there exist functions $f_{1,...,} f_k$ such that, for each i, f_i is P-degenerate on S^i and

$$U_n^k f = P^k f + P_n f_1 + \sum_{i=1}^{k} U_n^i f_i$$

Definition A1 (Euclidean (A,V) (Sherman (1994)). Let F be a class of real-valued functions on a set S and $D(\varepsilon, d_Q, F)$ be the packing number. Call F Euclidean for the envelop F if there exist positive constants A and V with the following property : if Q is a measure for which $QF^2 < \infty$, then

$$\begin{aligned} D(\varepsilon, d_Q, F) &\leq A\varepsilon^{-V}, \quad 0 < \varepsilon \leq 1, \\ \text{where, } for \;\; f, g &\in \mathcal{F}, \\ d_Q(f,g) &= [Q|f-g|^2/QF^2]^{1/2}. \end{aligned}$$

Proposition A1 (Sherman(1994), Lemma 6) If F is Euclidean for an envelope F satisfying $P^k F^2 < \infty$, then F_i is Euclidean for an envelope F_i satisfying $P^k F_i^2 < \infty$.

Proof of Theorem 2

Proof. $H_K^h = \{q : q_\psi^h(\Psi_i, \Psi_j) = K_{ij}^h [1_{W_i}^w - 1_{W_j}^w][1_{Y_i}^y 1_{Z_i}^z - 1_{Y_j}^y 1_{Z_j}^z]$, for $\psi \in \Psi\}$ is Euclidean with an envelope, constant 1. This is the case because K_{ij}^h is Euclidean[18] with the constant envelope 1, a constant function and the other part of $q(\Psi_i, \Psi_j)$ is a function of indicators. By applying the permanence property of Euclidean, we conclude that H_K^h is Euclidean. Then by Proposition A1 R is Euclidean, thus the results in Sherman (1994) can be applied to R.Then Sherman (1994) Corollary 4(ii) with $k = 2$ is adopted to conclude $\sup_{\psi \in \mathbf{\Psi}} |R_n(\psi)| = O_p(\frac{1}{n})$. ■

Proposition A2 Under Assumptions A1,A2,

$$\begin{aligned} \mathcal{F}_1 &= \{r_1 : r_1(\Psi_i) = \mathbf{1}(W_i \leq w)\mathbf{1}(Y_i \leq y)\mathbf{1}(Z_i \leq z)\int K(\frac{Y_i - s_2}{h})P(ds), \psi \in \mathbf{\Psi}\} \\ & \quad is\ stochastic\ equicontinuous \end{aligned}$$

Proof. F_1 is stochastic equicontinuous by the mix and match Theorem 3 in Andrews (1994) (or by the permanency property discussed in van der Vaart and Wellner (1996)), due to the fact that indicator functions are VC classes and by Lemma A3 ■

Proposition A3 Under Assumptions A1,A2,

$$\begin{aligned} \mathcal{F}_2 &= \{r_2 : r_2(\Psi_i) = \mathbf{1}(W_i \leq w)\mathbf{1}(Z_i \leq z)\int K(\frac{Y_i - s_2}{h})\mathbf{1}(s_2 \leq y)P(ds), \psi \in \mathbf{\Psi}\} \\ & \quad is\ \ stochastic\ equicontinuous \end{aligned}$$

Proof. F_2 is stochastic equicontinuous by the mix and match Theorem 3 in Andrews (1994), due to the fact that indicator functions are VC classes and by Lemma A4. ■

Proposition A4 Under Assumptions A1,A2,

$$\begin{aligned} \mathcal{F}_3 &= \{r_3 : r_3(\Psi_i) = \mathbf{1}(W_i \leq w)\int K(\frac{Y_i - s_2}{h})\mathbf{1}(s_2 \leq y)\mathbf{1}(s_3 \leq z)P(ds), \psi \in \mathbf{\Psi}\} \\ & \quad is\ \text{stochastic equicontinuous}\ . \end{aligned}$$

Proof. F_3 is stochastic equicontinuous by the mix and match Theorem 3 in Andrews (1994), due to the fact that indicator functions are VC classes and by Lemma A3 and A4. ■

[18] Refer to Gine(1996).

