Testing for Stochastic Marginal Effects

Shu Shen *
The University of Texas at Austin
August 25, 2010

Abstract

This paper proposes Kolmogorov-Smirnov type tests for testing whether a covariate has a uniformly positive (or negative) effect on the conditional distribution of a response variable given the covariate and other conditioning variables. This type of analysis is useful in situations where the econometrician or policy maker is interested in knowing whether a variable or policy would improve the distribution of the response outcomes in a stochastic dominance sense. The response variable is assumed to be continuous, while both discrete and continuous covariate cases are considered. I derive the asymptotic distribution of the test statistics and show that they have known simple asymptotic distributions by using and extending conditional empirical process results given by Horvath and Yandell (1988) [16]. Monte Carlo experiments are conducted and the tests are shown to have reasonable small sample behavior. The tests are applied to a study on child gender and parental income.

*E-mail: shu.shen@mail.utexas.edu. I am deeply indebted to my advisor Stephen Donald for his continuous support and encouragement. I would like to thank Jason Abrevaya, Maxwell Stinchcombe and Yu-chin Hsu for many valuable suggestions. All errors are mine.
1 Introduction

Empirical researchers in economics are often interested in the marginal effect of variables, such as those under control of policy-makers on economic outcomes. To date the standard approach is to focus on the mean outcome though recently authors have begun to consider other aspects of the distribution of the outcome. For example, Firpo (2007) [11] proposed estimators for quantile treatment effects. Abrevaya (2001) [1] looked at the impact of various maternal behaviors on the lower end of the infant birthweight distribution. In this paper, we study the stochastic marginal effect which is defined as the marginal effect of a covariate on the conditional distribution of a response given the covariate and other conditioning variables. The stochastic marginal effect is a function from the support of the response variable to the real line. It is of interest when the econometrician or policy maker wants to know not only a variable or policy’s effect on some specific characteristic (mean or specific quantiles) of the response outcomes but also its effect over the whole distribution of the response outcomes. This paper proposes Kolmogorov-Smirnov type tests for testing whether a stochastic marginal effect function has uniform sign, or a variable has a uniformly positive or negative effect on the conditional distribution of the response variable.

Let $Y$ be a continuous random variable with support $\mathcal{Y} \subseteq \mathbb{R}$ and random variable $X = (X_1, X_2)$ with support $\mathcal{X} \subseteq \mathbb{R}^p$. Suppose $X_1$ is a single dimensional dummy variable and that $X_2 \in \mathcal{W} \subseteq \mathbb{R}^{p-1}$ is continuous. The stochastic marginal effect of the discrete $X_1$ on the conditional distribution $F_{Y|X}$ is the difference between $F_{Y|X}$ evaluated at different $X_1$ values. The first null hypothesis we study in this paper is as follows

$$H^1_{0,x_2}: \text{For fixed } x_2 \in \mathcal{W}, F_{Y|X_1=1,X_2=x_2}(y|x_2) \leq F_{Y|X_1=0,X_2=x_2}(y|x_2) \text{ for all } y \in \mathcal{Y}.$$ 

The null hypothesis states that the conditional distribution of $Y$ at $X_1 = 1$, $X_2 = x_2$ stochastically dominates that at $X_1 = 0$, $X_2 = x_2$. In other words, for fixed $x_2$ the probability that $Y$ is
less than $y$ is always lower when $X_1 = 1$ than when $X_1 = 0$. The stochastic marginal effect of $X_1$ is equal to the *stochastic conditional treatment effect* (conditional on $X_2$) if selection of $X_1$ is based on observables ($X$). When $p=1$ then this hypothesis becomes identical to the null of first order stochastic dominance as considered in Barrett and Donald (2003) [3].

Lee and Whang (2009) [20] (later as LW) study a null hypothesis that is stronger than $H_{0,x_2}^1$ in the sense that their null is formulated for all possible $X_2$ values instead of at given fixed values. Though LW’s test has some nice features, such as power properties typical of parametric tests, in applications with many covariates the null hypothesis for their test is likely to be too strong and their test computationally too demanding to make it useful in practice. In contrast, our test enables econometricians to study the stochastic marginal effect of $X_1$ on $F_{Y|X}$ evaluated at any interesting values of conditioning $X_2$. For example, if a covariate is expected to affect the response variable through two mechanisms in opposite directions, it might be possible that one mechanism dominates for some $X_2$ values and the other for the rest. Then the econometrician may want to evaluate the stochastic marginal effect at different $X_2$ values.

Now consider the case where the covariate $X_1$ is continuous. The stochastic marginal effect of $X_1$ on the conditional distribution $F_{Y|X}$ is the partial derivative of $F_{Y|X}$ with respect to $X_1$. $H_{0,x}^2$ is a null hypothesis testing whether the stochastic marginal effect has a uniformly negative sign.

$$H_{0,x}^2 : \text{For fixed } x \in X, \frac{\partial}{\partial x_1} F_{Y|X=x}(y|x) \leq 0 \text{ for all } y \in Y.$$  

If $H_{0,x}^2$ is true, then at the value $x$ a marginal increase in $X_1$, holding $X_2$ fixed decreases the probability that $Y$ is less than $y$ uniformly across the support. As an example, suppose that $Y = H(X, e)$, where $e$ denote the unobserved influencing factors. Whether the stochastic marginal effect above gives a structural interpretation or not depends on the independence relationship between $X$ and $e$.

The closest existing literature to our second benchmark test is Lee, Linton and Whang (2009)’s
[19] (later as LLW) stochastic monotonicity test. Their test studies whether a variable has uniformly positive or negative stochastic marginal effects for all its possible values. LLW does not explicitly discuss the multi-dimensional conditional variable case. Therefore, their null hypothesis is a stronger version of $H^2_{0,x}$ when dimension of $X$ equals to 1: rejecting $H^2_{0,x}$ at any $X$ value rejects the stochastic monotonicity in LLW. In applications, our test is much faster in computation as we don’t need to use a grid for $X$ when calculating the test statistic. We will compare the small sample behavior of our second benchmark test when $p = 1$ with LLW’s in the Monte Carlo section.

In this paper, we use kernel estimation techniques to define test statistics, and show that they converge to simple distributions related to changed time Brownian Bridge. Critical values for our benchmark test statistics discussed above have explicit formulas and are easy to compute. Higher order stochastic dominance extensions of the tests in section 3 involves simulated critical values. All our test statistics are of Kolmogorov-Smirnov type tests \(^1\) based on maximal differences over the support of the distribution.

The remainder of the paper is organized as follows. In section 2, we propose test statistics and their asymptotic properties for the two benchmark tests with nulls discussed above. In section 3 we consider less robust but $\sqrt{n}$ consistent semiparametric tests as extensions to the benchmark tests. We also extend the tests to study higher order stochastic dominance of marginal effect functions. In section 4, we conduct Monte Carlo experiments to examine the small sample behavior of the proposed tests. In section 5 we apply our test to the empirical problem of child gender and parental labor income.

2 The Benchmark Tests: Statistics and Asymptotic Properties

In this paper, we only consider the case where we have independent data drawn from an identical distribution. The response variable must be continuous while the regressors could be continuous or discrete or a mix of continuous and discrete variables.

\(^1\)For advantages and disadvantages of Kolmogorov-Smirnov type tests compared to $L^2$-norm type statistics, please refer to McCaig and Yatchew (2007)[22]
Assumption 2.1

1. \{Y_i, X_i\}_{i=1}^n are independent observations randomly drawn from an identical joint distribution characterized by cumulative function \(G(y, x)\).

2. \(X = (X_1, X_2)\). \(Y, X\) and \(X_2\) have bounded supports \(Y = [0, \bar{z}]\), \(X \subseteq \mathbb{R}^p\) and \(W \subseteq \mathbb{R}^{p-1}\) respectively, \(\bar{z} < \infty\).

We are interested in testing whether \(X_1\) has uniformly negative stochastic marginal effect at \(X = x \in \mathcal{X}\). Both cases of discrete and continuous \(X_1\) variables will be studied. The conditioning \(X_2\) is not limited to continuous variables, however we will only discuss the case with continuous \(X_2\) since discrete variables can be treated by sample splitting. Though sample splitting is fully nonparametric it may result in a deterioration of the rate of convergence due to the shrinking of the sample size. We will discuss in the next section semiparametric extensions to get around sample splitting for discrete \(X_2\) elements.

2.1 Nonparametric Testing for Stochastic Marginal Effects: the Discrete Variable Case

First we study the uniform sign stochastic marginal effect test for a discrete \(X_1\) variable. Without loss of generality, assume further that \(X_1\) is a dummy variable. The stochastic marginal effect of \(X_1\) is equal to \(F_{Y|X_1=1, X_2=x_2} - F_{Y|X_1=0, X_2=x_2}\). The null hypothesis is formulated in the introduction section as \(H^1_{0,x_2}\).

Assume generally that \(Y = H(X, e)\) where \(H\) is an unknown structural function and \(e\) is the unobservable determinant of \(Y\). The interpretation of \(F_{Y|X_1=1, X_2=x_2} - F_{Y|X_1=0, X_2=x_2}\) in applications depends on the relationship between \(X\) and \(e\). The following assumption gives an instance where the marginal effect has a causal interpretation.

**Assumption A** Let \(Y(i) = H(X_1 = i, X_2, e)\), be the potential outcomes for \(X_1 = i\) where \(i = 0, 1\) and assume \((Y(0), Y(1)) \perp X_1 \mid X_2\).
Assumption A is the standard unconfoundedness assumption or selection on observables that has been commonly used in the treatment effect literature. Under Assumption A, the stochastic marginal effect of $X_1$ is equal to the stochastic conditional treatment effect (conditional on $X_2$) of the discrete $X_1$, which indicates whether a variable or policy improves or deteriorates the distribution of the response outcomes in a stochastic dominance sense.

$$
F_{Y|X_1=1,X_2=x_2} - F_{Y|X_1=0,X_2=x_2} = F_{Y(1)|X_1=1,X_2=x_2} - F_{Y(0)|X_1=0,X_2=x_2}
$$

$$
= F_{Y(1)|X_2=x_2} - F_{Y(0)|X_2=x_2}
$$

$$
= E[1(Y(1) \leq y) - 1(Y(0) \leq y)|X_2].
$$


Note that Assumption A is not required for the asymptotic results discussed in the rest of this section. It only provides a situation where the null and alternative have a direct structural interpretation. If it is not valid, the null hypothesis still gives information on the stochastic relationship between covariates and the response variable. Examples where this kind of stochastic relationship might be interesting include number of kids ($X_1$) versus parental labor income ($Y$), union status ($X_1$) versus workers’ salary, and etc.

Let $q = p - 1$ be the dimension of $X_2$. Assume the underlying joint distribution of $Y$ and $X$ satisfies the following conditions.

**Assumption 2.2**

1. $X_1 \in \{0, 1\}$, $P(X_1 = 1) \in (0, 1)$.

