Tests for detecting probability mass points*

Byung-hill Jun† Hosin Song‡

March 31, 2017

Abstract

The objective of this paper is developing a test to detect the presence of mass points. We consider a mixture model to represent a random variable which has probability mass points as well as non-mass points. To serve our purpose we propose two tests: the Kolmogorov-Smirnov (KS) test and the probability mass point (PMP) test. The KS test uses the well-known fact that the distribution function of a continuous random variable is uniformly distributed on a unit interval. The PMP test exploits the asymptotic difference between two kernel density estimators with different bandwidths. Depending on the choice of bandwidths, two versions of PMP tests are presented. Based on the PMP test and the idea of complete consistency in Andrews (1986), we propose a testing procedure to consistently construct a set of mass points. Numerical experiments and an application are conducted to demonstrate the validity of our proposed methods.

*An earlier version of this paper entitled "Consistent kernel density estimation in the presence of probability mass points" was presented at 2014 Asian Meeting of the Econometric Society, WEAI 89th annual conference, and 2016 China Meeting of the Econometric Society. We thank anonymous referees for helpful comments. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A5A8018660).

†Department of Int'l Economics and Law, Hankuk University of Foreign Studies, Seoul, Korea. E-mail: bjun@hufs.ac.kr.

‡Department of Economics, Ewha Womans University, Seoul, Korea. Corresponding author. E-mail: hsong@ewha.ac.kr.

Keywords: complete consistency, kernel density estimator, Kolmogorov-Smirnov test, probability mass point test, multiple hypothesis testing

JEL codes : C14, C18, C46, C52

1 Introduction

Empirical researchers are often confronted with variables massed at certain values. Examples of such variables include subjective probability (Bruin de Bruin et al. (2000,2002)), firm's earnings (Burgstahler and Dichev (1997)), income (Saez (2010), and Chetty et al. (2011)), job tenure (Ureta (1992)), household expenditure (Pudney (2008)), working hours (Otterbach and Sousa-Poza (2010)), and neonatal mortality (Arulampalam et al. (2014)). There are several arguments trying to explain these mass points (e.g., personal characteristics associated with rounding or recollection errors (Budd and Guinnane (1991), and Bruin de Bruin et al. (2000, 2002)), nature of survey question (Holbrook et al. (2014)), and optimization behavior associated with an incentive (Burgstahler and Dichev (1997), Saez (2010), and Chetty et al. (2011)). Regardless of the disagreement about the sources generating the mass points, it is well known that neglecting the presence of mass points leads to the failure of conventional statistical inference in many situations (Heitjan and Rubin (1991)). Many studies address the issues arising from the prevalence of mass points to develop ways to obtain correct estimates and inferences (Petoussi et al. (1997), Bruin de Bruin et al. (2002), Bar and Lillard (2012), Arulampalam et al. (2014), Crawford et al. (2015), Barreca et al. (2016), and Zinn and Würbachac (2016)).

Our work shares the concerns related to the mass points with these lines of studies. However, instead of suggesting a correction for problems resulting from the mass points in a specific model, we aim to provide a way to check whether or not data used in analysis are massed at some values in a general setting. Similar attempts have been made in previous works. For example, Burgstahler and Dichev (1997) and Saez (2010) examine some points around which individuals or firms are known to experience an abrupt change affecting their economic incentives. These studies examine whether excess observations are massed at those conjectured points. Takeuchi (2004) is more closely related to our work. Motivated by a histogram approach in Burgstahler and Dichev (1997), he proposes a simple statistical test to see if there is a jump in the distribution function by using a property of the smoothness of a distribution function.

Suggestions by the aforementioned works are often of limited use in practice since they presume the knowledge of possible location of mass points. Another limitation is that these studies consider a specific single point of a distribution while there are possibly multiple candidates of mass points all of which should be correctly classified as a mass point or a non-mass point. The objective of this paper is to develop a procedure which consistently detects the set of mass points. To this purpose two tests are proposed to identify the presence of such mass points. The first proposed test is the KS test. This test exploits the fact that a cumulative distribution function (CDF) of a continuous random variable is a uniform random variable on the unit interval $[0, 1]$. The null hypothesis of non-existence of mass points is rejected when the KS statistic is sufficiently large. The KS test can be used to globally detect the presence of mass points since it does not test the probability mass at a specific single point. Secondly, we propose the probability mass point (PMP) test. This test employs the asymptotic difference between two kernel density estimators with different bandwidths. We show that depending on bandwidths choice, the proposed PMP test statistic converges to the standard normal distribution or a linear transformation of a positive Poisson distribution at a non-mass point, while it diverges to infinity at a mass point. Compared to Takeuchi (2004), the PMP test is more general since the PMP test incorporates a general kernel density function estimator, while Takeuchi (2004) considers histograms with a fixed bin. Unlike the KS test, the PMP test locally tests the existence of probability mass at a specific point. Thus, this test can be used to find the location of mass points as a complement of the KS test.

These two tests can provide meaningful information on the existence of mass points. However, proposing them is not sufficient to serve our purpose since a consistent procedure should satisfy the following two conditions. Firstly, there is no false detection at least in the limit. In other words, it requires that both the probability that a mass point is not detected and the probability that a non-mass point is falsely detected converge to 0 as a sample size increases to infinity. To deal with the first requirement, we adopt a completely consistent test suggested by Andrews (1986). Secondly, any value on the empirical support might have probability mass. This feature requires a multiple hypothesis testing. Therefore, we modify the PMP test to incorporate the idea of a completely consistent test in Andrews (1986) and a multiple testing procedure proposed by Holm (1979).

The paper is organized as follows. In section 2, we consider a mixture model to represent the situation where mass points are present among non-mass points. In section 3 and 4, we present the KS test and the PMP test. Section 5 develops a procedure to consistently detect all mass points. In section 6, we conduct some numerical experiments to evaluate the performance of our proposed tests and procedure. In section 7, we apply our proposed methods to the PSID wage data. Section 8 concludes. All technical details including proofs are presented in appendix.

2 Model

Let $Y_1, \ldots, Y_n$ be random variables which are independently drawn from an identical distribution. We consider a case where the distribution of these random variables possibly has mass points on its support. The location of mass points is assumed to be unknown to researchers a priori. Specifically, we consider a mixture model stated in Assumption 1.

Assumption 1. $Y_1, \ldots, Y_n$ *are independently and identically distributed (i.i.d.) with the following distribution function* $F_0(y)$*:*

$$F_0(y) = pF_1(y) + (1-p)F_2(y) \tag{1}$$

where $0 \leq p \leq 1$*,* $F_1(y) = \sum_{j=1}^{J} f_1(d_j) I(y \geq d_j)$*,* $f_1(y)$ *is a probability mass function (pmf) whose support is* $D = \{d_1, \ldots, d_J\}$*, and* $F_2(y) = \int_{-\infty}^{y} f_2(t)dt$*,* $f_2(y)$ *is a probability density* *function (pdf) with support $C \subset \mathbb{R}$ such that $f_2(y)$ is twice continuously differentiable at any point $y \in C$. Also assume that $D \subset \overline{C}$ where $\overline{C}$ is the closure of C.*

Restrictions for a continuous pdf $f_2(y)$ are standard in many econometric studies. These restrictions are used to derive the properties of the PMP test in section 4. The mass points are assumed to be present on the closure of non-mass points. This fact complicates the identification of the set of mass points D.

As shown in the following examples, variables examined in related literature are described by the model (1).

Example 1. Bruin de Bruin et al. (2000, 2002) find that an excessive fraction of individuals choose 50% when they report their subjective probabilities. In this example, Y is a reported subjective probability, $D = \{0.5\}$ and $C = \{y \in \mathbb{R} : 0 \leq y \leq 1\}$.

Example 2. Saez (2010) notices the bunching of self-employed workers' earnings at the kinked points of income taxes. Denote the kinked points by $d_1, d_2, \ldots, d_J$. In this example, Y is the earnings, $D = \{d_1, d_2, \ldots, d_J\}$ and $C = \{y \in \mathbb{R} : y \geq 0\}$.

Example 3. Zinn and Würbach (2016) considers the digit preference patterns when individual reports their incomes. Specifically, individuals are assumed to tend to report multiples of 100, 500, and 1,000 instead of their true incomes. Y is the reported income, $D = \{d = kd_0 : d_0 = 100, 500, 1,000\ ; k = 0, 1, \ldots\}$ and $C = \{y \in \mathbb{R} : y \geq 0\}$.