Proposition A5 Under Assumptions A1,A2,

$$\mathcal{F}_4 = \{r_4 : r_4(\Psi_i) = \int K(\frac{Y_i - s_2}{h})\mathbf{1}(s_1 \leq w)\mathbf{1}(s_2 \leq y)\mathbf{1}(s_3 \leq z)P(ds), \psi \in \mathbf{\Psi}\}$$

is stochastic equicontinuous

Proof. F_4 is stochastic equicontinuous the mix and match Theorem 3 in Andrews (1994), by Lemma A3 and A4 ■

Proof of Theorem 3

Proof. This folllos because the permanency property and by Proposition A2 to A5. Note that the other two conditions in Andrew's propostion are satisfied. ■

Proof of Theorem 4

Proof. This follows from the decomposition of $T_n(\psi)$, Theorem 2 and Theorem 3. By Theorem 2 the remainder term is asymptotically uniformly negligible and the projection term is an empirical process that is P-Donsker, thus, $\sqrt{n}T_n(\psi)$ weakly converges to a Brownian Bridge with the mean and the covariance defined as stated in Theorem 4. ■

Proof of Corollary 1

Proof. This is by Theorem 4 and by the continuous mapping theorem. ■

Proof of Theorem 5

Proof. Under the alternative, an additional term in the projection needs to be considered.

$$\begin{aligned} & \sqrt{n}\widehat{T}_n(\psi) \\ = & \sqrt{n}E[q^h_\psi(\Psi_i, \Psi_j)] + 2\sqrt{n}P_n\left\{E_{\Psi_j}[q^h_\psi(\Psi_i, \Psi_j)|\Psi_i] - E[q^h_\psi(\Psi_i, \Psi_j)]\right\} \end{aligned}$$

The critical values of the null distribution are not known since the variance is not known, but we know that the variance is finite. Suppose the alternative is true. Then the statistic $\sqrt{n}\widehat{T}_n(\psi)$ goes to infinity, while the unknown critical values are finite, the test rejects the null with probability 1. ■

Proof of Theorem 6

Proof. This is direct from the fact that in (14)$\sqrt{n}\widehat{T}_n(\psi)$ weakly converges to $O(1) + G_P$. ■

Proof of Theorem 7

Proof. The proof is composed of several steps. In the first step it is shown that the remainder term for the bootstrapped object, $T_n^{(*k)}(\psi)$, is asymptotically negligible. In the second step, the limit processes of the centered bootstrapped empirical processes, $\widehat{T_n^{*k}}(\psi)$, is the same as that of $\widehat{T_n}(\psi)$, for $\psi \in \Psi$. In the final step, the continuous mapping theorem is used to discuss the asymptotic distributions of KS_n^{*k} and KS_n. Consider the sample of $(\Psi_1, \Psi_2, ..., \Psi_n)$. Let $(\Psi_1^*, \Psi_2^*, ..., \Psi_n^*)$, where $\Psi_i^* = (W_i^*, Y_i^*, Z_i^*)$ be the bootstrapped sample. Now $T_n^{(*k)}(\psi)$ is found by using the bootstrapped sample $(\Psi_1^*, \Psi_2^*, ..., \Psi_n^*)$, by (8)