2. Let $G_i(y, x_2) = G(y, i, x_2)$ be continuous on $\mathbb{R}^{q+1}$, $q \geq 1$, and $s_i(y, x_2)$ be the probability distribution function corresponding to $G_i(y, x_2)$, $i = 0, 1$. 
3. Let \( g_i(y, x_2) = \int 1(u \leq y) s_i(u, x_2) du, \ i = 0, 1. \) \( g_i(y, x_2) \) is uniformly bounded up until the second derivative: \( \sup_{y,u} |g_i^{(j)}(y,u)| < \infty, \ i = 0, 1, \ j = 0, 1, 2. \)

4. Let \( f_i(x_2) = g_i(\infty, x_2), \ i = 0, 1. \) \( f_i(x_2) \) is uniformly bounded up until the second derivative: \( \sup_u |f_i^{(j)}(u)| < \infty, \ i = 0, 1, \ j = 0, 1, 2. \)

It is obvious that \( f_0, f_1 \) are probability density functions of \( X_2 \) at \( X_1 = 0 \) and \( 1. \) Let \( K : \mathbb{R}^q \rightarrow \mathbb{R} \) be the kernel function. Kernel density estimators of \( f_0 \) and \( f_1 \) are defined as follows,

\[
\hat{f}_{0n}(x_2) = \frac{1}{n_0 h_{0n}^q} \sum_{i=1}^{n} 1(X_{1i} = 0) K \left( \frac{X_{2i} - x_2}{h_{0n}} \right),
\]

\[
\hat{f}_{1n}(x_2) = \frac{1}{n_1 h_{1n}^q} \sum_{i=1}^{n} 1(X_{1i} = 1) K \left( \frac{X_{2i} - x_2}{h_{1n}} \right),
\]

where \( h_{0n}, h_{1n} \) are bandwidths of the two nonparametric estimators. Let \( s_0 \) and \( s_1 \) be joint probability density functions of \((Y, X_2)\) at \( X_1 = 0 \) and \( 1. \) \( g_i(y, x_2) = P(Y \leq y, X_i = i, X_2 = x_2), \ i = 0, 1. \) Define

\[
m_i(y|x_2) = \frac{g_i(y, x_2)}{f_i(x_2)} = F_{Y|X_1=i, X_2=x_2}(y|x_2),
\]

then \( m_i(y|x_2) \) is the conditional distribution of \( Y \) given \( X_2 = i \) and \( X_2 = x_2, \ i = 0, 1. \) Nadaraya-Waston kernel estimators of \( m_0(y|x_2) \) and \( m_1(y|x_2) \) are,

\[
\hat{m}_{0n}(y|x_2) = \frac{\sum_{i=1}^{n} 1(X_{1i} = 0) K \left( \frac{X_{2i} - x_2}{h_{0n}} \right) 1(Y_i \leq y)}{\sum_{i=1}^{n} 1(X_{1i} = 0) K \left( \frac{X_{2i} - x_2}{h_{0n}} \right)},
\]

\[
\hat{m}_{1n}(y|x_2) = \frac{\sum_{i=1}^{n} 1(X_{1i} = 1) K \left( \frac{X_{2i} - x_2}{h_{1n}} \right) 1(Y_i \leq y)}{\sum_{i=1}^{n} 1(X_{1i} = 1) K \left( \frac{X_{2i} - x_2}{h_{1n}} \right)}.
\]

For the fixed \( x_2 \) values in \( W, \) i.e. values with positive density, we propose the following test
statistic for $H_{0,x_2}^1$:

$$
\hat{S}_1(x_2) = A_1 \left( \frac{n_1 h_{1n}^q \hat{f}_{1n}(x_2) n_0 h_{0n}^q \hat{f}_{0n}(x_2)}{n_1 h_{1n}^q f_{1n}(x_2) + n_0 h_{0n}^q f_{0n}(x_2)} \right)^{\frac{1}{2}} \sup_{y \in Y} \left( \hat{m}_{1n}(y|x_2) - \hat{m}_{0n}(y|x_2) \right),
$$

where $A_1 = (\int K(\phi)^2 d\phi)^{-\frac{1}{2}}$ is a constant term that adjusts with the choice of kernel function. When $q = 1$, $A_1 = (5/3)^{\frac{1}{2}}$ for the Epanechnikov kernel function and $(2\sqrt{\pi})^{\frac{1}{2}}$ for the Gaussian kernel function. We will be characterizing the limiting distribution of $\hat{S}_1(x_2)$ under the null using the asymptotic properties of kernel based conditional empirical process developed in the statistics literature. See, among many others, Stute (1986, 1986a) [32] [31] and Horvath and Yandell (1988) [16].

**Assumption 2.3** Suppose we have a bounded symmetric kernel that satisfy:

1. $\int K(\phi) d\phi = 1$, $\int \phi K(\phi) d\phi = 0$;
2. $\sup |K(\phi)| < \infty$, $\int |K(\phi)| d\phi < \infty$, $\int |\phi K(\phi)| d\phi < \infty$;
3. $\mu_2 = \int \phi^2 K(\phi) d\phi < \infty$

**Assumption 2.4** The two bandwidths satisfy the following conditions:

1. $n_0 h_{0n}^q$, $n_1 h_{1n}^q \to \infty$ as $n \to \infty$;
2. $n_0 h_{0n}^{q+4}$, $n_1 h_{1n}^{q+4} \to 0$ as $n \to \infty$.

Note that Assumption 2.4 implies undersmoothing. The optimal bandwidth in the Asymptotic Mean Integrated Squared Error (AMISE) sense, which is of order $O \left( n^{-\frac{1}{q+4}} \right)$ (see for example Pagan and Ullah (1999) [25]), does not satisfy the above criteria since it will result in some bias in the centering for the limiting distribution of kernel based empirical processes. Bandwidth’s that satisfy this assumption should be of order $O \left( n^{-\frac{1}{q+4}+k} \right)$ for some small positive $k$ value. These requirements
for the bandwidth are the same as those for asymptotic normality of pointwise kernel regression estimators.

Let $D[0, \bar{z}]$ denote the space of all cadlag functions ("right continuous with left limits") on $[0, \bar{z}]$, and $\mathcal{B}(D[0, \bar{z}])$ the generated Borel $\sigma$ field. For fixed $x_2$, both $\hat{m}_{0n}(\cdot|x_2)$ and $\hat{m}_{1n}(\cdot|x_2)$ are random elements in $(D[0, \bar{z}], \mathcal{B}(D[0, \bar{z}]))$.

**Theorem 2.1** Under Assumption 2.1-2.4, we have that

$$\left[ n_0 h_{0n}^q \right]^{1/2} (\hat{m}_{0n}(\cdot|x_2) - m_0(\cdot|x_2)) \Rightarrow A_1^{-1} f_0(x_2)^{-1/2} \mathcal{B} (m_0(\cdot|x_2))$$

in $D([0, \bar{z}])$.

Here $\mathcal{B}(\cdot)$ is a standard Brownian Bridge Process on the unit interval $[0,1]$. Let $C([0, \bar{z}])$ denote the space of all continuous functions on $[0, \bar{z}]$. Then $P[\mathcal{B} (m_0(\cdot|x_2)) \in C([0, \bar{z}])] = 1$. A corresponding result holds for the process concerning the $X_1 = 1$ observations. We know from theorem 2.1 that, for fixed $x_2$ values, conditional empirical process sequences $\hat{m}_{0n}(\cdot|x_2)$ and $\hat{m}_{1n}(\cdot|x_2)$, after proper centering, converge to scaled changed time Brownian Bridge processes at rate $\sqrt{n_0 h_{0n}^p}$ and $\sqrt{n_1 h_{1n}^p}$ respectively. Notice that $\text{Var}[\mathcal{B} (m(y|x_2))] = m(y|x_2)[1 - m(y|x_2)] = \text{Var}(1(Y \leq y)|X_2 = x_2)$. If we fix the $y$ value, Theorem 2.1 degenerates to the standard asymptotic distribution of the pointwise Nadaraya-Waston estimator for $E[1(Y \leq y)|X_2 = x_2]$. The variance of asymptotic stochastic process is influenced by the density at the fixed regressor value $x_2$, a property inherited from the nature of Nadaraya-Waston estimators. The smaller is the density, the larger is the variance. Therefore, although the test is theoretically applicable to all fixed regressor values that have positive density, it is more precise if a fixed value in the high density region is chosen in empirical studies with limited sample size.

The proof of Theorem 1 follows the idea in Horvath and Yandell (1988) [16], who showed a stronger asymptotic result for the conditional empirical process with dimension $q = 1$. Since
\[ f_n(x_2) \overset{p}{\rightarrow} f_1(x_2), \ f_0n(x_2) \overset{p}{\rightarrow} f_0(x_2) \] under Assumption 2.1-2.4, we get the following Corollary.

**Corollary 2.1** Under Assumption 2.1-2.4, we have that

\[ A_1[n_0h_0^2f_0n(x_2)]^{1/2}(\hat{m}_0n(.,|x_2) - m_0(.,|x_2)) \Rightarrow \mathcal{B}(m_0(.,|x_2)) \]

in \( D([0, \tilde{z}]) \).

Again, a corresponding result holds for the process concerning the \( X_1 = 1 \) observations. Define test decision rule as

"reject \( H_{0,x_2}^1 \) if \( \hat{S}_1(x_2) > c_1 \)

where \( c_1 \) is critical value of the first benchmark test that we will discuss in the following. The convergence results discussed above give the following proposition that characterizes the properties of the first benchmark test.

**Proposition 2.1** Given Assumption 2.1-2.4 and that \( c_1 \) is a positive finite constant, we have:

1. If \( H_{0,x_2}^1 \) is true, \( \lim_{n \to \infty} P(\text{reject } H_{0,x_2}^1) \leq P(\sup_t \mathcal{B}(t) > c_1) = \exp(-2c_1^2) \),

with the equality holds when the equality in \( H_{0,x_2}^1 \) holds.

2. If \( H_{0,x_2}^1 \) is false, \( \lim_{n \to \infty} P(\text{reject } H_{0,x_2}^1) = 1 \)

The inequality in Proposition 2.1.1 implies that the test will never reject more often than \( P(\sup_t \mathcal{B}(t) > c_1) \) if the null hypothesis is true. The probability of rejection will be asymptotically equal to \( P(\sup_z \mathcal{B}(z) > c_1) \) if and only if \( m_1(.,|x_2) = m_0(.,|x_2) \) for the fixed \( x_2 \in \mathcal{W} \). The null is rejected with probability converging to 1 when it is not true. As is well known in the literature (see McFadden (1989) [23] for instance) that \( P(\sup_t \mathcal{B}(t) > c_1) = \exp(-2c_1^2) \), the p-value of \( \hat{S}_1 \) is \( \exp(-2(\hat{S}_1)^2) \). Then the null hypothesis is rejected if the p-value obtained is larger than \( \alpha \) which is the nominal size of the test. The decision rule can also be stated in terms of critical values. Some
standard critical values are 1.073 for the 10% significance level, 1.2239 for the 5% and 1.5174 for the 1%.

It is important to make clear that the convergence rate of our test statistic is slower than $\sqrt{n}$ and is dependent on the bandwidth because the Nadaraya Waston pointwise estimator we use to construct the test statistic relies heavily on information in a small window around the focused point. LW’s test enjoys $\sqrt{n}$ convergence rate because they study whether the sign of stochastic marginal effect of $X_1$ is uniform for all $X_2$ values and hence construct the test statistic using all data points of $X_2$. However, we argue that although the convergence rate of their test does not deteriorate with the increase of regressor dimension, it seems that it might be implausible to expect their null hypothesis to be satisfied in practical applications where there are many covariates.