Our primary interest is identifying the set of mass points D. Depending on a data generating process, a research might have an additional interest to recover the distribution of a true variable. As mentioned in Barreca, Lindo, and Waddell (2016) if the mass points are regarded as the results of measurement errors (e.g., the digit preference, recollection errors, and rounding errors), one might want to obtain the distribution of a true variable

Y^*.

3 KS test

In this section, we present a way to apply the KS statistic to detect the presence of mass points. We exploit the well-known fact that the distribution function of a continuous random variable is known to follow a uniform distribution on the unit interval. Therefore, the distribution function $F_0(y)$ in (1) is a uniform random variable on $[0, 1]$ when there are no mass points ($p = 0$). Otherwise ($p > 0$), the distribution function $F_0(y)$ has jumps and does not follow a uniform distribution.

In this sense, we use the KS statistic to globally detect the presence of mass points. Specifically, we test $H_0 : p = 0$ versus $H_1 : p > 0$. Then, the KS test statistic is defined as follows.

$$
\begin{aligned}
KS &= \sup_y \sqrt{n} \left| \hat{F}(y) - F_2(y) \right| \\
&= \sup_y \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} I(Y_i \leq y) - F_2(y) \right| \\
&= \sup_y \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} I(F_2(Y_i) \leq F_2(y)) - F_2(y) \right| \\
&= \sup_{u \in (0,1)} \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} I(F_2(Y_i) \leq u) - u \right| \qquad (2)
\end{aligned}
$$

where $\hat{F}(y) = \frac{1}{n} \sum_{i=1}^{n} I(Y_i \leq y)$ and $I(\cdot)$ is the indicator function. The third equality follows from the monotone property of a distribution function, and the fourth equality follows from the fact that any continuous distribution function follows a uniform distribution on the unit interval.

Under the null hypothesis $H_0 : p = 0$, $KS \xrightarrow{d} \sup_{u \in [0,1]} B(u)$, $0 \leq u \leq 1$ where $B(\cdot)$ is a Brownian bridge process. The KS test statistic is infeasible since F_2 is unknown in general. Therefore, replacing F_2 with its empirical distribution $\hat{F}$ constructs the following feasible test statistic.

$$KS_n = \sup_{u \in (0,1)} \sqrt{n} \left| \frac{1}{n} \sum_{i=1}^{n} I(\hat{F}(Y_i) \leq u) - u \right| \tag{3}$$

where $\hat{F}(Y_i) = \dfrac{1}{n-1} \displaystyle\sum_{j=1, j \neq i}^{n} I(Y_j \leq Y_i)$.

KS_n converges to the supremum of the Brownian bridge process in distribution since $\hat{F}$ strongly converges to F_0 as n increases. Critical values can be obtained according to the following theorem in Serfling (1980).

Theorem (Kolmogorov) *Let F_0 be one-dimensional and continuous. Then for any $d > 0$,*

$$\lim_{n \to \infty} P[KS_n \leq d] = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j+1} \exp(-2j^2 d^2).$$

Instead of $1 - 2\sum_{j=1}^{\infty}(-1)^{j+1}\exp(-2j^2d^2)$, we can use $1 - 2\sum_{j=1}^{M}(-1)^{j+1}\exp(-2j^2d^2)$ with sufficiently large M. When $M = 1000$, $P[KS \leq 1.23] = 0.90, P[KS \leq 1.36] = 0.95$, and $P[KS \leq 1.63] = 0.99$.

4 PMP test

4.1 Properties of a kernel density estimator

In this section, we present the PMP test which employs two kernel density estimators using different bandwidths. Specifically, the PMP test uses the asymptotic properties of $\widehat{f}(y_0; h_1) - \widehat{f}(y_0; h_2)$, $h_1 \neq h_2$ where $\widehat{f}(y_0; h_j)$, $j = 1, 2$ is a kernel density estimator of $f(y_0)$ with a bandwidth h_j. To define the PMP test formally, let us state the assumptions for a kernel $K(\cdot)$ and a bandwidth h.

Assumption 2A. *The kernel $K(\cdot)$ satisfies the following conditions:*

(i) $K : \mathbb{R} \to \mathbb{R}$, $K(\cdot)$ is Lebesgue integrable,

(ii) $\int K(u)du = 1$,

(iii) $K(\cdot)$ is uniformly bounded,

(iv) $\lim_{|u|\to\infty} |u|K(u) = 0$, and

(v) $K(\cdot)$ is symmetric.

Assumption 3A. *A bandwidth h satisfies $h \downarrow 0$ and $nh \to \infty$ as $n \to \infty$.*

Conditions 2A and 3A are standard in kernel estimation. Notice that the fourth condition in Assumption 2A implies $\lim_{|u|\to\infty} K(u) = 0$.

When p in (1) is 0, Assumptions 1, 2A, and 3A are reduced to the standard ones in the kernel density estimation. A kernel density estimator $\widehat{f}(y_0; h)$ defined as $\widehat{f}(y_0; h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y_0 - Y_i}{h}\right)$ is consistent in such a setup. However, the consistency of a kernel density estimator fails when the data generating process is given by (1) with $p > 0$ (Zinde-Walsh (2008)). This finding is related to the following results on the mean and variance of a kernel density estimator.

$$E\left[\widehat{f}(y_0; h)\right] = \frac{1}{h}K(0)p(y_0) + (1-p)\left(f_2(y_0) + \kappa_\mu h^2\right) + o\left(p(y_0) + h^2\right) \tag{4}$$

and

$$V\left[\widehat{f}(y_0; h)\right] = \frac{1}{nh^2}K^2(0)p(y_0)(1 - p(y_0)) + \frac{1}{nh}(1-p)\kappa_V + o\left(p(y_0)\frac{1}{nh^2} + \frac{1}{nh}\right) \tag{5}$$

where $p(y_0) = pf_1(y_0)$, $\kappa_\mu = \frac{1}{2}f_2^{(2)}(y_0)\int s^2K(s)ds$, $\kappa_V = f_2(y_0)\int K^2(s)ds$, and $f_2^{(2)}(\cdot)$ is the second derivative of $f_2(\cdot)$. Recognize that when $p = 0$, equations (4) and (5) are reduced to the well-known expressions of the asymptotic mean and variance of a kernel density estimator in literature. However, when $p > 0$, a kernel estimator at a mass point y_0 ($p(y_0) > 0$) has exploding terms in the asymptotic bias and variance. As a result,

$$\widehat{f}(y_0; h) \quad \xrightarrow{p} \quad \infty \quad \text{for} \quad y_0 \in D \tag{6}$$

and

$$\widehat{f}(y_0;h) \quad \xrightarrow{p} \quad (1-p)f_2(y_0) \quad \text{for} \quad y_0 \in C. \tag{7}$$

4.2 PMP test

Burgstahler and Dichev(1997) and Takeuchi(2004) suggest using the smoothness of a density function to test whether a distribution function is continuous at a specific point. Their suggestions are implicitly based on the facts that if there is no probability mass at y_0, a kernel density estimator with a suitable bandwidth is consistent and the difference between two kernel density estimators using different suitable bandwidths is well approximated by a normal distribution.

We attempt to generalize this approach. Unlike the KS test, our goal is testing whether a specific point y_0 is a mass point or not. Remember that $p(y_0)$ is the probability that a random variable Y takes on a value y. In other words, $p(y_0) \equiv pf_1(y_0)$. Hence, $p(y_0) = 0$ is equivalent to $p = 0$ or $f_1(y_0) = 0$. The null and alternative hypotheses are expressed as $H_0 : p(y_0) = 0$ vs. $H_1 : p(y_0) \neq 0$, respectively. Let $\widehat{f}(y_0; h_1)$ and $\widehat{f}(y_0; h_2)$ be estimators for $f(y_0)$ with different bandwidths h_1 and h_2. Hereafter, suppress y_0 so that $\widehat{f(h_1)}$ and $\widehat{f(h_2)}$ denote these two estimators for $f(y_0)$.

Let us introduce a constant c characterizing the relative magnitude of h_1 and h_2. Without loss of generality, assume $h_1 < h_2$. Define c as the limit of two bandwidths, that is, $c = \lim h_1/h_2 \in [0, 1)$. In the following part, we present two versions of the PMP test depending on c.