From the B_n- value found from the above, define the mean of the values of the bootstrapped statistic, $T_n^{(*k)}(\psi)$,

$$
\begin{aligned}
T_n^{(*k)}(\psi) &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n-1} q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k}), \\
\text{where } q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k}) &= K_{ij}^h[\mathbf{1}_{W_i^{*k}}^w - \mathbf{1}_{W_j^{*k}}^w][\mathbf{1}_{Y_i^{*k}}^y \mathbf{1}_{Z_i^{*k}}^z - \mathbf{1}_{Y_j^{*k}}^y \mathbf{1}_{Z_j^{*k}}^z], \\
with\ \mathbf{1}_B^b &= \mathbf{1}(B \le b), \quad K_{ij}^h \equiv K(\frac{Y_i^{*k} - Y_j^{*k}}{h}), \\
\psi &= (w, y, z), \text{for } k = 1, 2, ..., B_n.
\end{aligned}
$$

Define the empirical measure of the bootstrapped observation, P_n^*. Note that $T_n^{(*k)}(\psi)$ can be decomposed into the projection and the remainder term of the following :

$$
\begin{aligned}
T_n^{(*k)}(\psi) &= \widehat{T_n^{(*k)}}(\psi) + R_n^{(*k)}(\psi), \quad \text{where} \\
\widehat{T_n^{(*k)}}(\psi) &= E[q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k})] + 2P_n^*\{E[q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k})|\Psi_i^*] - E[q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k})]\} \\
R_n^{(*k)}(\psi) &= T_n^{(*k)}(\psi) - \widehat{T_n^{(*k)}}(\psi), \\
&\quad \text{where } P_n^* \text{ is the bootstrap empirical measure} \\
\text{for } q &\in \mathcal{H}_K^h.
\end{aligned}
$$

Recall that $\overline{T}_{B_n}^{(*)}(\psi), T_n^{*k}(\psi)$, and KS_n^{*k} are defined as follows :

$$
\overline{T}_{B_n}^{(*)}(\psi) = \frac{1}{B_n} \sum_{k=1}^{B_n} T_n^{(*k)}(\psi)
$$

$$
T_n^{*k}(\psi) = T_n^{(*k)}(\psi) - \overline{T}_{B_n}^{(*)}(\psi)
$$

$$KS_n^{*k} = \sup_{\psi} |\sqrt{n} T_n^{*k}(\psi)|.$$

The form obtained in Step 4 is

$$\begin{aligned}\sqrt{n} T_n^{*k}(\psi) &= \sqrt{n} T_n^{(*k)}(\psi) - \sqrt{n}\overline{T}_{B_n}^{(*)}(\psi) \\ &= \sqrt{n} T_n^{(*k)}(\psi) - \underbrace{\frac{1}{B_n}\sum_{k=1}^{B_n}\sqrt{n} T_n^{(*k)}(\psi)}_{(A)} \qquad \text{(Boot-1)}\end{aligned}$$

Then consider (A).

$$\begin{aligned}(A) \quad &: \quad \frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n} T_n^{(*k)}(\psi) \\ &= \frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n}[\ \widehat{T_n^{(*k)}}(\psi) + R_n^{(*k)}(\psi)] \\ &= \underbrace{\frac{1}{B_n}\sum_{k=1}^{B_n}\sqrt{n}\ \widehat{T_n^{(*k)}}(\psi)}_{(B)} + \underbrace{\frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n} R_n^{(*k)}(\psi)}_{(C)}\end{aligned}$$

To examine the asymptotic behaviour of (A), the two components, (B) and (C) will be separately considered. Once it is shown that $\sup R_n^{(*k)}(\psi) = O_p(\frac{1}{n})$, (C) is asymptotically negligible. By Theorem 2, the remainder term is uniformly asymptotically negligible. Thus, the asymptotic behavior of $T_n^{(*k)}(\psi)$ will depend on (B).