2.2 Nonparametric Testing for Stochastic Marginal Effects: the Continuous Variable Case

In this section, we propose a test to decide whether the stochastic marginal effect of a continuous variable is uniformly positive or negative given a dataset. Assume the joint density of $(Y, X)$ satisfies the following smoothness conditions, which are slightly stronger than those in section 2.1 as we will see that the test statistic in this section includes kernel derivative estimators.

Assumption 2.5

1. Let $G(y, x)$ be continuous on $\mathbb{R}^{p+1}$, $q \geq 1$, and $s(y, x)$ be the corresponding probability distribution function.

2. Let $g(y, x) = \int 1(u \leq y)s(u, x)du$ be uniformly bounded up until third derivatives: 
   $$\sup_{y,u}|g^{(j)}(y, u)| < \infty, \quad j = 0, 1, 2, 3.$$ 

3. Let $f(x) = g(\infty, x)$ be uniformly bounded up until third derivatives: $\sup_u |f^{(j)}(u)| < \infty$, 
   $j = 0, 1, 2, 3$. 

11
Let \( m(y|x) = \frac{g(y,x)}{f(x)} = F_{Y|X=x}(y|x) \) be the conditional distribution of \( Y \) given \( X = x \). Define
\[
m^{(1)}(y|x) = \frac{\partial}{\partial x_1} m(y|x) = \frac{\partial}{\partial X_1} F_{Y|X=x}(y|x).
\]
Then \( m^{(1)}(y|x) \) is the stochastic marginal effect of the continuous \( X_1 \). The null hypothesis on whether \( X_1 \) has uniform negative stochastic marginal effect is formulated in the introduction section as \( H_{0,x}^2 \).

Suppose \( Y = H(X,e) \) where \( H \) is an unknown structural function and \( e \) is the unobservable determinants of \( Y \). Similar to the previous test, interpretation of \( \frac{\partial}{\partial X_1} F_{Y|X}(y|x) \) in the null depends on the relationship between \( X \) and \( e \).

**Assumption B** \( e \perp X_1 \mid X_2 \).

**Assumption C** \( e \perp X \).

Assumption B is the continuous counterpart of the unconfoundedness assumption in section 2.1. Under Assumption B, the partial derivative of \( F_{Y|X} \) again equals to the stochastic treatment effect.
\[
\frac{\partial}{\partial X_1} F_{Y|X}(y|X) = \frac{\partial}{\partial X_1} \int 1[H(X,e) \leq y]dF_{e|X} = \frac{\partial}{\partial X_1} \int 1[H(X,e) \leq y]dF_{e|X_2}
\]
\[
= \int \frac{\partial}{\partial X_1} 1[H(X,e) \leq y]dF_{e|X_2} = \int \frac{\partial}{\partial X_1} 1[H(X,e) \leq y]dF_{e|X}
\]
\[
= E \left[ \frac{\partial}{\partial X_1} 1(Y \leq y) \mid X \right].
\]
The second and fourth equality holds from Assumption B. The third holds from boundedness of indicator function.
Under Assumption C, the partial derivative of $F_{Y|X}$ equals to the average partial effect of $X_1$ on $1(Y \leq q)$ (averaging over $e$). We call it \textit{stochastic partial effect} of $X_1$.

$$
\frac{\partial}{\partial X_1} F_{Y|X}(y|X) = \frac{\partial}{\partial X_1} \int 1[H(X, e) \leq y]dF_{e|X} = \frac{\partial}{\partial X_1} \int 1[H(X, e) \leq y]dF_e \\
= \int \frac{\partial}{\partial X_1} 1[H(X, e) \leq y]dF_e
$$

The stochastic treatment effect averages the effect of $X_1$ over the conditional distribution of $e$ given $X$ while the stochastic partial effect averages the effect over marginal distribution of $e$. Therefore, although both effects gives causal meanings, the former relies on the sample distribution while the latter is distribution free. Also notice that neither Assumption B nor C affects the asymptotic properties of the test statistic to be discussed. They only provide situations where the null and alternative have a direct structural interpretation. If neither assumption is valid, the stochastic marginal effect still gives information in the relationship between the covariate and the response variable. Interesting economic relationships where stochastic marginal effect could be of interest include other family income versus father’s labor income, sons’ wealth versus parents’ wealth and vehicle mpg versus annual mileage. Following the definitions in the previous section, let $\hat{f}_n(x), \hat{m}_n(y|x)$ be the kernel estimators of $f(x)$ and $m(y|x)$. Define $\hat{f}^{(1)}_n(x), \hat{g}^{(1)}_n(y, x)$ as kernel derivative estimators of $f(x)$ and $g(y, x)$ with respect to $x_1$.

$$
\hat{f}^{(1)}_n(x) = -\frac{1}{nh_n^{p+1}} \sum_{i=1}^{n} K_1^{(1)} \left( \frac{X_i - x}{h_n} \right), \\
\hat{g}^{(1)}_n(y, x) = -\frac{1}{nh_n^{p+1}} \sum_{i=1}^{n} K_1^{(1)} \left( \frac{X_i - x}{h_n} \right) 1(Y_i \leq y).
$$

Here $K : \mathbb{R}^p \rightarrow \mathbb{R}$ is the kernel function, $K_1^{(1)}$ is the partial derivative of $K$ with respect to its first argument and $h_n$ is the bandwidth that goes to zero when $n$ goes to infinity. The kernel function $K$ and the bandwidth $h_n$ are assumed to follow the conditions in below.
**Assumption 2.6** Let $K$ be a symmetric kernel and $K^{(1)}_1$ its partial derivative with respect to the first argument. The kernel function satisfies:

1. $\int K(\phi)d\phi = 1$, $\int \phi K(\phi)d\phi = 0$;
2. $\sup |K(\phi)| < \infty$, $\int |K(\phi)|d\phi < \infty$, $\int |\phi K(\phi)|d\phi < \infty$;
3. $\mu_2 = \int \phi^2 K(\phi)d\phi < \infty$
4. $\sup |K^{(1)}_1(\phi)| < \infty$, $\int |K^{(1)}_1(\phi)|d\phi < \infty$.

**Assumption 2.7** The bandwidth is assumed to satisfy:

1. $nh_n^{p+2} \to \infty$ as $n \to \infty$;
2. $nh_n^{p+6} \to 0$ as $n \to \infty$.

Under Assumption 2.5-2.7, Vinod and Ullah’s (1988) [34] kernel derivative estimator

$$\hat{m}_n^{(1)}(y|x) = \hat{f}_n(x)^{-1} \left[ \hat{g}_n^{(1)}(y, x) - \hat{m}_n(y|x) \hat{f}_n^{(1)}(x) \right]$$

is a consistent estimator of $m^{(1)}(y|x)$ with convergence rate $\sqrt{nh_n^{p+2}}$ for all $y \in Y$. Actually the bandwidth assumption in 2.7 is stronger than that is necessary for consistency. Undersmoothing is required so that the bias resulted from kernel derivative estimation will vanish with the increase of sample size. Pagan and Ullah (1999) [25] has a detailed discussion about various kernel derivative estimators and their properties including the Vinod and Ullah estimator that we are using.

For fixed $x$ values satisfying the positive density assumption in 2.5.3, we propose the test statistic for the null hypothesis $H_{0,x}^2$ as:

$$\hat{S}_2(x) = A_2 \left[ nh_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \sup_{y \in Y} \hat{m}_n^{(1)}(y|x),$$
where \( A_2 = \left( \int K_1^{(1)}(\phi)^2d\phi \right)^{-\frac{1}{2}} \). When \( p = 1 \), \( A_2 = \left( \frac{3}{\pi} \right)^{\frac{1}{2}} \) for the Epanechnikov kernel function and \( (4\sqrt{\pi})^{\frac{1}{2}} \) for the Gaussian kernel function.

For fixed \( x \) values, we know that \( \hat{m}_n^{(1)}(.|x) \) is a random element in \( (D[0, \bar{z}], \mathcal{B}(D[0, \bar{z}])) \). The following result characterizes the asymptotic behavior of the process \( \hat{m}_n^{(1)}(.|x) \).

**Theorem 2.2** Under Assumption 2.1 and 2.5-2.7,

\[
\left[ n h_n^{p+2} \right]^{\frac{1}{2}} \left( \hat{m}_n^{(1)}(.|x) - m^{(1)}(.|x) \right) \Rightarrow A_2^{-1} f(x)^{-\frac{1}{2}} \mathcal{B}(m(.|x))
\]

in \( D([0, \bar{z}]) \).

Theorem 2.2 degenerates to the standard asymptotic distribution of pointwise kernel derivative estimator of \( E[1(Y \leq y)|X = x] \) if we fix the \( y \) value. The rate of convergence here is smaller than that in Theorem 2.1 because, for any fixed \( y \), kernel derivative estimators converge slower than kernel regression estimators. The higher is the order of the derivative, the slower is the convergence rate. Since \( \hat{f}_n(x) \overset{p}{\to} f(x) \) under the assumptions, we have the following corollary:

**Corollary 2.2** Under Assumption 2.1 and 2.5-2.7,

\[
A_2 \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \left( \hat{m}_n^{(1)}(.|x) - m^{(1)}(.|x) \right) \Rightarrow \mathcal{B}(m(.|x)).
\]

in \( D([0, \bar{z}]) \).

Let \( c_2 \) be the critical value of the first benchmark test. We define decision rule of the test as

“reject \( H_{0,x}^2 \) if \( \hat{S}_2(x) > c_2 \).”

From Corollary 2.2, we find that our second benchmark test shares the same rules for calculating p-values and critical values with the first test.
Proposition 2.2 Given Assumption 2.1 and 2.5-2.7 and that $c_2$ is a positive finite constant, we have:

1. If $H_{0,x}^2$ is true, $\lim_{n \to \infty} P(\text{reject } H_{0,x}^2) \leq P(\sup_z \mathfrak{B}(z) > c_2) = \exp(-2c_2^2)$, with the equality holds when the equality in $H_{0,x}^2$ holds.

2. If $H_{0,x}^2$ is true, $\lim_{n \to \infty} P(\text{reject } H_{0,x}^2) = 1$.

3 Some Useful Extensions

3.1 Single Index Models

Although the above two benchmark tests are quite robust (test statistics and decision rules are provided as long as the response variable is continuous and its joint distribution with the regressors satisfies modest assumptions), they suffer from “curse of dimensionality” just as all other non-parametric analysis methods do. In this section, we consider semiparametric extensions for both benchmark tests.

Assumption 3.1 Let $\theta$ be a $p$ dimensional parameter vector lying in the parameter space $\Theta$.

1. The conditional distribution of $Y$ given $X$ is of some single index form, or $F_{Y|X=x}(\cdot|x) = \tilde{m}(\cdot|x\theta)$.

2. Let $\tilde{G}(y,x\theta)$ be the cumulative distribution function of $(Y,X\theta)$. $\tilde{G}(\cdot)$ is continuous on $\mathbb{R}^2$.

The continuity assumption in 3.1.2 only requires that at least one element of $X$ is continuous. It is much weaker than its counterpart in Section 2. Therefore, for most empirical situations with discrete or mixed conditioning variables, sample splitting is no longer necessary under the single index setup. One sufficient condition for the single index assumption to hold is that the cumulative distribution function $G(y,x)$ of $(Y,X)$ in Assumption 2.1 equals to $\tilde{G}(y,x\theta)$.