4.2.1 PMP test with $c > 0$

Assumption 2B. *In addition to Assumption 2A, a kernel $K(\cdot)$ satisfies the following two conditions:*
(i) $K(\cdot)$ is second order, that is, $\int uK(u)du = 0$ and $\int u^2K(u)du < \infty$, and

(ii) $K(\cdot)$ is Lipschitz-continuous. In other words, there exists a finite constant d^* satisfying

$$|K(x) - K(y)| \leq d^*|x - y| \quad \textit{for all } x, y.$$

Assumption 3B *Suppose that h_1 and h_2 satisfy Assumption 3A with $c \in (0,1)$ as $n \to \infty$ and $nh_1h_2^4 \to 0$ as $n \longrightarrow \infty$.*

The Lipschitz-continuous condition is required to evaluate the variance of $\widehat{f(h_1)} - \widehat{f(h_2)}$. As mentioned by Newey and West (1994), a wide class of kernels satisfy the Lipschitz-continuity so that Assumption 2B is not so restrictive.Assumption 3B requires $nh_1h_2^4 \to 0$. Let $h_2 = n^\alpha$ and $h_1 = ch_2$ with $\alpha < 0$. Then conditions $nh_1h_2^4 \to 0$ and $nh_2 \to \infty$ imply a restriction on α. In comparison with the optimal bandwidth minimizing the mean-squared errors, Assumption 3B is interpreted as a undersmoothing condition.

Theorem 1 states that the standardized difference between two kernel density estimators is well approximated by a standard normal distribution under $H_0 : p(y_0) = 0$, and explodes under $H_1 : p(y_0) > 0$.

Theorem 1. *Suppose Assumptions 1, 2B and 3B hold. Then, for any y_0,*

$$T_L = \frac{\sqrt{nh_1}(\widehat{f(h_1)} - \widehat{f(h_2)})}{\sqrt{\hat{V}_0}} \xrightarrow{d} N(0,1) \textit{ under } H_0 : p(y_0) = 0,$$

where $\hat{V}_0$ is given by $\hat{V}_0 = k_c\widehat{f(h_2)}$ for a constant

$$k_c = (1 + c)\int K^2(s)ds - 2c\int K(s)K(cs)ds.$$

But, under $H_1 : p(y_0) > 0$, T_L tends to infinity with probability 1.

Theorem 1 provides a way to construct a consistent test for $H_0 : p(y_0) = 0$ against $H_1 : p(y_0) > 0$. Under H_0, the order of the asymptotic bias and variance are h_2^2 and $1/(nh_1)$, respectively. Thus, the condition $nh_1h_2^4 \to 0$ is related to undersmoothing in constructing kernel density estimators $\widehat{f(h_1)}$ and $\widehat{f(h_2)}$, which guarantees that neglecting the asymptotic bias in T_L has no effect on the limiting distribution as long as the first-order is concerned. With such undersmoothing, the bias term will be asymptotically negligible. Then, Theorem 2 can be modified as

$$\frac{\sqrt{nh_1}\left(\widehat{f(h_1)}-\widehat{f(h_2)}-E\left[\widehat{f(h_1)}-\widehat{f(h_2)}\right]\right)}{\sqrt{\hat{V}_0}} \xrightarrow{d} N(0,1) \text{ under } H_0 : p(y_0)=0.$$

Burgstahler and Dichev (1997) and Takeuchi (2004) compare the difference between $\hat{p}_j$ and the simple average of $\hat{p}_{j-1}$ and $\hat{p}_{j+1}$ where $\hat{p}_j$ is the empirical frequency of the j-th bin whose center is y_0 and radius is $h/2$. More specifically, the test statistic of Burgstahler and Dichev (1997) and Takeuchi (2004) is given by

$$T^{\dagger} = \frac{(\hat{p}_{j-1}+\hat{p}_{j+1})/2-\hat{p}_j}{\sqrt{V\left((\hat{p}_{j-1}+\hat{p}_{j+1})/2-\hat{p}_j\right)}}.$$

A direct calculation provides

$$\frac{\hat{p}_{j-1}+\hat{p}_{j+1}}{2}-\hat{p}_j = \frac{3}{2}h\left(\widehat{f(3h/2)}-\widehat{f(h/2)}\right)$$

where $\widehat{f(h')}$, $h'=h/2$ or $3/2h$ is a uniform kernel density estimator of $f(y_0)$ with a bandwidth h'. It can be also shown that

$$V\left((\hat{p}_{j-1}+\hat{p}_{j+1})/2-\hat{p}_j\right) = \frac{3}{2n}hf(y_0)+o\left(\frac{h}{n}\right)$$

and

$$V\left(\frac{3}{2}h\left(\hat{f}(3h/2)-\hat{f}(h/2)\right)\right) = \frac{3}{2n}hf(y_0)+o\left(\frac{h}{n}\right).$$

Therefore, $T^{\dagger}$ can be interpreted as a statistic T_L using a uniform kernel and a set of bandwidths with $c=1/3$. This interpretation reflects the selection of the set of bandwidth. A close look at the proof of Theorem 1 implies that a test statistic $T^{\dagger}$ does not converge to a standard normal distribution without undersmoothing even under H_0. Thus, conventional critical values fail to generate the desired rejection probabilities under H_0.

Although c has no first order effect on the distribution of T_L under H_0, a close look at the proof of Theorem 1 hints at how the choice of c influences the size property of T_L in a finite sample. Note that the normal approximation in Theorem 1 becomes less precise as the difference between T_L and $N(0,1)$ becomes larger. Under H_0, the leading term in the difference between two random variables is proportional to $\frac{0.5}{\sqrt{k_c f(y_0)}}(1-1/c^2)f_2^{(2)}(y_0)\int s^2K(s)ds$. Therefore, size distortion is large when $\frac{1}{\sqrt{k_c}}(1-1/c^2)$ is large. Similarly, c affects the power property since the leading term of T_L under H_1 is proportional to $\frac{1}{\sqrt{k_c}}(1-c)$.[1] Thus, T_L has greater power when this leading term is large. Despite the clear dependence of size and power properties on c, the directional relationship between c and size distortion (or power) is not unanimous since the effects of c relies on a kernel $K(\cdot)$.

In implementing the PMP test, k_c needs to be computed. However, the computation of k_c may be cumbersome depending on the choice of kernel. In such a case, using bootstrap standard error is a convenient alternative. Let $\widehat{V}^*(\widehat{f_b^*(h_1)}-\widehat{f_b^*(h_2)})$ be a bootstrap variance estimator of $\widehat{f(h_1)}-\widehat{f(h_2)}$. Specifically,

$$\widehat{V}^*(\widehat{f_b^*(h_1)}-\widehat{f_b^*(h_2)}) = \frac{1}{B-1}\sum_{b=1}^{B}\left[\widehat{f_b^*(h_1)}-\widehat{f_b^*(h_2)}-\left(\overline{\widehat{f^*(h_1)}}-\overline{\widehat{f^*(h_2)}}\right)\right]^2 \tag{8}$$

where $(Y_{1,b}^*, Y_{2,b}^*, \ldots, Y_{n,b}^*)$ is the b-th bootstrap replication, B is the number of replications, and for $r=1,2$,

$$\widehat{f_b^*(h_r)} = \frac{1}{nh_r}\sum_{i=1}^{n}K\left(\frac{y_0-Y_{i,b}^*}{h_r}\right), \quad \overline{\widehat{f^*(h_r)}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{f_b^*(h_r)}.$$

The following test statistic $T_{L,B}$ can be used to implement the PMP test:

$$T_{L,B} = \frac{\widehat{f(h_1)}-\widehat{f(h_2)}}{\sqrt{\widehat{V}^*(\widehat{f_b^*(h_1)}-\widehat{f_b^*(h_2)})}}. \tag{9}$$

Proposition 1 states that $T_{L,B}$ can be used to detect the presence of the probability mass points.

Proposition 1. *Suppose Assumptions 1, 2B and 3B hold. Moreover, assume that* $B\to\infty$ *and* $n/B\to d$, $d\in[0,\infty)$ *as* $n\to\infty$. *Then, for any* y_0, $T_{L,B}\xrightarrow{d}N(0,1)$ *under* $H_0: p(y_0) =$

[1] More precisely, $\sqrt{\frac{h_1}{n}}T_L \xrightarrow{p} (1-c)K(0)\sqrt{\frac{p(y_0)}{k_c}}$ under H_1.

0 *and* $T_{L,B}$ *tends to infinity with probability 1 under* $H_1 : p(y_0) > 0$.