Firstly, consider (B). Under the null, since $E[q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k})] = 0$,(B) becomes

$$\frac{1}{B_n}\sum_{k=1}^{B_n}\sqrt{n}\ \widehat{T_n^{(*k)}}(\psi) = \frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n} 2P_n^*\{E[q_\psi^h(\Psi_i^{*k}, \Psi_j^{*k})|\Psi_i^{*k}],$$

on the other hand, if the alternative is true, we have

$$\begin{aligned}&\frac{1}{B_n}\sum_{k=1}^{B_n}\ \sqrt{n}\widehat{T_n^{(*k)}}(\psi) \\ = \quad &\sqrt{n}E[q(\Psi_i^{*k}, \Psi_j^{*k}; \psi, h)] + \\ &\frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n} 2P_n^*\{E[q(\Psi_i^{*k}, \Psi_j^{*k}; \psi, h)|\Psi_i^{*k}] - E[q(\Psi_i^{*k}, \Psi_j^{*k}; \psi, h)]\}].\end{aligned}$$

Therefore, (Boot-1) can be reexpressed as follows

$$\begin{aligned}\sqrt{n}T_n^{*k}(\psi) &= \sqrt{n}T_n^{(*k)}(\psi) - \sqrt{n}\overline{T}_{B_n}^{(*)}(\psi) \\ &= \sqrt{n}\widehat{T_n^{(*k)}}(\psi) + R_n^{(*k)}(\psi) - \frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n}[\ \widehat{T_n^{(*k)}}(\psi) + R_n^{(*k)}(\psi)] \\ &= \sqrt{n}\widehat{T_n^{(*k)}}(\psi) - \frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n}\ \widehat{T_n^{(*k)}}(\psi) + o_p(1)\end{aligned}$$

since the remainder term is asymptotically uniformly negligible.

We now consider the projection terms of the bootstrapped object to show that they have the same form irrespective of whether the null is true or not.

$$\sqrt{n}\widehat{T_n^{(*k)}}(\psi) - \frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n}\ \widehat{T_n^{(*k)}}(\psi) \qquad \text{(Boot-2)}$$

$$= \left\{\begin{array}{cc} \begin{array}{c} 2\sqrt{n}P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] \\ -\frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n} 2P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] \end{array} & \text{under } H_0 \\ \begin{array}{c} 2\sqrt{n}P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] - P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})] \\ -\frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n}[2P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] - P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})]] \end{array} & \text{under } H_1 \end{array}\right\}$$

$$= \left\{\begin{array}{cc} \begin{array}{c} 2\sqrt{n}P_n^*\{E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] \\ -\frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n} 2P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] \end{array} & \text{under } H_0 \\ \begin{array}{c} 2\sqrt{n}P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] \\ -\frac{\sqrt{n}}{B_n}\sum_{k=1}^{B_n} 2P_n^* E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})|\Psi_i^{*k}] \end{array} & \text{under } H_1 \end{array}\right\}$$

The first equality is because of the decomposition and the definition of $\overline{T}_{B_n}^{(*)}(\psi)$,and the second equality is due to the fact that if H_1 were true, the additional term, $E[q_\psi^h(\Psi_i^{*k},\Psi_j^{*k})]$, needs to be subtracted to impose the null distribution. By recentering the proposed bootstrapped object could be used to approximate the null distribution even if the alternative were true.

Then note that from (Boot-2) the second term converges to zero by Glivenko-Cantelli Theorem, so the asymptotic behavior is determined by the first term.

Secondly, we show consistency of $T_n^{*k}(\psi)$

Since F is Donsker and the envelop of F is finite, by the extended version that allow for nonmeasurability of Gine and Zinn's (1990) bootstrap central limit theorem, the two processes $\sqrt{n}\widehat{T^*}_n(\psi)$ and $\sqrt{n}\widehat{T}_n(\psi)$ converge to the same Gaussian process.

Thirdly, since sup is a continuous operator, by the continuous mapping theorem we can conclude that the limiting distributions of KS_n^* and KS_n are the same, thus the quantiles should be the same. This guarantees that $c_n^* \to c_p(\alpha)$.

To show (ii) when the null is not true, $\sqrt{n}\widehat{T}_n(\psi)$ tends to infinity as n goes to infinity, whereas c_n^* is bounded, thus, the test rejects the null with probability one when the sample size is large. ■