Let $X\theta = X_1\theta_1 + X_2\theta_2$. Denote $\tilde{m}_1(\cdot|x_2\theta_2) = \tilde{m}(\cdot|\theta_1 + x_2\theta_2)$ and $\tilde{m}_0(\cdot|x_2\theta_2) = \tilde{m}(\cdot|x_2\theta_2).$ Under the single index assumptions, the stochastic marginal effect of $X_1$ equals to $\tilde{m}_1(\cdot|x_2\theta_2) - \tilde{m}_0(\cdot|x_2\theta_2)$.
when it is a dummy variable and $\frac{\partial}{\partial x_1} \tilde{m}(y|x\theta)$ when it is a continuous variable. When $X_1$ is a dummy variable, we normalize the first argument of $\theta_2$ to 1. When $X_1$ is continuous, we normalize the first argument of $\theta$ to 1. Then, the null hypothesis that we are interested in becomes:

$$H_{0,x_2}^3: \text{For fixed } x_2 \in \mathcal{W}, \tilde{m}_0(y|x_2\theta_2) \leq \tilde{m}_1(y|x_2\theta_2) \text{ for all } y \in \mathcal{Y},$$

$$H_{0,x}^4: \text{For fixed } x \in \mathcal{X}, \frac{\partial}{\partial x_1} \tilde{m}(y|x\theta) \leq 0 \text{ for all } y \in \mathcal{Y}.$$

When the true values of the coefficient parameters are known, our previous benchmark tests could be easily applied by substituting $X_2\theta_2$ ($X\theta$) for $X_2$ ($X$). When the coefficients are not known, a two step testing approach could be used where $\theta_2$ ($\theta$) must be first estimated up to scale. Hall and Yao (2005) proposed an iterative estimator that is $\sqrt{n}$ consistent and enables researchers to approximate conditional distributions using the single index dimension reduction technique. Meanwhile, given the fact that under Assumption 3.1 $E(Y|X = x) = \int yd\tilde{m}(y|x\theta)$ is in the single index form, we could also use the non-iterative (and hopefully easier to compute) semiparametric identification estimators provided by single index mean estimation literatures. Among many are non-iterative average partial derivative estimator developed by Hardle and Stoker (1989) [14] and Powell, Stock, and Stoker (1989) [26] for coefficients of continuous variables, by Horowitz and Hardle (1996) [15] for coefficients of dummy variables.

**Assumption 3.2** The estimator $\hat{\theta}$ is $\sqrt{n}$ consistent, that is for all $\theta$ in a compact space $\Theta$, $\hat{\theta} - \theta = O_p(n^{-\frac{1}{2}})$.

It is easy to conjecture that the use of the $\sqrt{n}$ consistent estimator $\hat{\theta}$ will not change the asymptotic properties of the test statistics as the estimator has faster rate of convergence than the nonparametric test statistic itself.

**Corollary 3.1** Under the additional assumption 3.1, results in Proposition 2.1 and 2.2 could be
applied to the following test statistics respectively:

\[ \hat{S}_3(x_2) = A_1 \left[ \frac{n_1 h_1 f_1(x_2 \hat{\theta}_2) n_0 h_0 f_0(x_2 \hat{\theta}_2)}{n_1 h_1 f_1(x_2 \hat{\theta}_2) + n_0 h_0 f_0(x_2 \hat{\theta}_2)} \right]^{\frac{1}{2}} \sup_{y \in Y} \left( \hat{m}_{1n}(y|x_2 \hat{\theta}_2) - \hat{m}_{0n}(y|x_2 \hat{\theta}_2) \right), \text{ and} \]

\[ \hat{S}_4(x) = A_2 \left[ n h_3 f_3(x \hat{\theta}) \right]^{\frac{1}{2}} \sup_{y \in Y} \hat{m}_{n}^{(1)}(y|x \hat{\theta}). \]

Sequences \( \hat{f}_n, \hat{f}_1, \hat{m}_0, \hat{m}_1 \), and \( \hat{m}_n^{(1)} \) are defined the same as in the previous sections with the independent variables replaced by the single index \( X_2 \hat{\theta}_2 \) or \( X \hat{\theta} \) respectively. The single index assumption in Assumption 3.1 reduces the dimension in test statistics. The bandwidth requirements are the same as in the Assumption 2.4 and 2.7 with dimension \( q = 1 \). This dimension reduction increases efficiency of the tests given the validity of the single index assumption.

### 3.2 Higher Order Stochastic Dominance Tests

Sometimes, first order stochastic dominance tests could be too restrictive if policy makers care a lot about the stochastic marginal effect at the lower end of the distribution of the response variable. Define \( \mathcal{I}_j(y; h(u|x)) \) as the function that integrates \( h(u|x) \) to order \( j - 1 \) with respect to \( u \) up to value \( y \):

\[ \mathcal{I}_1(y; h(u|x)) = h(y|x), \]

\[ \mathcal{I}_2(y; h(u|x)) = \int_0^y h(u|x)du, \]

\[ \mathcal{I}_3(y; h(u|x)) = \int_0^y \int_0^s h(u|x)du ds, \]

...
Then, depending on whether $X_1$ is discrete or continuous, the null hypothesis of interest becomes:

\[ H^5_{0,x_2} : \text{For fixed } x_2 \in W, I_j(y; F_{Y|X_1=1, X_2=x_2}(u|x_2)) \leq I_j(y; F_{Y|X_1=0, X_2=x_2}(u|x_2)) \text{ for all } y \in \mathcal{Y}, \]
\[ H^6_{0,x} : \text{For fixed } x \in \mathcal{X}, \frac{\partial}{\partial x_1} I_j(y; F_{Y|x=x}(u|x)) \leq 0 \text{ for all } y \in \mathcal{Y}. \]

Define test statistics:

\[
\hat{S}_5(x_2) = A_1 \left( \frac{n_1 h_{1n}^q \hat{f}_{1n}(x_2) n_0 h_{0n}^q \hat{f}_{0n}(x_2)}{n_1 h_{1n}^q \hat{f}_{1n}(x_2) + n_0 (h_{0n})^q \hat{f}_{0n}(x_2)} \right)^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} \{ I_j(y; \hat{m}_{1n}(u|x_2)) - I_j(y; \hat{m}_{0n}(u|x_2)) \},
\]
\[
\hat{S}_6(x) = A_2 \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \sup_{y \in \mathcal{Y}} I_j(y; \hat{m}_n^{(1)}(u|x))
\]

By Theorem 2.1, 2.2, the Continuous Mapping Theorem and the fact that the integral function $I_j(.)$ is continuous, the following corollary summarizes the asymptotic properties of the above integrated processes.

**Corollary 3.2**

\[
A_1 \left[ n_0 h_{0n}^q \hat{f}_{0n}(x_2) \right]^{\frac{1}{2}} \{ I_j(y; \hat{m}_{0n}(u|x_2)) - I_j(y; m_0(u|x_2)) \} \Rightarrow I_j(y; \mathfrak{N}(m_0(u|x_2))),
\]
\[
A_2 \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \{ I_j(y; \hat{m}_n^{(1)}(u|x)) \} \Rightarrow I_j(y; \mathfrak{N}(m(u|x))).
\]

in $C([0, \bar{z}])$.

A corresponding result holds for the process concerning the $X_1 = 1$ observations as well. For these higher order stochastic dominance tests, the asymptotic distributions of test statistics depend on the underlying conditional distributions of the response variables. In the following, we discuss a simulation approach for obtaining p-values. It involves the use of artificial random numbers and exploits the construction of Wiener process with random change of time, discussed in Billingsley (1999) [5] on page 146-154, to simulate a process that is identical to but (asymptotically) inde-
pendent of $\mathfrak{B} \circ m(.|x)$. Let $U^n_{i=1}$ denote a sequence of i.i.d. $N(0, 1)$ random variables that are independent of the samples. We denote the simulated process by the notation $\mathfrak{B}^* \circ \hat{m}_n(.|x)$ and use $\mathfrak{B}^* (\hat{m}_n(y|x))$ to represent the process $\mathfrak{B}^* \circ \hat{m}_n(.|x)$ evaluated at the point $y \in \mathcal{Y}$. Then the process is generated by letting

$$\mathfrak{B}^* (\hat{m}_n(y|x)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1(i \leq n \times \hat{m}_n(y|x)) - \hat{m}_n(y|x)) U_i.$$ 

The p-values of our higher order stochastic dominance tests of the marginal effects could be obtained from appropriate functionals of the simulated processes:

$$\hat{p}_{5,j} = P_U \left[ \sup_y I_j(y; \mathfrak{B}^* \circ \hat{m}_0n(.|x)) > \hat{S}_5(x_2) \right],$$

$$\hat{p}_{6,j} = P_U \left[ \sup_y I_j(y; \mathfrak{B}^* \circ \hat{m}_n(.|x)) > \hat{S}_6(x) \right],$$

where $P_U(.)$ is the probability function associated with the normal random variables $U_i$ and is, in real calculation, equal to the fraction of simulations with the interested supremum functional larger than the test statistics. The standard errors of the simulated p values are related to the number of simulations performed. See Barrett and Donald (2002) [3] or Hansen (1996) [13] for discussions on simulation number and standard errors of the simulated p values. The following results describe the asymptotic characteristics of higher order stochastic dominance tests and the simulated p-values.

**Proposition 3.1** Given Assumption 2.1-2.4 (2.1, 2.5-1.7) and assuming that $\alpha < \frac{1}{2}$, a test for $j$-th order stochastic dominance test for discrete (continuous) marginal effect based on the rule

"reject $H_{0,x_2}^{5,j}(H_{0,x}^{6,j})$ if $\hat{p}_{5,j}(\hat{p}_{6,j}) < \alpha$"

satisfies the following:

1. If $H_{0,x_2}^{5,j}(H_{0,x}^{6,j})$ is true, $\lim_{n \to \infty} P \left[ \text{reject } H_{0,x_2}^{5,j}(H_{0,x}^{6,j}) \right] \leq \alpha,$
with the equality holds when the equality in $H_{0,x}^{5,j}(H_{0,x}^{6,j})$ holds.

2. If $H_{0,x}^{5,j}(H_{0,x}^{6,j})$ is false, $\lim_{n\to\infty} P[\text{reject } H_{0,x}^{5,j}(H_{0,x}^{6,j})] = 1$.

4 Monte Carlo Results

First we conduct a set of experiments to study the small sample behavior of our tests for the stochastic marginal effect of discrete variables. For each experiment, independent variable $X_1$ is randomly drawn to take value 0 and 1 both with probability $1/2$ while $X_2, X_3 \sim N(0, 0.3^2)$. To evaluate test performance when the null hypothesis of uniformly negative stochastic marginal effect is true and the equality in null is satisfied, the response variable $Y$ is generated independently from $N(0, 0.3^2)$. The stochastic marginal effect of $X_1$ hence equals to zero at any $X_2, X_3$ values. We test the null that the stochastic marginal effect of $X_1$ is uniformly negative with the effect evaluated at $(X_2, X_3) = (0, 0)$. To see test performance when null hypothesis is false, $Y$ is generated following a simple random coefficient model $Y = X_1b - X_1(X_2 + X_3)$, where $b \sim U[0, 1]$ is the random coefficient. The stochastic marginal effect is NOT uniformly smaller than 0 when $X_2 + X_3 > 0$. We test the null that the stochastic marginal effect of $X_1$ is uniformly negative with the effect evaluated at $(X_2, X_3) = (0.2, 0.2)$.