4.2.2 PMP test with $c = 0$

Assumption 3B requires that bandwidths h_1 and h_2 have the same order. The asymptotic behavior of the PMP test statistic given by $\sqrt{nh_1}(\widehat{f(h_1)} - \widehat{f(h_2)})/\sqrt{\hat{V}_0}$ is changed when the bandwidths have different orders. Now consider a special case where $h_1/h_2 \to c = 0$. For this special case we impose an additional restriction on the class of kernels. Specifically, we suggest using a uniform kernel in constructing $\widehat{f(h_1)}$ and $\widehat{f(h_2)}$.

Assumption 2C. *The kernel* $K(\cdot)$ *is defined as* $K(u) = I(|u| \leq 1/2)$ *where* $I(\cdot)$ *is the indicator function.*

Assumption 3C. *Assume that* $h_1 = c_0/n$, $c_0 > 0$, *and* h_2 *satisfies Assumption 3A.*

A uniform kernel in Assumption 2C satisfies Assumption 2B. Under Assumption 3C, k_c is 1. Then, $\widehat{V}_0$ is $\widehat{f(h_2)}$. Under Assumption 3C, only $\widehat{f(h_2)}$ is a consistent estimator. When $p(y_0) = 0$, on the empirical support an inconsistent estimator of $\widehat{f(h_1)}$ is shown to have the following limiting distribution:

$$\widehat{f(h_1)} \xrightarrow{d} \frac{1}{c_0} W^+(k^*) \text{ for } k^* = c_0 f_2(y_0) \tag{10}$$

where $W^+(k^*)$ is a positive Poisson distribution with a parameter k^*[2]. Theorem 3 combines these results to present the asymptotic distributions of a test statistic T under an alternative set of assumptions.

Theorem 2. *Suppose Assumptions 1, 2C, and 3C hold. Let* y_0 *be a point in its empirical*

[2]That is, $W^+(k^*)$ represents the conditional Poisson distribution $W(k^*)$ with mean k^* given that $W(k^*)$ is positive.

support. Then under $H_0 : p(y_0) = 0$*,*

$$T'_S = \frac{\sqrt{nh_1}\left(\widehat{f(h_1)} - \widehat{f(h_2)}\right)}{\sqrt{\widehat{V}_0}} \xrightarrow{d} \frac{W^+(k^*) - k^*}{\sqrt{k^*}}$$

as $n \to \infty$*. Under* $H_1 : p(y_0) > 0$*,* T'_S *diverges to* ∞ *with probability 1 as* $n \to \infty$*.*

Unlike the test statistic in Theorem 2, the limiting distribution in Theorem 3 is not pivotal. Thus, we need k^* to find a theoretically correct asymptotic critical value. To deal with this issue, we attempt to construct a conservative test. Specifically, we modify T'_S as follows.

$$T_S = I\left(\widehat{f(h_2)} > d_0\right) T'_S$$

where d_0 is a finite positive constant and $I(\cdot)$ is the indicator function. Let $t'_\alpha(k^*)$ be the $(1-\alpha)$th quantile of T'_S. Define $t'_{\alpha,0}$ as

$$t'_{\alpha,0} = \sup_{k^* \in (d_1, \infty)} t'_\alpha(k^*)$$

where d_1 is a given constant.

Corollary 1 shows that the test rejecting $H_0 : p(y_0) = 0$ against $H_1 : p(y_0) > 0$ when $T' > t'_{\alpha,0}$ has an asymptotic size not greater than α and is consistent.

Corollary 1. *Suppose Assumptions 1, 2C, and 3C hold and* $d_1 \leq c_0 d_0$*. Then*

$$Pr\left(T_S > t'_{\alpha,0} \mid p(y_0) = 0\right) \leq \alpha \tag{11}$$

and

$$Pr\left(T_S > t'_{\alpha,0} \mid p(y_0) > 0\right) \to 1 \tag{12}$$

as $n \to \infty$*.*

In implementing this test, two constants d_0 and d_1 are used. To understand the role of these constants, set d_1 to be 0. Then $t'_{\alpha,0} = \sup_{k^* \in \mathcal{R}^+} t'_\alpha(k^*)$. Let q_{1,k^*} be the probability that $W^+(k^*)$ is 1. That is, $q_{1,k^*} = Pr(W^+(k^*) = 1)$. The $(1-\alpha)$th quantile of $W^+(k^*)$ is 1 for any $\alpha \geq 1 - q_{1,k^*}$ As a result, we have

$$t'_\alpha(k^*) = \frac{1-k^*}{\sqrt{k^*}} \quad \text{for} \quad \alpha \geq 1 - q_{1,k^*} \ , \ k^* \approx 0,$$

which increases to infinity as $k^* \to 0$. For each $\alpha \in (0,1)$, there exists $\bar{k}^*$ such that $\alpha \geq 1 - q_{1,k^*}$ for all $k^* \leq \bar{k}^*$. Therefore, $t'_{\alpha,0} = \sup_{k^* \in \mathcal{R}^+} t'_\alpha(k^*) = \infty$ for all $\alpha \in (0,1)$ when $d_1 = 0$. To resolve this problem, we take the supremum of $t'_\alpha(k^*)$ over the set of k^* greater than d_1. This modification generates a finite critical value, which guarantees the required rejection probability (11) for any $k^* \geq d_1$, that is, $f_2(y_0) \geq d_1/c_0$ under $H_0 : p(y_0) = 0$. For $k^* < c_0 d_0$, that is, $f_2(y_0) < d_0/c_0$, the null hypothesis is rejected with probability approaching 0, and thus (11) is also satisfied. However, the inequality (11) is not guaranteed for $k^* \in (c_0 d_0, d_1)$ when d_1 is greater than $c_0 d_0$. Such possibility is removed by setting $d_1 \leq c_0 d_0$.

5 Consistent procedure to estimate a mass points set

As implied by (6) and (7), a standard kernel method does not deliver consistent estimators of the pmf $f_1(y)$ and pdf $f_2(y)$ associated with the true data generating process (1) if there are probability mass points. The discontinuity of the underlying distribution function must be taken care of in order to obtain consistent estimators of these density functions. Therefore, detecting the set of discontinuity points, that is, the probability mass points, is essential to construct consistent estimators of $f_1(y)$ and $f_2(y)$. To estimate the set of probability mass points D consistently, we use the PMP tests T_L and T_S.

Let $S_Y = \{y_{s,1}, \ldots, y_{s,n_s}\}$ be the empirical support where n_s is the number of distinct elements of S_Y. At each element of S_Y, we test the probability mass. Specifically, we consider the following n_s multiple hypotheses testing problem.

$$H_0^{(i)} : p(y_{s,i}) = 0 \quad \text{vs.} \quad H_1^{(i)} : p(y_{s,i}) > 0 \quad , \quad i = 1, 2, \cdots, n_s$$

After conducting this test, we collect the rejected hypotheses to define $\hat{D}$. That is,

$$\hat{D} = \{y_{s,i}, i = 1, \ldots, n_s : H_0^{(i)} \text{ is rejected } \} \tag{13}$$

Recognize that

$$D = \{y_{s,i}, i = 1, \ldots, n_s : p(y_{s,i}) > 0\}.$$

Therefore, consistent estimation of D requires that as $n \longrightarrow \infty$,

$$P\left(\text{ Reject } H_0^{(i)} \;\mid\; H_0^{(i)} \right) \longrightarrow 0 \tag{14}$$

and

$$P\left(\text{ Reject } H_0^{(i)} \;\mid\; H_1^{(i)} \right) \longrightarrow 1 \tag{15}$$

for each $i, i = 1, 2, \ldots, n_s$. A test satisfying (14) and (15) is said to be completely consistent (Andrews (1986)). The following testing procedure is suggested for the construction of $\hat{D}$.

1. Following Andrews (1986), set $N_1 = 1$ and N_ν, $\nu = 2, 3, \ldots$ as the smallest integer greater than $N_{\nu-1}$ such that for all $n \geq N_\nu$,

$$\Phi\left(z_{1-\frac{1}{\nu^{3/2}}} - \frac{\sqrt{n}}{\nu} \right) \leq \frac{1}{\nu^{3/2}}. \tag{16}$$

   Solving (16) provides that

$$N_\nu \geq 4\nu^2 z^2_{1-\frac{1}{\nu^{3/2}}}. \tag{17}$$

2. For each single i, compute the p-value of the null hypothesis $H_0^{(i)}$ and denote it by $\hat{p}_i$. and arrange the p-values in ascending order. That is, $\hat{p}_{(1)} \leq \hat{p}_{(2)} \leq \ldots \leq \hat{p}_{(n_s)}$.

3. Select the first ν hypotheses.

4. Among the chosen ν hypotheses, find the largest (positive) integer j satisfying

$$\hat{p}_{(i)} < \frac{\alpha_n}{\nu - i + 1} \text{ for all } i = 1, \ldots, j$$

   and denote it by j^* where $\alpha_n = 1/\sqrt{\nu}$ for $n \in [N_\nu, N_{\nu+1})$. Let $j^* = 0$ if

$$\hat{p}_{(1)} \geq \frac{\alpha_n}{\nu}.$$

5. Reject the null hypotheses associated with $\hat{p}_{(j)}$, $j \leq j^*$ and construct $\hat{D}$ as in (13). Do not reject any hypothesis and let $\hat{D} = \phi$ if $j^* = 0$.

Theorem 3. *Suppose that assumptions for Theorem 1 (or Theorem 2, resp.) are satisfied. Then $\hat{D}$ constructed by T_L (or T_S, resp.) is consistent to D in the sense of*

$$P(\hat{D} \neq D) \longrightarrow 0 \quad as \quad n \to \infty$$

This proposed procedure can be easily conducted when using the first version of the PMP test T_L. However, there is a practical issue in using the second version of the PMP test P_S since the distribution of the PMP test statistic depends on an unknown nuisance parameter. As a result, the exact expressions for theoretical size and power at a given sample size are unknown so that finding N_ν best suiting to P_S is tricky. So we use the same inequalities (16) and (17) to find N_ν. This solution does not harm the consistency of our procedure with T_S as provided in the proof although there might be a better choice of N_ν to control for the size and power in a balanced way.

6 Simulations

In this section, we numerically evaluate the performance of proposed tests in this work. Samples are drawn independently from a model (1) where $F_1(\cdot)$ is a discrete uniform distribution function with $D = \{-1, 0, 1\}$ and $F_2(y)$ is a standard normal distribution function. We independently conduct 1000 replications for various combinations of sample size n and p: $p = 0, 0.05, 0.1, 0.2$, and $n = 500, 1000, 2000$.

Table 1 shows the performance of the KS test to globally detect mass points. The null hypothesis $H_0 : p = 0$ is rejected at 5% significance level if $KS > 1.36$ and at 1% significance level if $KS > 1.63$. When $p = 0.2$, the KS test detects well the presence of mass points even at moderate sample sizes. However, for small values of p or small sample size, the KS test does not exhibit satisfactory rejection probabilities. Especially, the KS test has no power at the given sample sizes when $p = 0.05$. Last two columns in Table 1 are the actual size of the KS test. Interestingly, all rejection probabilities are 0 in the current set-up.

Table 2 displays the actual rejection probabilities of two versions of PMP tests at mass points 0 and 1, and a non-mass point 0.5.[3] For the bandwidth of T_L, we choose $h_1 = 0.5n^{-0.4}$ and $h_2 = n^{-0.4}$. For the bandwidth of T_S, we set $h_1 = 1/n$ and $h_2 = n^{-0.2}$. As expected, the rejection probabilities of T_L are observed to be large when p is large. For example, when $p = 0.2$, the rejection probabilities are close to 1 for all sample sizes n. The rejection probability at mass points 0 and 1 increases with the sample size. As a result, the power at these points is also close to 1 for all p when $n = 2000$. In contrast to the KS test, T_L has nontrivial rejection probabilities even when $p = 0.05$. Under the null hypothesis $H_0 : p = 0$, the actual rejection probabilities turn out to be close to nominal sizes. Rejection probabilities of T_S exhibit a similar pattern. However, interestingly T_S has greater power at mass points and actual size of nearly zero at a continuous point in most cases.

The performance of bootstrap PMP test $T_{L,B}$ in (9) is summarized in Table 3. As predicted by Proposition 1, the rejection probabilities of $T_{L,B}$ are similar to ones of T_L. Thus, there is no significant improvement in power and size distortion.

Table 4 presents the size properties of T_L and T_S at non-mass points neighboring to a mass point 0. Especially, we consider five non-mass points $0.00001, 0.0001, 0.001, 0.01$, and 0.1. When a non-mass point is sufficiently away from 0, T_S has at least as small false detection rates as T_L for all combinations of p and n in the experiment. Such better performance of T_S is most striking at a point 0.01. At this point T_S is not likely to falsely detect this point while T_L generates false detection with a rate close to 1. As a non-mass point becomes close to 0, both tests T_S and T_L make more errors, and thus suffer from large size distortion. At one glance, larger sample size seems to aggravate the size distortion problem as shown at a point 0.0001. However, this observation misleads the true effect of the sample size n. In fact, the

[3] Because of the symmetry of the distribution, the rejection probabilities at -1 turn out to be similar to ones at 1. Thus, we do not report the results for another mass point -1.

size distortion disappears eventually as the sample size n increases so that the bandwidths h_1 and h_2 becomes sufficiently smaller than the distance between a non-mass point and 0.[4] This fact explains the better performance of T_S at points such as 0.001, 0.01 and 0.1.

The performance of proposed procedure in section 5 is summarized in Table 5. We report the actual probability that the mass points set is correctly detected (CT)[5]. We also report two other measures FR and TR where FR is the actual probability that some non-mass points are contained in the estimated mass points set $\widehat{D}$, and TR is the actual probability that all mass points are contained in the estimated mass points set $\widehat{D}$[6]. In 1000 replications, the procedure using T_L never correctly detects the set of mass points. Such failure mainly results from the false rejection rate FR, which is explained by the poor size property of T_L around a mass point presented in Table 5. In contrast, the correct detection rate CT improves significantly when T_S is used. At the current sample sizes, the correct detection rates CT are relatively low. However, the correct detection rate CT is observed to increase with the sample size n as implied in Theorem 3. These results advocate using T_S in the procedure for detecting the set of mass points D.

7 Application to PSID wage data

[4]In our example, a kernel density estimator of a non-mass point y_0 is influenced by observations on 0 only when $|y_0| \leq h_1/2$. Recall that h_1 decreases with the sample size n. Therefore, h_1 becomes eventually less than $|y_0|$, and thus the related size distortion problem disappears.

[5]Let $\hat{D}_r$ be the estimated set of mass points in the r-th replication, and R be the total number of replications in the experiment. Then the actual probability of consistent detection of D is computed as $CT = R^{-1}\sum_{r=1}^{R} I(\widehat{D}_r = D)$.

[6]Despite the failure, the procedure using T_L could still provide useful information on the presence of mass points. Table A.1. implies that non-mass points in a neighborhood of a mass point would be collectively detected as mass points. Therefore, if $\widehat{D}$ includes most points in an interval, this fact should be regarded as a warning for the presence of a mass point.

As an empirical application, we apply the proposed procedure in Section 5 to the Panel Study of Income Dynamics (PSID) wave 2013 income data. Many studies recognized that the distribution of reported incomes is massed at certain values (e.g., Hanisch(2005), Zinn and Würbachac (2016), Barreca, Lindo, and Waddell (2016), Chetty et al. (2011)). Motivated by these studies, we apply the KS test and the proposed procedure to the Panel Study of Income Dynamics (PSID) income data. Specifically, we use the wages and salaries of a head of a household in the PSID 2013 wave. The PSID 2013 waves includes 9063 household heads. Among them, 336 observations are dropped since their wages or salaries are missing. Finally our sample includes 8727 observations. In the following analysis, we divide the reported wages by 100000.

The solid line in Figure 1 shows the profile of kernel density estimates without considering the presence of mass points. A high spike is observed near 0 since approximately 25% individuals report zero income. The KS statistic is 23.1, which strongly supports the idea of mass points in the wage distribution. In implementing the procedure, we use T_S since the numerical experiment shows the better performance of this version of PMP test.

According to the proposed procedure, 121 unique elements are detected as mass points. Most of these values turn out to be multiples of \$1000. This result confirms the digit preference pattern in self-reported income (Zinn and Würbach (2016)). The linking parameter p is estimated to be 0.824. After removing the estimated mass points in $\hat{D}$, we construct the estimates of the density at non-mass points $(1-p)f_2(y)$. The dotted line presents these kernel density estimates taking care of mass points. A noticeable distance between those two density estimates is found in the PSID wage data. It implies that ignoring mass points may cause a large finite sample bias.