For each of the two data generating processes (DGPs), we consider three sets of conditioning variables in testing: the single dimensional conditioning variable $SI = X_2 + X_3$, the two dimensional conditioning variable set $(X_2, X_3)$ and the single index conditioning variable $\hat{SI} = X_2 + X_3\hat{b}$, where $\hat{b}$ is Powell and Stoker’s (1989) [26] up-to-scale weighted average derivative estimator obtained under the single index assumption in Section 3.1. In this section, we use the the . Guassian kernel is used for the estimator and bandwidth is $1.4n^{-\frac{1}{5}}\sigma$ following the optimal rule discussed in Powell and Stoker (1996) [27]. Epanechnikov kernel is used for the test and bandwidth are chosen to be $cn_1^{-\frac{1}{75}}\sigma$, $cn_1^{-\frac{1}{75}}\sigma$ for the first and third tests with one dimensional conditioning variable and $cn_2^{-\frac{1}{75}}\sigma$, $cn_2^{-\frac{1}{75}}\sigma$ for the second case. $\sigma$ is the standard deviation of the conditioning variables and the constant $c$ is allowed to take values 2, 2.5 or 3 to see whether the performance of the tests is
sensitive to the bandwidth choice.

Table 1 reports rejection proportions of each tests using 5% confidence level and 2000 simulations. We see from the table that when the null is correct and equality in null is held, the rejection probability is close to 5% in the first two nonparametric testing cases but exhibits underrejection in the single index testing case. When the null is false, the rejection probability goes to 1 when sample size gets larger. Test results are not very sensitive to bandwidth choices. Comparing performance of the first two nonparametric tests we note that the rejection proportion goes to 1 slower when the dimension of conditioning variable gets larger. Therefore, larger sample size is needed for good test performance when the dimension of explanatory variable increases in empirical studies. Comparing performance of the single index test with the others we notice that the semiparametric coefficient estimator does not affect the test performance as is predicted by the theory when the null hypothesis is false: the rejection rate of the single index test converges to 1 obviously faster than its nonparametric counterpart.

Using the same DGPs, we could also study the small sample behavior of our second order stochastic dominance test for the stochastic marginal effects of discrete variables. When the response variable $Y$ is generated independently from $N(0, 0.3^2)$, the null hypothesis that the stochastic marginal effect function at $X_1 = 1$ second order stochastic dominates the function at $X_1 = 0$ is satisfied and the equality in null is held for any $(X_2, X_3)$ values. We again evaluate the effects at $(X_2, X_3) = (0, 0)$. When $Y$ is generated following the simple random coefficient model, the null hypothesis of second order stochastic dominance is violated if $X_2 + X_3 > 0$. We evaluate the effects at $(X_2, X_3) = (0.2, 0.2)$. We see from the table that the second order stochastic dominance tests also have good small sample behavior: when the null is correct and equality in null is held, the rejection probabilities are close to 5%; when the null hypothesis is false, the rejection rate converges to 1.

Then we study the small sample behavior of our second benchmark test concerning stochastic marginal effects of continuous variables. We also compare performance of our test to that of
Table 1: Rejection Proportions of Stochastic Marginal Effect Tests: Discrete Variable Case

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Conditioning Set 1</th>
<th>Conditioning Set 2</th>
<th>Conditioning Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c=2</td>
<td>c=2.5</td>
<td>c=3</td>
</tr>
<tr>
<td>N=100</td>
<td>0.040</td>
<td>0.046</td>
<td>0.050</td>
</tr>
<tr>
<td>N=250</td>
<td>0.043</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>N=500</td>
<td>0.047</td>
<td>0.048</td>
<td>0.048</td>
</tr>
</tbody>
</table>

First Order Stochastic Dominance Testing

When the null is true and equality in null held:

<table>
<thead>
<tr>
<th>Condition</th>
<th>N=100</th>
<th>N=250</th>
<th>N=500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.193</td>
<td>0.755</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>0.185</td>
<td>0.773</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>0.170</td>
<td>0.770</td>
<td>0.994</td>
</tr>
</tbody>
</table>

When the null is not true:

<table>
<thead>
<tr>
<th>Condition</th>
<th>N=100</th>
<th>N=250</th>
<th>N=500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.650</td>
<td>0.937</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.680</td>
<td>0.953</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.669</td>
<td>0.957</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Second Order Stochastic Dominance Testing

When the null is true and equality in null held:

<table>
<thead>
<tr>
<th>Condition</th>
<th>N=100</th>
<th>N=250</th>
<th>N=500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.064</td>
<td>0.052</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.062</td>
<td>0.052</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>0.064</td>
<td>0.059</td>
<td>0.053</td>
</tr>
</tbody>
</table>

When the null is not true:

<table>
<thead>
<tr>
<th>Condition</th>
<th>N=100</th>
<th>N=250</th>
<th>N=500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.650</td>
<td>0.937</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.680</td>
<td>0.953</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.669</td>
<td>0.957</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Simulation for our test is done 2000 times but only 200 times for LLW’s test due to the computational speed constraint of their test. Consider the DGP with \( X \sim U(0, 1), u \sim U(0, 0.3^2) \) and \( Y = Xb + 2(X - 0.25)^2 \). First we test whether the marginal effect of \( X \) on \( F_{u|x} \) is uniformly negative at \( X = 0.5 \) for our test, and whether it is uniformly negative for all \( X \) values for LLW’s test. The null hypothesis for both tests are true and equality in null held, therefore both tests shall reject the null with probability equaling to the confidence level 5%. Then we test the performance of both tests when the null hypotheses are false. We test whether the marginal effect of \( X \) on \( F_{Y|x} \) is uniformly negative at \( X = 0.1 \) for our test and whether it is uniformly negative for all \( X \) values for LLW’s test. Rejection rates for both tests shall go to 1 when the sample size increases.

Table 2 reports rejection proportions of the above Monte Carlo experiments using 5% confidence level. Bandwidths are chosen to be \( h = cn^{-6/7} \) with \( c=1, 1.25 \) or 1.5 in all experiments. We see that our tests perform as predicted. The rejection proportion results are not sensitive to bandwidth choices. The test performance of LLW’s test is almost as good as our test when the null is false but is not very stable when the null is true and equality in null is held.

Using the same DGPs, we also compare the stochastic marginal effects of \( X \) with zero in a second order stochastic dominance sense. The second order stochastic dominance test also has good small sample behavior: when the null is correct and equality in null is held, the rejection probability is close to 5%; when the null hypothesis is false, the rejection rate converges to 1. LLW’s test does not include a higher order stochastic dominance version.

In conclusion, our tests on stochastic marginal effects have reasonable small sample behaviors and the test results are not very sensitive to the bandwidth choice as long as the undersmoothing requirement is satisfied. For the nonparametric benchmark cases, the sample size needed for good rejection probability increases with the increase of regressor dimension. But if econometricians are willing to assume that the conditional cumulative distribution of the response variable is of single index form, they could use semiparametric tests that reduce the dimension of conditional variable and improve test performance in small samples.
Table 2: Rejection Proportions of Stochastic Marginal Effect Tests: Discrete Variable Case

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>c=1</th>
<th>c=1.25</th>
<th>c=1.5</th>
<th>LLIW1*</th>
<th>LLW2*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>c=1</td>
<td>c=1.25</td>
<td>c=1.5</td>
<td>c=1</td>
<td>c=1.25</td>
</tr>
<tr>
<td></td>
<td>Our Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>First Order Stochastic Dominance Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When the null is true and equality in null held:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=250</td>
<td>0.039</td>
<td>0.037</td>
<td>0.040</td>
<td>0.235</td>
<td>0.070</td>
</tr>
<tr>
<td>N=500</td>
<td>0.038</td>
<td>0.042</td>
<td>0.041</td>
<td>0.295</td>
<td>0.080</td>
</tr>
<tr>
<td>N=1000</td>
<td>0.041</td>
<td>0.046</td>
<td>0.037</td>
<td>0.345</td>
<td>0.155</td>
</tr>
<tr>
<td>When the null is not true:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=250</td>
<td>0.371</td>
<td>0.201</td>
<td>0.114</td>
<td>0.190</td>
<td>0.165</td>
</tr>
<tr>
<td>N=500</td>
<td>0.860</td>
<td>0.661</td>
<td>0.464</td>
<td>0.420</td>
<td>0.595</td>
</tr>
<tr>
<td>N=1000</td>
<td>0.998</td>
<td>0.990</td>
<td>0.963</td>
<td>0.950</td>
<td>0.990</td>
</tr>
<tr>
<td></td>
<td>Second Order Stochastic Dominance Testing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When the null is true and equality in null held:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=250</td>
<td>0.026</td>
<td>0.020</td>
<td>0.036</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>N=500</td>
<td>0.042</td>
<td>0.032</td>
<td>0.034</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>N=1000</td>
<td>0.062</td>
<td>0.048</td>
<td>0.038</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>When the null is not true:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=250</td>
<td>0.374</td>
<td>0.222</td>
<td>0.100</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>N=500</td>
<td>0.552</td>
<td>0.518</td>
<td>0.344</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>N=1000</td>
<td>0.898</td>
<td>0.836</td>
<td>0.750</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

*LLW1 calculates the rejection proportions using the asymptotic expansion approach while LLW2 calculates the rejection proportions using asymptotic distribution approach. See LLW (2009) [19] for details.
5 Empirical Example

In this section, we present an empirical example concerning children and father’s labor supply. We test on whether child gender and other family income have uniform stochastic marginal effects on father’s labor income. The data we use comes from Agrist and Evans (1998) [2], which is from the 5% Public Use Microdata Sample (PUMS) of 1990 Census. Our sample is limited to white households with working parents and 2 same sex kids less than 10 years old. ² Fathers are 30-31 years old at the time of survey. The kernel function used in this section is the same as that in the Monte Carlo experiments.

Let $X$ be other family income, $Z$ be a dummy equals to one when both kids are boys and $Y$ be father’s labor income. First we report marginal effects of kid gender on expected father labor income in Graph 1.a, i.e., $E[Y|X, Z = 1] - E[Y|X, Z = 0]$. The bandwidth are chosen to optimize AIMSE of the nonparametric estimations (see Pagan and Ullah (1999) Page 103[25]). The dashed lines are bootstrapped 95% confidence intervals of the nonparametric estimates. Lundberg and Rose (2002) [21] found that fathers’ wage rates increase more in response to the births of sons than to the births of daughters. However, we do not observe that fathers with two sons earn more on average: kid gender is found generally irrelevant to fathers’ labor income in our sample. Uniform sign tests for the stochastic marginal effect of child gender conditioning on fathers’ labor income are reported in Graph 2.a. We cannot reject the null that kid gender has uniformly positive stochastic marginal effect. Neither can we reject the null that it has uniformly negative stochastic marginal effect except at the right tail of other family income distribution. Bandwidth used for the tests is $2.5N^{-1/4.75}$.

Then we study the marginal effect and the stochastic marginal effect of other family income. Income effect and positive sorting of marriage are expected to play in the relationship between other family income and father’s labor income toward opposite directions. Graph 1.b shows that

²Observations with allocated age, income and years of schooling are excluded following Agrist and Evans (1998) [2].
the relationship between other family income and expected father’s labor income is not monotonic along the support of other family income. At larger quantiles of other family income, the marginal effect is positive, or the marriage sorting effect dominates. At smaller quantiles of other family income, the marginal effect is negative, or the income effect dominates. The first conclusion from the nonparametric mean estimation confirmed by stochastic marginal effect tests. We find that we cannot reject the null hypothesis that the probability of a father earning less than a certain amount $y$ decreases with other family income (i.e., marriage sorting effect dominates) for the large other family income values. Meanwhile, we do not find support for the second nonparametric mean estimation conclusion that negative income effect dominates for poorer households since we generally reject the null that the probability of a father earning less than certain amount $y$ increases with other family income. The rejection is especially predominant when the other family income takes large values. Bandwidth used for the tests is $3N^{-1/6.75}$. 