8 Concluding Remarks

The purpose of this paper is to develop a consistent procedure to find the mass points set among observations. To achieve one part of the purpose, two tests are proposed to detect the presence of mass points in the data. One is the KS test, which enables us to globally detect the existence of mass points. The other is the PMP test, which can be used to distinguish whether a specific point of interest has mass or not. Depending on the order of two bandwidths, two versions of PMP tests are proposed. For the other part of the purpose, we propose a consistent detection procedure by combining the idea of a completely consistent test in Andrews (1986) and multiple testing procedure in Holm (1979). The PMP tests are used for the implementation of the procedure. Simulation results show that the PMP tests and the consistent detection procedure work fine even even when the sample size is not so large.

We expect that PMP tests and the detection procedure can be useful as a pretest when researchers are doing some kernel based nonparametric analysis. For example, the PMP tests can be used to test the discontinuity of a running variable at a threshold when doing regression discontinuity (RD) design analysis. Depending on the PMP test result at the threshold, the researcher can make decision about whether observations are appropriate for RD analysis. In this sense, this study complements the recent RD related nonparametric estimation research such as McCrary (2008) and Otsu et al.(2013). As another example, if some observations are suspected to have mass points for some known or unknown reasons and researchers are interested in kernel density estimation, then the consistent detection procedure can be very useful as shown in the application of household income density.

Appendix

A Proofs

Lemma A1. *Suppose Assumptions 1, 2A, and 3A hold. Then equations (4) and (5) are true.*

Proof. The mean of $\widehat{f(h)}$ is given by

$$E\left[\widehat{f(h)}\right] = \frac{1}{h}\sum_{d_j \in D} K\left(\frac{y_0 - d_j}{h}\right) p(d_j) + (1-p)\int \frac{1}{h}K\left(\frac{y_0 - y}{h}\right) f_2(y)dy$$

Recognize that

$$\frac{1}{h}K\left(\frac{y_0 - d_j}{h}\right) = \frac{1}{h}K(0) \qquad \text{for } y_0 = d_j, \tag{18}$$

$$\frac{1}{h}K\left(\frac{y_0 - d_j}{h}\right) \longrightarrow 0 \qquad \text{for } y_0 \neq d_j. \tag{19}$$

Then an equation (4) follows from combining (18), (19), and the well-known result for the mean of a kernel density estimator without the presence of mass points.

The variance of $\widehat{f(h)}$ can be written as

$$V\left[\widehat{f(h)}\right] = \frac{1}{nh^2}E\left[K^2\left(\frac{y_0 - Y_i}{h}\right)\right] - \frac{1}{n}\left(E\left[\frac{1}{h}K\left(\frac{y_0 - Y_i}{h}\right)\right]\right)^2 \tag{20}$$

Follow the same way in the derivation of the mean component. Then we have

$$\frac{1}{h}E\left[K^2\left(\frac{y_0 - Y_i}{h}\right)\right] = \frac{1}{h}K^2(0)p(y_0) + (1-p)f_2(y_0)\int k^2(s)ds + o(1)$$

Plugging this result and (4) into (20), and rearranging terms provides (5). ■

To prove the Theorem 1, we will investigate the asymptotic properties of the difference of the two kernel density estimators in Lemmas A2 and A3. With (4) and (5), let us find the asymptotic mean and variance of $\widehat{f(h_1)} - \widehat{f(h_2)}$.

Lemma A2. *Suppose that Assumptions 1 2B and 3A are satisfied. The asymptotic mean and variance of $\widehat{f(h_1)} - \widehat{f(h_2)}$ are given by*

$$E\left[\widehat{f(h_1)} - \widehat{f(h_2)}\right] = \frac{1}{h_1}\mu_1 + (1-p)\frac{h_1^2}{2}\left(1 - \frac{1}{c^2}\right) f_2^{(2)}(y_0)\int s^2 K(s)ds + o\left(p(y_0) + h_1^2\right)$$

and

$$V\left(\widehat{f(h_1)}-\widehat{f(h_2)}\right)=\frac{1}{nh_1^2}V_1+\frac{1}{nh_1}V_0+o\left(p(y_0)\frac{1}{nh_1^2}+\frac{1}{nh_1}\right)$$

where $\mu_1=(1-c)K(0)p(y_0)$, $V_0=(1-p)f_2(y_0)\left((1+c)\int K^2(s)ds-2c\int K(s)K(cs)ds\right)$, *and* $V_1=(1-c)^2K^2(0)p(y_0)(1-p(y_0))$.

Proof. By using an equation (4), we have

$$E\left[\widehat{f(h_1)}-\widehat{f(h_2)}\right]=\frac{1}{h_1}(1-c)K(0)p(y_0)+(1-p)\frac{h_1^2}{2}\left(1-\frac{1}{c^2}\right)f_2^{(2)}(y_0)\int s^2K(s)ds+o\left(p(y_0)+h_1^2\right)$$

The variance of $\widehat{f(h_1)}-\widehat{f(h_2)}$ is written as

$$V\left(\widehat{f(h_1)}-\widehat{f(h_2)}\right)=V\left(\widehat{f(h_1)}\right)+V\left(\widehat{f(h_2)}\right)-2Cov\left(\widehat{f(h_1)},\widehat{f(h_2)}\right)$$

The first two terms immediately follow from (5). Let $w_{1i}=\frac{1}{h_1}K\left(\frac{y_i-y_0}{h_1}\right)$ and $w_{2i}=\frac{1}{h_2}K\left(\frac{y_i-y_0}{h_2}\right)$. The last term is written as

$$\begin{aligned}Cov\left(\widehat{f(h_1)},\widehat{f(h_2)}\right)=&\frac{1}{n}E\left(w_{1i}w_{2i}\right)-\frac{1}{n}E\left(w_{1i}\right)E\left(w_{2i}\right)\\=&\frac{1}{n}E\left(w_{1i}w_{2i}\right)-\frac{1}{nh_1h_2}K^2(0)p(y_0)^2+o\left(p(y_0)\frac{1}{nh_1^2}+\frac{1}{nh_1}\right)\end{aligned}\tag{21}$$

where the second equality follows from (4). To evaluate $E(w_{1i}w_{2i})$, apply the law of iterated expectation. Then,

$$E\left(w_1w_2\right)=\frac{1}{h_1h_2}K(0)^2p(y_0)+(1-p)\frac{1}{h_2}\int K\left(s\right)K\left(\frac{h_1}{h_2}s\right)f_2(y_0-h_1s)ds+o(1)\tag{22}$$

because of (18) and (19). The second term is decomposed into two parts.

$$\begin{aligned}&(1-p)\frac{1}{h_2}\int K\left(s\right)K\left(\frac{h_1}{h_2}s\right)f_2(y_0-h_1s)ds\\&=(1-p)\frac{1}{h_2}\int K\left(s\right)K\left(cs\right)f_2(y_0-h_1s)ds+(1-p)\frac{1}{h_2}\int K\left(s\right)\left(K\left(\frac{h_1}{h_2}s\right)-K(cs)\right)\right)f_2(y_0+h_1s)ds\end{aligned}$$

Firstly,

$$\begin{aligned}\frac{1}{h_2}\int K\left(s\right)K\left(cs\right)f_2(y_0-h_1s)ds=&\frac{1}{h_2}\int K\left(s\right)K\left(cs\right)\left(f_2(y_0)-h_1sf_2'(y^*)\right)ds\\=&\frac{1}{h_2}f_2(y_0)\int K\left(s\right)K\left(cs\right)ds+o\left(\frac{1}{h_2}\right)\end{aligned}$$

where y^* lies between y_0 and $y_0 - h_1 s$ Secondly, notice that the Lipschitz-continuity of a kernel $K(\cdot)$ implies

$$\left| K\left(\frac{h_1}{h_2}s\right) - K(cs)\right| \leq d^* \left|\frac{h_1}{h_2} - c\right| |s|$$

for a constant d^*. Therefore, we have

$$\begin{aligned}\left|\frac{1}{h_2}\int K(s)\left(K\left(\frac{h_1}{h_2}s\right) - K(cs)\right) f(y_0 + h_1 s)ds\right| \leq & \frac{d^*}{h_2}\left|\frac{h_1}{h_2} - c\right| \int |s||K(s)|\, f(y_0 + h_1 s)ds \\ =& o\left(\frac{1}{h_2}\right)\end{aligned}$$

where the last equality comes from properties of a kernel $K(\cdot)$ described in the definition. Plug these results into (22) to obtain

$$E(w_1 w_2) = \frac{1}{h_1 h_2} K(0)^2 p(y_0) + (1-p)\frac{1}{h_2} f_2(y_0) \int K(s)\, K(cs)\, ds + o\left(\frac{1}{h_2}\right)$$