27
6 Appendix

Proof of Theorem 2.1

The proof generalizes part of Horvath and Yandel’ s (1988) asymptotic results on conditional empirical processes to higher dimensional case. First, define empirical distributions corresponding to $G_0(y, x_2)$ and $F_0(x_2)$.

$$
\hat{G}_{0n}(y, x_2) = \frac{1}{n} \sum_{i=1}^{n} 1(X_{1i} = 0)1(Y_i \leq y, X_{2i} \leq x_2); \quad \hat{F}_{0n}(y, x_2) = \hat{G}_{0n}(\infty, x_2) = \frac{1}{n} \sum_{i=1}^{n} 1(X_{1i} = 0)1(X_{2i} \leq x_2).
$$

Then define two sequences that we will use frequently in proof based on the empirical processes:

$$
\alpha^0_n(y, x_2) = n^{1/2} \left( \hat{G}_{0n}(y, x_2) - G_0(y, x_2) \right); \quad t^0_n(x_2) = n^{1/2} \left( \hat{F}_{0n}(x_2) - F_0(x_2) \right)
$$

To prove Theorem 1, it is sufficient to show that the asymptotic property holds for the $Z_1 = 0$ subsample. In the rest of the proof of Theorem 1, we suppress the underscipts denoting the subsample. By
Neuhaus (1971) [24] which extends Billingsley’s (1999) [5] (first edition in 1968) weak convergence results on \(D([0,1])\) to the space of all cadlag functions on \([0,1]^q\), we have the following lemmas.

**Lemma 1** \( t_n \Rightarrow \gamma, \) where \( \gamma \) is a centered Gaussian process on \([0,1]^q\) with covariance \( \text{cov}(\gamma(x_2), \gamma(x_2')) = F(x_2 \wedge x'_2) - F(x_2)F(x_2') \) for any \( x_2, x'_2 \in \mathcal{W} \).

where \( x_2 \wedge x'_2 = (x_2 \wedge x'_2, ..., x_2 \wedge x'_2) \) \((x_2 = (x_2, ..., x_2))\) and \( P[\gamma \in C] = 1. \)

**Lemma 2** \( \alpha_n \Rightarrow \tau, \) where \( \tau \) is a centered Gaussian process on \([0,1]^{q+1}\) with covariance \( \text{cov}(\tau(y, x_2), \tau(y', x_2')) = G(y \wedge y', x_2 \wedge x_2') - G(y, x_2)G(y', x_2') \) for any \((y, x_2), (y', x_2') \in \mathcal{Y} \times \mathcal{W}, P[\tau \in C] = 1. \)

Define,

\[
\hat{g}_n(y, x_2) = h_n^{-q} \int_{-\infty}^{\infty} K \left( \frac{u - x_2}{h_n} \right) d_u \hat{G}_n(y, u) = (nh_n^q)^{-1} \sum_{i=1}^{n} K \left( \frac{X_{2i} - x_2}{h_n} \right) 1(Y_i \leq y)
\]

\[
\hat{f}_n(x_2) = \hat{g}_n(\infty, x_2) = h_n^{-q} \int_{-\infty}^{\infty} K \left( \frac{u - x_2}{h_n} \right) d_u \hat{F}_n(u) = (nh_n^q)^{-1} \sum_{i=1}^{n} K \left( \frac{X_{2i} - x_2}{h_n} \right)
\]

\[
\hat{m}_n(y|x_2) = \hat{g}_n(y, x_2)/\hat{f}_n(x_2)
\]

\[
g_n(y, x_2) = h_n^{-q} \int_{-\infty}^{\infty} K \left( \frac{u - x_2}{h_n} \right) d_u G(y, u)
\]

\[
f_n(x_2) = g_n(\infty, x_2) = h_n^{-q} \int_{-\infty}^{\infty} K \left( \frac{u - x_2}{h_n} \right) dF(u)
\]

\[
m_n(y, x_2) = g_n(y, x_2)/f_n(x_2).
\]

Notice that \( g_n(y, x_2), f_n(x_2) \) are expectations of \( \hat{g}_n(y, x_2) \) and \( \hat{f}_n(x_2) \) since we have random samplings of \( X \) and \( Y \). Now, we first want to show that

**Lemma 3** Under Assumption 2.1-2.3 and 2.4.1,

\[
(nh_n^q)^{\frac{1}{2}} (\hat{m}_n(\cdot|x_2) - m_n(\cdot|x_2)) \Rightarrow \left[ \frac{\int K(\phi)^2 d\phi}{f(x_2)} \right]^{\frac{1}{2}} \mathbb{B}(m(\cdot|x_2))
\]

When we get Lemma 3, what left for showing Theorem 2.1 is that
imposing the bandwidth assumption in 2.4.2 gives us the result of Theorem 2.1.

by the smoothness requirement in Assumption in 2.2 and the kernel requirement in Assumption 2.3. Then

$$\sup_y \left| m_{(n)}(y|x_2) - m(y|x_2) \right| = \sup_y \left| \frac{h_n^2}{2f(x)} \mu_2 \left( m^{(2)}(y|x)f(x) + 2f^{(1)}(x)m^{(1)}(y|x) \right) + o(h_n^2) \right|$$

Proof of Lemma 3:

Define \( \beta_n(y|x_2) = (nh_n^q)^{\frac{1}{2}} (\tilde{m}_n(y|x_2) - m_{(n)}(y|x_2)) \). First we want to show that

(1)

$$\{ \beta_n(y, x_2), y \in \mathcal{Y} \} = \left\{ \frac{h_n^{\frac{q}{2}}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u) - \frac{h_n^{\frac{q}{2}} m(y|x_2)}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) dt_n(u), y \in \mathcal{Y} \right\}$$

Decompose \( \beta_n(y|x) \), we get

$$\beta_n(y|x_2) = (nh_n^q)^{\frac{1}{2}} \left[ \frac{\hat{g}_n(y, x_2) - g_n(y, x_2)}{f_n(x_2)} \right]$$

$$= (nh_n^q)^{\frac{1}{2}} \frac{\hat{g}_n(y, x_2) - g_n(y, x_2)}{f_n(x_2)} - (nh_n^q)^{\frac{1}{2}} \frac{g_n(y, x_2) \left( \hat{f}_n(x_2) - f_n(x_2) \right)}{f^2_n(x_2)}$$

$$\quad - (nh_n^q)^{\frac{1}{2}} \frac{\hat{g}_n(y, x_2) - g_n(y, x_2)}{f_n(x_2)} \left( \hat{f}_n(x_2) - f_n(x_2) \right) + (nh_n^q)^{\frac{1}{2}} \frac{g_n(y, x_2) \left( \hat{f}_n(x_2) - f_n(x_2) \right)^2}{f^2_n(x_2) f_n(x_2)}$$

Since \( f_n(x_2) \rightarrow f(x_2) \) and \( m_n(x_2) \rightarrow m(x_2) \) under Assumption 2.1-2.3 and 2.4.1, we have the first two parts of the expression

$$\left\{ \left( nh_n^q \right)^{\frac{1}{2}} \frac{\hat{g}_n(y, x_2) - g_n(y, x_2)}{f_n(x_2)} - \left( nh_n^q \right)^{\frac{1}{2}} \frac{g_n(y, x_2) \left( \hat{f}_n(x_2) - f_n(x_2) \right)}{f^2_n(x_2)} \right\}, y \in \mathcal{Y}$$

$$= \left\{ \frac{h_n^{\frac{q}{2}}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u) - \frac{h_n^{\frac{q}{2}} m(y|x_2)}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) dt_n(u), y \in \mathcal{Y} \right\}.$$
In the third part of the expression, we know that $(n h_n^q)^{\frac{1}{2}} \left( \hat{f}_n(x_2) - f_n(x_2) \right) = O_p(1)$ and that

\[
\sup_y \left| \hat{g}_n(y, x_2) - g_n(y, x_2) \right| = \sup_y \left| n^{-\frac{1}{2}} h_n^{-q} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u) \right|
\]

\[
= (nh_n^q)^{-\frac{1}{2}} \sup_y |h_n^{-\frac{q}{2}} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u)|
\]

\[
= \sup_y |h_n^{-\frac{q}{2}} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u)|
\]

\[
\overset{D}{=} C(n h_n^q)^{-\frac{1}{2}} \sup_y |W(m(y|x_2)|
\]

\[
p, 0,
\]

where $C$ is a constant and $W$ is a standard Wiener process. The third equality will be shown in a moment in (3), and the convergence is due to the fact that $P(\sup_{t \in [0,1]} |W(t)| < s) = P(|B(1)| < s)^2 = (2\Phi(s) - 1)^2$. Therefore, the third part goes to zero uniformly. The fourth part also goes to zero uniformly since $(n h_n^q)^{\frac{1}{2}} \left( \hat{f}_n(x_2) - f_n(x_2) \right) = O_p(1)$ and $\hat{f}_n(x_2) - f_n(x_2) = o_p(1)$. The result of (1) hence follows.

Let $H$ be the distribution function of $y$ and $I_j$ be the marginal distribution function of $X^{j}_2$, $j = 1, 2, ..., p$. Define Copula function of cumulative distribution function $G(y, x_2) = J(H(y), I(x_2))$, where $I(x_2) = (I^{1}_2, ..., I^{q}_2)$. Then the Gaussian process $\tau$ in Lemma 2 satisfies

\[
\{\tau(y, x_2), (y, x_2) \in \mathcal{Y} \times \mathcal{W}\} \overset{D}{=} \{W_J(H(y), I(x_2)) - G(y, x_2)W_J(1, \iota), (y, x_2) \in \mathcal{Y} \times \mathcal{W}\}
\]

where $\iota$ is a vector of $q$ ones, $W_J$ is a $q+1$ dimensional Wiener process with $E[W_J(s, t)] = 0$ and $E[W_J(s, t)W_J(s', t')] = J(s \wedge s', t \wedge t')$, for $(s, t), (s', t') \in \mathcal{Y} \times \mathcal{W}$.

Lemma 1 gives

\[
\left\{ \frac{h_n^{-\frac{q}{2}}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u), y \in \mathcal{Y}\right\} \overset{D}{=} \left\{ \frac{h_n^{-\frac{q}{2}}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u \tau(y, u), y \in \mathcal{Y}\right\}
\]

\[
\overset{D}{=} \left\{ \frac{h_n^{-\frac{q}{2}}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u W_J(H(y), I(u)) - \frac{h_n^{-\frac{q}{2}}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) g(y, u) du, y \in \mathcal{Y}\right\}
\]

Since $\sup_y \left| \int K \left( \frac{u - x_2}{h_n} \right) g(y, u) du \right| \leq h_n^{\alpha} \sup_{y, u} |g(y, u)| \int |K(\phi)| d\phi = O(h_n^q)$ under Assumptions 2.2.2. 