Plugging this equation into (21) provides

$$\begin{aligned}Cov\left(\widehat{f(h_1)}, \widehat{f(h_2)}\right) =& \frac{1}{nh_1 h_2} K(0)^2 p(y_0)(1 - p(y_0)) \\ &+ (1-p)\frac{1}{nh_2} f_2(y_0) \int K(s)\, K(cs)\, ds + o\left(p(y_0)\frac{1}{nh_1^2} + \frac{1}{nh_1}\right)\end{aligned}$$

This result together with (5) imply that

$$\begin{aligned}V\left(\widehat{f(h_1)} - \widehat{f(h_2)}\right) =& \frac{1}{nh_1^2}(1-c)^2 K(0)^2 p(y_0)(1-p(y_0)) \\ &+ \frac{1}{nh_1}(1-p) f_2(y_0)\left((1+c)\int K(s)^2 ds - 2c\int K(s)K(cs)ds\right) \\ &+ o\left(p(y_0)\frac{1}{nh_1^2} + \frac{1}{nh_1}\right) \qquad \blacksquare\end{aligned}$$

Lemma A3. *Suppose that Assumptions 1, 2B, and 3A are satisfied. In addition, assume,* $h_1/h_2 \to c \in (0,1)$ *as* $n \to \infty$. *Also assume that for some* $\delta > 0$, *a kernel* $K(\cdot)$ *satisfies*

$$\int |K(s)|^{2+\delta}\, ds < \infty$$

Then for $c \in (0,1)$

$$T_L'' = \left(\widehat{f(ch)} - \widehat{f(h)} - E\left[\widehat{f(ch)} - \widehat{f(h)}\right]\right) / \sqrt{V\left(\widehat{f(ch)} - \widehat{f(h)}\right)} \longrightarrow_d N(0,1) \tag{23}$$

as $n \to \infty$.

Proof. Write T_L'' as follows.

$$\begin{aligned} T_L'' =& \sum_{i=1}^{n} \frac{w_{1i} - w_{2i}}{\sqrt{nV(w_{1i} - w_{2i})}} \\ :=& \sum_{i=1}^{n} L_{n,i} \end{aligned}$$

where $w_{1i} = \frac{1}{h_1} K\left(\frac{y_0 - Y_i}{h_1}\right) - E\left[\frac{1}{h_1} K\left(\frac{y_0 - Y_i}{h_1}\right)\right]$ and $w_{2i} = \frac{1}{h_2} K\left(\frac{y_0 - Y_i}{h_2}\right) - E\left[\frac{1}{h_2} K\left(\frac{y_0 - Y_i}{h_2}\right)\right]$.

By construction, $L_{n,i}$, $i = 1, 2, \cdots, n$ are *i.i.d.* random variables satisfying $E(L_{n,i}) = 0$ and $V(L_{n,i}) = 1/n$. From these facts it follows that

$$\frac{V(L_{n,i})}{V(T_L'')} = \frac{1}{n} \longrightarrow 0 \quad \text{as} \quad n \to \infty$$

Notice that this result together with a condition

$$\sum_{i=1}^{n} E|L_{n,i}|^{2+\delta} < \infty, \quad \delta > 0 \tag{24}$$

are the conditions of Liapounov's central limit theorem so that (23) is shown to be satisfied. Therefore, let us show that (24) is satisfied.

We have

$$\begin{aligned} \sum_{i=1}^{n} E|L_{n,i}|^{2+\delta} =& \left(\frac{1}{\sqrt{nV(w_{1i} - w_{2i})}}\right)^{2+\delta} \sum_{i=1}^{n} E|w_{1i} - w_{2i}|^{2+\delta} \\ \leq& n2^{1+\delta} \left(\frac{1}{\sqrt{nV(w_{1i} - w_{2i})}}\right)^{2+\delta} \left(E|w_{1i}|^{2+\delta} + E|w_{2i}|^{2+\delta}\right) \end{aligned} \tag{25}$$

where the inequality comes from the c_r inequality. Use the the c_r inequality and the law of iterated expectation to obtain

$$\begin{aligned} E|w_{1i}|^{2+\delta} \leq& 2^{2+\delta} E\left|K\left(\frac{y_0 - Y_i}{h_1}\right)\right|^{2+\delta} \\ =& 2^{2+\delta}\left(\frac{1}{h_1^{2+\delta}} K(0)^{2+\delta} p(y_0) + (1-p)\left(\frac{1}{h_1}\right)^{1+\delta} f_2(y_0) \int |K(s)|^{2+\delta} ds + o\left(h_1^{-1-\delta}\right)\right) \end{aligned}$$

where the last equality results from Lemma 2.1 in Pagan and Ullah (1999). A similar result is obtained for $E|w_{2i}|^{2+\delta}$. Thus, the term in the right-hand side in (25) is bounded by

$$n2^{3+\delta}\left(\frac{1}{\sqrt{nV(w_{1i}-w_{2i})}}\right)^{2+\delta}$$
$$\times\left(p(y_0)K(0)^{2+\delta}\left(h_1^{-2-\delta}+h_2^{-2-\delta}\right)+(1-p)f_2(y_0)\int|K(s)|^{2+\delta}ds\left(h_1^{-1-\delta}+h_2^{-1-\delta}\right)+o\left(h_1^{-1-\delta}\right)\right). \tag{26}$$

Firstly, consider a case where $p(y_0)=0$. In such a case, (26) is written as

$$n2^{3+\delta}\left(\frac{1}{\sqrt{nV(w_{1i}-w_{2i})}}\right)^{2+\delta}\left((1-p)f_2(y_0)\int|K(s)|^{2+\delta}ds\left(h_1^{-1-\delta}+h_2^{-1-\delta}\right)+o\left(h_1^{-1-\delta}\right)\right),$$

which is again bounded by

$$\begin{aligned}
&2^{3+2\delta}n\left(\frac{1}{\sqrt{nV(w_{1i}-w_{2i})}}\right)^{2+\delta}(h_1^{-1-\delta}+h_2^{-1-\delta})\left((1-p)f_2(y_0)\int|K(s)|^{2+\delta}ds+o(1)\right)\\
=&2^{3+2\delta}\left(\frac{1}{\sqrt{h_1V(w_{1i}-w_{2i})}}\right)^{2+\delta}(nh_1)^{-\delta/2}\left((1-p)f_2(y_0)\int|K(s)|^{2+\delta}ds+o(1)\right)\\
&+2^{3+2\delta}\left(\frac{1}{\sqrt{h_2V(w_{1i}-w_{2i})}}\right)^{2+\delta}(nh_2)^{-\delta/2}\left((1-p)f_2(y_0)\int|K(s)|^{2+\delta}ds+o(1)\right).
\end{aligned}$$

Lemma A.2 implies $h_jV(w_{1i}-w_{2i})=nh_jV(\widehat{f(h_1)}-\widehat{f(h_2)})=O(1)$, $j=1,2$. Now recall the assumptions $\int|K(s)|^{2+\delta}ds<\infty$ and $nh_j\to\infty$, $j=1,2$ as $n\to\infty$ to conclude that

$$\lim_{n\to\infty}\sum_{i=1}^{n}E|L_{n,i}|^{2+\delta}=0<\infty$$

when $p(y_0)=0$.

Secondly, consider a case where $p(y_0)\neq 0$. (26) is written as

$$n2^{3+\delta}\left(\frac{1}{\sqrt{nV(w_{1i}-w_{2i})}}\right)^{2+\delta}\left(p(y_0)K(0)^{2+\delta}\left(h_1^{-2-\delta}+h_2^{-2-\delta}\right)+o\left(h_1^{-1-\delta}\right)\right)$$

Similar with the previous case, this term is shown to be bounded by

$$2^{3+2\delta} n^{-\delta/2} \left(\frac{1}{\sqrt{h_1^2 V(w_{1i} - w_{2i})}} \right)^{2+\delta} \big(p(y_0) K(0)^{2+\delta} + o(1) \big)$$
$$+ 2^{3+2\delta} n^{-\delta/2} \left(\frac{1}{\sqrt{h_2^2 V(w_{1i} - w_{2i})}} \right)^{2+\delta} \big(p(y_0) K(0)^{2+\delta} + o(1) \big)$$

Remember that $nh_j^2 V(\hat{f}(h_1) - \hat{f}(h_2)) = h_j^2 V(w_{1i} - w_{2i}) = O(1)$ when $p(y_0) \neq 0$. Therefore, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n} E|L_{n,i}|^{2+\delta} = 0 < \infty$$

when $p(y_0) \neq 0$. So the conditions for Liapounov's central limit theorem are satisfied for the case where $p(y_0) > 0$ and applying Liapounov's central limit theorem completes the proof. ■

Proof of Theorem 1.