31
and 2.3, we have that
\[ \left\{ \frac{h_n^{-q}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u), y \in \mathcal{Y} \right\} \overset{D}{=} \left\{ \frac{h_n^{-q}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u W_J(H(y), I(u)), y \in \mathcal{Y} \right\}. \]

Calculation shows,
\[
E \left( \frac{h_n^{-q}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u W_J(H(y), I(u)) \right) = 0
\]
\[
E \left( \frac{h_n^{-q}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u W_J(H(y), I(u)) \int K \left( \frac{u - x_2}{h_n} \right) d_u W_J(H(y'), I(u)) \right)
= h_n^{-q} \int K \left( \frac{u - x_2}{h_n} \right)^2 g(y \wedge y', u) du
\]
\[
\overset{\text{let}}{=} l_{(n)}(y \wedge y')
\]

Then, we know from the Gaussian Characterization of the Brownian Motion that
\[ \left\{ \frac{h_n^{-q}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u W_J(H(y), I(u)), y \in \mathcal{Y} \right\} \overset{D}{=} \left\{ W(l_{(n)}(y)), y \in \mathcal{Y} \right\} \]
where \( W \) is a standard Wiener process.

Let \( l(y) = g(y, x_2) \int K(\phi)^2 d\phi \), we know that \( \sup_y |l_{(n)}(y) - l(y)| = O(h_n) \) under Assumption 2.2.2, 2.3 and 2.4.1. By Theorem 1.1.1 in Csorgo and Revesz (1981) [7], we know that
\[
\sup_y |W(l_{(n)}(y)) - W(l(y))| \overset{a.s.}{=} O(h_n \log \frac{1}{h_n}) \overset{a.s.}{\rightarrow} 0
\]
Together with the rescaling property of Brownian Motion, we have
\[ \left\{ \frac{h_n^{-q}}{f(x_2)} \int K \left( \frac{u - x_2}{h_n} \right) d_u \alpha_n(y, u), y \in \mathcal{Y} \right\} \overset{D}{=} \left\{ \left[ \int K(\phi)^2 d\phi \right]^{\frac{1}{2}} W(m(y|x_2)), y \in \mathcal{Y} \right\} \]
Likewise, we have
\[ \left\{ \frac{h_n^{-q}}{f(x_2)} m(y|x_2) \int K \left( \frac{u - x_2}{h_n} \right) dt_n(u), y \in \mathcal{Y} \right\} \overset{D}{=} \left\{ m(y|x_2) \left[ \int K(\phi)^2 d\phi \right]^{\frac{1}{2}} W(m(\infty|x_2)), y \in \mathcal{Y} \right\}. \]
By Lemma 2 we finally have that

\[ \{ \beta_n(y|x_2), y \in \mathcal{Y} \} \overset{D}{=} \left\{ \sqrt{\frac{\int K(\phi)^2 d\phi}{f(x_2)}} \left[ W(m(y|x_2)) - m(y|x_2)W(m(\infty|x_2)) \right], y \in \mathcal{Y} \right\}. \]

It is obvious that \( m(\infty|x_2) = 1 \). Therefore, we get the result of Lemma 3 by definition of the Brownian Bridge. Theorem 2.1 follows as is discussed.

**Proof of Theorem 2.2**

The proof is carried out using the same idea as that for Theorem 2.1. First define empirical distributions corresponding to \( G(y, x) \) and \( F(x) \) and two sequences that we will use frequently in proof based on the empirical processes:

\[ \hat{G}_n(y, x) = \frac{1}{n} \sum_{i=1}^{n} 1(Y_i \leq y, X_i \leq x); \quad \hat{F}_n(y, x) = \hat{G}_n(\infty, x) = \frac{1}{n} \sum_{i=1}^{n} 1(X_i \leq x). \]

Then define,

\[ \alpha_n(y, x) = n^{\frac{1}{2}} \left( \hat{G}_n(y, x) - G(y, x) \right); \quad t_n(x) = n^{\frac{1}{2}} \left( \hat{F}_n(x) - F(x) \right) \]

Then define,

\[ g^{(1)}_{(n)}(y, x) = -h_n^{-(p+1)} \int_{-\infty}^{\infty} K^{(1)}_1 \left( \frac{u - x}{h_n} \right) d_u \hat{G}_n(y, u) = -\left( nh_n^{p+1} \right)^{-1} \sum_{i=1}^{n} K^{(1)}_1 \left( \frac{X_i - x}{h_n} \right) 1(Y_i \leq y) \]

\[ f^{(1)}_{(n)}(x) = \hat{g}^{(1)}_{(n)}(\infty, x) = -h_n^{-(p+1)} \int_{-\infty}^{\infty} K^{(1)}_1 \left( \frac{u - x}{h_n} \right) d_u \hat{F}_n(u) = -\left( nh_n^{p+1} \right)^{-1} \sum_{i=1}^{n} K^{(1)}_1 \left( \frac{X_i - x}{h_n} \right) \]

\[ m^{(1)}_{(n)}(y|x) = f^{(1)}_{(n)}(x)^{-1} \left[ g^{(1)}_{(n)}(y, x) - m_{(n)}(y|x)f^{(1)}_{(n)}(x) \right] \]

Again we first want to show that

**Lemma 4** Under Assumption 2.1, 2.5-2.6 and 2.7.1,

\[ (nh_n^{p+2})^{\frac{1}{2}} \left( \hat{m}^{(1)}_{(n)}(.|x) - m^{(1)}_{(n)}(.|x) \right) \Rightarrow \left[ \frac{\int K^{(1)}_1(\phi)^2 d\phi}{f(x)} \right]^{\frac{1}{2}} \mathcal{B}(m(.|x)) \]

The result of Theorem 2.2 follows Lemma 4 and the fact that \( \sup_y |m^{(1)}_{(n)}(y|x) - m^{(1)}(y|x)| = O \left( h_n^2 \right) \) under
assumption 2.5-2.7.

Proof of Lemma 4:

Define \( r_n(y|x) = (nh_n^{p+2})^{\frac{1}{2}} \left( \hat{m}_n^{(1)}(y|x) - m_{(n)}^{(1)}(y|x) \right) \). First we show that

(4)

\[
\{ r_n(y, x), y \in \mathcal{Y} \} \overset{D}{=} \left\{ \frac{\hat{h_n}^{-\frac{p}{2}}}{f(x)} \int -K_1^{(1)} \left( \frac{u - x}{h_n} \right) d_u \alpha_n(y, u) - \frac{\hat{h_n}^{-\frac{p}{2}} m(y|x)}{f(x)} \int -K_1^{(1)} \left( \frac{u - x}{h_n} \right) dt_u(y), y \in \mathcal{Y} \right\}
\]

Decompose \( r_n(y|x) \), we get

\[
\begin{align*}
\hat{r}_n(y|x) &= (nh_n^{p+2})^{\frac{1}{2}} \left[ \frac{\hat{g}^{(1)}_n(y, x)}{f_n(x)} - \frac{g^{(1)}_n(y, x)}{f_n(x)} \right] - (nh_n^{p+2})^{\frac{1}{2}} m_{(n)}(y|x) \left( \frac{\hat{f}^{(1)}_n(x)}{f_n(x)} - \frac{f^{(1)}_n(x)}{f_n(x)} \right) \\
& \quad - (nh_n^{p+2})^{\frac{1}{2}} f_n(x) \left( \frac{\hat{m}_n(y|x) - m_{(n)}(y|x)}{f_n(x)} \right)
\end{align*}
\]

(5)

By Lemma 3 and the fact that Assumption 2.5-2.7.1 are stronger than Assumption 2.2-2.4.1, we have

\[
\sup_y \left| (nh_n^p)^{\frac{1}{2}} \left( \hat{m}_n(y|x) - m_{(n)}(y|x) \right) \right| = O_p(1)
\]

Together with the smoothness assumption in 2.5.3, we know the last term in the RHS of equation (5) satisfies:

(6)

\[
\sup_y \left| (nh_n^{p+2})^{\frac{1}{2}} f_n(x) \left( \frac{\hat{m}_n(y|x) - m_{(n)}(y|x)}{f_n(x)} \right) \right| \overset{p}{\to} 0
\]
Now we consider the first term in the RHS of equation (5).

\[
(nh_n^{p+2})^{\frac{1}{2}} \left[ \frac{g_n^{(1)}(y, x)}{f_n(x)} - \frac{g_n^{(1)}(y, x)}{f_n(x)} \right] = \left( nh_n^{p+2} \right)^{\frac{1}{2}} \frac{g_n^{(1)}(y, x) - g_n^{(1)}(y, x)}{f_n(x)}
\]

\[
- \left( nh_n^{p+2} \right)^{\frac{1}{2}} \frac{g_n^{(1)}(y, x)}{f_n(x)} \left( \hat{f}_n(x) - f_n(x) \right)
\]

\[
- \left( nh_n^{p+2} \right)^{\frac{1}{2}} \frac{g_n^{(1)}(y, x) - g_n^{(1)}(y, x)}{f_n(x)} \left( \hat{f}_n(x) - f_n(x) \right)
\]

\[
+ \left( nh_n^{p+2} \right)^{\frac{1}{2}} \frac{g_n^{(1)}(y, x)}{f_n(x)} \left( \hat{f}_n(x) - f_n(x) \right)^2
\]

We know that \( (nh_n^{p+2})^{\frac{1}{2}} \left( g_n^{(1)}(y, x) - g_n^{(1)}(y, x) \right) = nh_n^{\frac{p}{2}} \int -K_1^{(1)} \left( \frac{u - x}{h_n} \right) d_u \alpha_n(y, u) \). Meanwhile, the other three parts of the expansion could be shown uniformly converging to zero since we have, under Assumptions 2.5-2.7.1, \( \hat{f}_n(x) - f_n(x) \xrightarrow{p} 0 \), \( f_n(x) \to f(x) \), \( (nh_n^{p+2})^{\frac{1}{2}} \left( \hat{f}_n(x) - f_n(x) \right) \xrightarrow{p} 0 \), \( \sup_y \left| g_n^{(1)}(y, x) - g_n^{(1)}(y, x) \right| \xrightarrow{p} 0 \) (could be shown in the same way as equation (2) in the previous proof.) and \( \sup_y \left| g_n^{(1)}(y, x) \right| < \infty \).

Thus, we have that when \( n \to \infty \),

\[
\left\{ (nh_n^{p+2})^{\frac{1}{2}} \left[ \frac{g_n^{(1)}(y, x)}{f_n(x)} - \frac{g_n^{(1)}(y, x)}{f_n(x)} \right], y \in \mathcal{Y} \right\} \xrightarrow{D} \left\{ \frac{h_n^{-\frac{p}{2}}}{f(x)} \int -K_1^{(1)} \left( \frac{u - x}{h_n} \right) d_u \alpha_n(y, u), y \in \mathcal{Y} \right\}.
\]

Likewise, the second term in the RHS of equation (5) satisfy that as \( n \to \infty \),

\[
\left\{ (nh_n^{p+2})^{\frac{1}{2}} \left( \frac{\hat{f}_n^{(1)}(y, x)}{f_n(x)} - \frac{f_n^{(1)}(y, x)}{f_n(x)} \right) m_n(y|x), y \in \mathcal{Y} \right\} \xrightarrow{D} \left\{ \frac{h_n^{-\frac{p}{2}}}{f(x)} \int -K_1^{(1)} \left( \frac{u - x}{h_n} \right) dt_n(u), y \in \mathcal{Y} \right\}.
\]

(5) (7), (8), (6) together give the result of (4). Using the same trick as in the proof of Lemma 3 after showing (1), we then get the result of Lemma 4 and hence Theorem 2.2.
Proof of Proposition 2.1

First we show Proposition 2.1.1, the asymptotic property of the test statistic when the null hypothesis is true. Define

$$
\hat{T}_1(\cdot|x_2) = A_1 \left( \frac{n_1 h_{1n}^q \hat{f}_{1n}(x_2) n_0 h_{0n}^q \hat{f}_{0n}(x_2)}{n_1 h_{1n}^q \hat{f}_{1n}(x_2) + n_0 h_{0n}^q \hat{f}_{0n}(x_2)} \right)^{\frac{1}{2}} \left[ (\hat{m}_{1n}(\cdot|x_2) - m_1(\cdot|x_2)) - (\hat{m}_{0n}(\cdot|x_2) - m_0(\cdot|x_2)) \right].
$$

The test statistic satisfies

$$
\hat{S}_1(x_2) \leq \sup_{y \in \mathcal{Y}} \hat{T}_1(y|x_2) + \sup_{y \in \mathcal{Y}} (m_1(y|x_2) - m_0(y|x_2))
\leq \sup_{y \in \mathcal{Y}} \hat{T}_1(y|x_2).
$$

with equality holds when \( m_1(\cdot|x_2) = m_0(\cdot|x_2) \). Let bandwidths be \( h_{0n} = \delta_0 n_0^{-\lambda} \) and \( h_{1n} = \delta_1 n_1^{-\lambda} \). If \( \Delta_0 > \Delta_1 \), \( n_0 h_{0n}^q \) goes to infinity faster than \( n_1 h_{1n}^q \). Together with Corollary 2.1,

$$
\hat{T}_1(\cdot|x_2) = \frac{n_0 h_{0n}^q \hat{f}_{0n}(x_2)}{n_1 h_{1n}^q \hat{f}_{1n}(x_2) + n_0 h_{0n}^q \hat{f}_{0n}(x_2)} \left\{ A_1 \left[ n_1 h_{1n}^q \hat{f}_{1n}(x_2) \right]^{\frac{1}{2}} \left( \hat{m}_{1n}(\cdot|x_2) - m_1(\cdot|x_2) \right) 
- A_1 \left[ n_0 h_{0n}^q \hat{f}_{0n}(x_2) \right]^{\frac{1}{2}} \left( \hat{m}_{0n}(\cdot|x_2) - m_0(\cdot|x_2) \right) \right\} 
\rightarrow A_1 \left[ n_1 h_{1n}^q \hat{f}_{1n}(x_2) \right]^{\frac{1}{2}} \left( \hat{m}_{1n}(\cdot|x_2) - m_1(\cdot|x_2) \right) 
\Rightarrow \mathfrak{B}(m_1(\cdot|x_2)).
$$

Therefore by Continuous Mapping Theorem,

$$
P(\hat{S}_1(x_2 > c_1)) \leq P \left( \sup_{y \in \mathcal{Y}} \mathfrak{B}(m_1(y|x_2)) > c_1 \right) = P \left( \sup_{t} \mathfrak{B}(t) > c_1 \right).
$$

Likewise, if \( \Delta_0 < \Delta_1 \), \( \hat{T}_1(x_2) \) converges weakly to \( \mathfrak{B}(m_0(\cdot|x_2)) \) and we also get Proposition 2.1.1.

Let \( n_0 = \sum_{i=1}^n 1(X_i = 0), n_1 = \sum_{i=1}^n 1(X_i = 1) \) be the number of observations with \( X_1 = 0 \) or \( 1 \), and \( \lambda_n = \lim_{n \to \infty} \lambda_n = P(X_1 = 1) \equiv \lambda \in (0,1) \). If \( \Delta_1 = \Delta_2, \frac{n_1 h_{1n}^q \hat{f}_{1n}(x_2)}{n_0 h_{0n}^q \hat{f}_{0n}(x_2)} P \left( \frac{\hat{x}}{y_2} \right)^q \left( \frac{\lambda}{1-\lambda} \right)^{1-\frac{q}{2}} \frac{f_1(x_2)}{f_0(x_2)} \equiv \xi. \)
Then \( \hat{T}_1(x_2) \) converges weakly to a linear combination of two changed time random Brownian Bridge processes:

\[
\hat{T}_1(.|x_2) \Rightarrow \sqrt{\frac{1}{\xi + 1}} B(m_1(.|x_2)) - \sqrt{\frac{\xi}{\xi + 1}} B(m_0(.|x_2)) \equiv \bar{T}_1(.|x_2).
\]

Denote the limiting random variable corresponding to \( \bar{T}_1(.|x_2) \) in the case where \( m_1(.|x_2) = m_0(.|x_2) \) by \( \bar{T}_1^0(.|x_2) \). Notice the fact that \( \bar{T}_1^0(.|x_2) \overset{D}{=} B(m_0(.|x_2)). \) Let \( Y^* \) denote the set of \( y \) values for which \( m_1(y|x_2) = m_0(y|x_2) \). Then all we need to show for Proposition 2.1.1 when \( \Delta_0 = \Delta_1 \) is that

\[
P(\hat{S}_1 > c_1) \rightarrow P\left( \sup_{y \in Y^*} \bar{T}_1(y|x_2) > c_1 \right)
\]

and that

\[
P\left( \sup_{y \in Y^*} \bar{T}_1(y|x_2) > c_1 \right) \leq P\left( \sup_{y \in Y^*} \bar{T}_1^0(y|x_2) > c_1 \right) \leq P\left( \sup_{y \in Y} \bar{T}_1^0(y|x_2) > c_1 \right).
\]

Proof for equation (7) and (8) are similar to those provided for equation (18) and (26) in Barrett and Donald (2003) [3] and is omitted here. The tightness condition of the weak convergence shown in Corollary 2.1 is the workhorse in proving equation (7).

To prove the second part of Proposition 2.1, we note that if the alternative hypothesis is true, there exists some \( y^+ \in Y \) such that

\[
m_1(y^+|x) - m_0(y^+|x) = \delta > 0.
\]

Then we know

\[
\hat{S}_1(x_2) \geq \bar{T}_1(y^+|x_2) + (m_1(y^+|x_2) - m_0(y^+|x_2))
\]

\[
\rightarrow \infty
\]

The first inequality holds because \( y^+ \in Y \); the second holds from the bandwidth assumption 2.4.1 and equation (9). The result of the proposition hence follows.
Proof of Proposition 2.2

To prove part 1 of the proposition, we note that
\[ \hat{S}_2 \leq \sup_y \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \left( \hat{m}_n^{(1)}(y|x) - m^{(1)}(y|x) \right) + \sup_y \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} m^{(1)}(y|x) \]
\[ \leq \sup_y \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \left( \hat{m}_n^{(1)}(y|x) - m^{(1)}(y|x) \right) \]

The first inequality is due to the property of supremum and the second uses the fact that under \( H_{0,x}^2 \), \( m^{(1)}(y|x) \leq 0 \) for all \( y \in \mathcal{Y} \). Both inequalities turn to equality if and only if \( m^{(1)}(y|x) = 0 \) for all \( y \in \mathcal{Y} \).

Therefore, the result follows from Corollary 2.2.

To prove the second part, we note that if the alternative hypothesis is true, there exists some \( y^* \in \mathcal{Y} \) such that \( m^{(1)}(y^+|x) > 0 \). Then we know
\[ \hat{S}_2 \geq \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \hat{m}_n^{(1)}(y^+|x) \]
\[ \geq \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} \left( \hat{m}_n^{(1)}(y^+|x) - m^{(1)}(y^+|x) \right) + \left[ n h_n^{p+2} \hat{f}_n(x) \right]^{\frac{1}{2}} m^{(1)}(y^+|x) \]
\[ \to \infty \quad \text{as} \quad n \to \infty \]

The first inequality holds because \( y^* \in \mathcal{Y} \); the second holds from the bandwidth assumption in 2.7.1. The result of the proposition hence follows.

Proof of Corollary 3.1

We the corollary for the discrete variable case by show that

Lemma 5 Under Assumption 2.1-2.4, we have that
\[ \left[ n_0 h_{0n} \right]^{\frac{1}{2}} \left( \hat{m}_{0n}(.|x_2^2) - m_0(.|x_2^2) \right) \Rightarrow A_{1}^{-1} \tilde{f}_0(x_2\theta_2)^{-\frac{1}{2}} \mathbb{V} \left( m_0(.|x_2\theta_2) \right) \]
in \( D([0,\tilde{z}]) \) with \( P(\mathbb{V} (m_0(.|x_2)) \in C([0,\tilde{z}])) = 1 \).

A corresponding result holds for the process concerning the \( X_1 = 1 \) observations.
Do a second order Taylor’s expansion of \( \hat{m}_n(y|z_2\hat{\theta}_2) \) around \( \theta \),

\[
\hat{m}_n(y|x_2\hat{\theta}_2) = \hat{m}_0n(y|x_2\theta) + \sum_{i=1}^{q}(\hat{\theta}_2 - \theta_i^2)\hat{m}_0^{(1)}(y|x_2\theta_2)\theta_i^2 + R(\theta_2)
\]

where the remainder term \( R(\theta_2) \) is less than \( A_3 \sum_{i=1}^{q}(\hat{\theta}_2 - \theta_i^2)^2 \) with \( 2A_3 \) being the constant term that bound the second derivative of \( \hat{m}_0n(y|x_2\hat{\theta}) \). Since we know that \( \hat{\theta} \xrightarrow{p} \theta \), \( h_n \to 0 \) and that \( \sqrt{n}(\hat{\theta} - \theta) \) converges to a normal distribution, we can conclude the following uniform convergence result

\[
\sup_y \left| \frac{1}{n}\sum_{i=1}^{n}1(i = n \times t)U_i(\omega) \right| \xrightarrow{p} 0.
\]

Together with Theorem 2.1, we have the result of Lemma 5. The corollary for the continuous variable case could be shown in the same fashion.

**Proof of Proposition 3.1**

We only proof the proposition for the continuous variable case. The discrete variable case easily follows. Given normal random variables \( U_1, U_2,... \) on an \((\Omega, \mathcal{F}, P)\), let \( M^n(w) \) be the function in \( D([0,1]) \) with value

\[
M^n_t(\omega) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}1(i \leq n \times t)U_i(\omega)
\]

By Billingsley (1999) [5] Theorem 14.1, we know that

\[
M^n \Rightarrow W.
\]

where \( W \) is standard Wiener process on \( D([0,1]) \) and \( P(W \in C) = 1 \). Define,

\[
B^*_t(\omega) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[1(i \leq n \times t) - t]U_i(\omega).
\]

Then \( B^* \) converges to a standard Brownian Bridge process \( B \) on \( D([0,1]) \) and \( P(B \in C) = 1 \).
Also, given the assumptions for kernels and bandwidths, we have

$$\hat{m}_n(.|x) \Rightarrow m(.|x).$$

Therefore, by the Lemma on page 151 of Billingsley (1999) [5], we know that

$$B^*(\hat{m}_n(.|x)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [1(i \leq n \times \hat{m}_n(.|x)) - \hat{m}_n(.|x)]U_i \Rightarrow B(m(.|x)).$$

The rest of the proof for Proposition 3.1 is the same as in the proof for Proposition 2 in Barrett and Donald (2003) [3] and is hence omitted.
References