From Lemmas A2 and A3, it immediately follows that under $H_0 : p(y_0) = 0$,

$$\begin{aligned} T_L =& T_L'' + \frac{\sqrt{nh_1}}{\sqrt{\hat{V}_0}} \Big(E(\widehat{f(h_1)}) - E(\widehat{f(h_2)}) \Big) \\ =& T_L'' + o(1) \end{aligned}$$

where the second equality follows from Lemma A2, and $nh_1h_2^4 = o(1)$ by assumption.

Suppose that $p(y_0) > 0$, which is equivalent to $p > 0$ and $f_1(y_0) > 0$. In this case, it can be shown that

$$\begin{aligned} T_L =& T_L'' + \frac{\sqrt{nh_1}}{\sqrt{\hat{V}_0}} \Big(E(\widehat{f(h_1)}) - E(\widehat{f(h_2)}) \Big) \\ =& T_L'' + \frac{\sqrt{n}}{\sqrt{h_1 \hat{V}_0}} (\mu_1 + o(1)) \end{aligned} \tag{27}$$

where the second equality comes from Lemma A2. (27) is shown to diverge as n by using Lemma A3 and $h\widehat{f(h_2)} = K(0)p(y_0) + o_p(1)$.

■

To prove the Proposition 1, we firstly show that the bootstrap estimator $\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)})$ is consistent.

Lemma A4. *Let p^* be the probability measure associated with the empirical distribution of a given sample $(Y_1, Y_2, \cdots, Y_n)$. Under Assumptions 1A, 2B and 3B,*

$$\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}) \xrightarrow{p^*} \frac{1}{n^2}\sum_i \left(\frac{1}{h_1}K_i(h_1) - \frac{1}{h_2}K_i(h_2)\right)^2 - \frac{1}{n}\left(\widehat{f(h_1)} - \widehat{f(h_2)}\right)^2 + O_{p^*}\left(\frac{1}{B}\right)$$

as $B \to \infty$.

Proof. Let E^* and V^* be the expectation and variance involved with p^*, respectively. The bootstrap estimator $\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)})$ is written as

$$\begin{aligned}\widehat{V}^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right) =& \frac{1}{B-1}\sum_b \left[\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)} - E^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right)\right]^2 \\ & - \frac{B}{B-1}\left[\overline{\widehat{f^*(h_1)}} - \overline{\widehat{f^*(h_2)}} - E^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right)\right]^2 \end{aligned} \tag{28}$$

The WLLN implies

$$\overline{\widehat{f^*(h_1)}} \xrightarrow{p^*} E^*\left(\widehat{f_b^*(h_r)}\right) \tag{29}$$

and

$$\frac{1}{B-1}\sum_b \left[\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)} - E^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right)\right]^2 \xrightarrow{p^*} V^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right) \tag{30}$$

for any given $(Y_1, Y_2, \cdots, Y_n)$. Moreover,

$$\overline{\widehat{f_b^*(h_1)}} - \overline{\widehat{f_b^*(h_2)}} - E^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right) = O_{p^*}\left(\frac{1}{\sqrt{B}}\right).$$

Let $N_i^*(b)$ be the number of random drawings of Y_i from the sample with replacement to obtain the b-th bootstrapped sample. Then $(N_1^*(b), \cdots, N_n^*(b))$ follows a multinomial distribution with n trials and population proportion $\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$. With the use of the fact, $E^*\left(\widehat{f_b^*(h_r)}\right)$ and $V^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right)$can be evaluated.

Firstly, for $r = 1, 2$

$$
\begin{aligned}
E^*\left[\widehat{f_b^*(h_r)}\right] =& \frac{1}{nh_r} \sum_i E^*\left[N_i^*(b)\right] K\left(\frac{y_0 - Y_i}{h_r}\right) \\
=& \widehat{f(h_r)}
\end{aligned} \tag{31}
$$

Secondly,

$$
\begin{aligned}
V^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right) =& \frac{1}{n^2} \sum_{i,j} \left(\frac{1}{h_1}K_i(h_1) - \frac{1}{h_2}K_i(h_2)\right)\left(\frac{1}{h_1}K_j(h_1) - \frac{1}{h_2}K_j(h_2)\right) Cov^*\left(N_i^*(b), N_j^*(b)\right) \\
=& \frac{1}{n^2} \sum_i \left(\frac{1}{h_1}K_i(h_1) - \frac{1}{h_2}K_i(h_2)\right)^2 \left(1 - \frac{1}{n}\right) \\
&+ \frac{1}{n^2} \sum_{i \neq j} \left(\frac{1}{h_1}K_i(h_1) - \frac{1}{h_2}K_i(h_2)\right)\left(\frac{1}{h_1}K_j(h_1) - \frac{1}{h_2}K_j(h_2)\right)\left(-\frac{1}{n}\right) \\
=& \frac{1}{n^2} \sum_i \left(\frac{1}{h_1}K_i(h_1) - \frac{1}{h_2}K_i(h_2)\right)^2 - \frac{1}{n^3}\left(\sum_i \frac{1}{h_1}K_i(h_1) - \frac{1}{h_2}K_i(h_2)\right)^2
\end{aligned} \tag{32}
$$

Use equations (29)-(32) together with (28) to have

$$
\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}) \xrightarrow{p^*} \frac{1}{n^2} \sum_i \left(\frac{1}{h_1}K_i(h_1) - \frac{1}{h_2}K_i(h_2)\right)^2 - \frac{1}{n}\left(\widehat{f(h_1)} - \widehat{f(h_2)}\right)^2 + O_{p^*}\left(\frac{1}{B}\right)
$$

as $B \longrightarrow \infty$. ■

Proof of Proposition 1.

Rewrite $T_{L,B}$ as follows.

$$
T_{L,B} = T_L \sqrt{\frac{\hat{V}_0}{nh_1 \widehat{V}^*\left(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}\right)}}
$$

By expanding the terms in Lemma A4, we have

$$\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}) = \frac{1}{n^2h_1^2}\sum_i K_i^2(h_1) - \frac{2}{n^2h_1h_2}\sum_i K_i(h_1)K_i(h_2) + \frac{1}{n^2h_2^2}\sum_i K_i^2(h_2) - \frac{1}{n}\left(\widehat{f(h_1)} - \widehat{f(h_2)}\right)^2 + O_{p^*}\left(\frac{1}{B}\right).$$

From this equation, it follows that

$$nh_1\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}) = \frac{1}{nh_1}\sum_i K_i^2(h_1) - \frac{2}{nh_2}\sum_i K_i(h_1)K_i(h_2) + \frac{h_1}{nh_2^2}\sum_i K_i^2(h_2) + o_p(1)$$

Note that both $\widehat{f(h_1)}$ and $\widehat{f(h_2)}$ are consistent under H_0. Moreover, Lemma A2 with the law of large numbers imply

$$\frac{1}{nh_1}\sum_i K_i^2(h_1) \xrightarrow{p} f(y_0)\int K^2(s)ds$$

$$\frac{1}{nh_2}\sum_i K_i(h_1)K_i(h_2) \xrightarrow{p} cf(y_0)\int K(s)K(cs)ds$$

,and

$$\frac{h_1}{nh_2^2}\sum_i K_i^2(h_2) \xrightarrow{p} cf(y_0)\int K^2(s)ds$$

By plugging these results and arranging terms, we have

$$nh_1\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}) \xrightarrow{p} V_0$$

Finally, remember that $\hat{V}_0 \xrightarrow{p} V_0$ under $H_0 : p(y_0) = 0$. Therefore,

$$\frac{\widehat{V}_0}{nh_1\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)})} \xrightarrow{p} 1 \quad \text{under} \quad H_0.$$

Hence, $T_{L,B}$ and T_L have the same asymptotic distribution under H_0.

Now consider the asymptotic distribution of $T_{L,B}$ under $H_1 : p(y_0) > 0$. From Lemma A4, it follows that

$$nh_1^2\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)}) = \frac{1}{n}\sum_i K_i^2(h_1) - \frac{2h_1}{nh_2}\sum_i K_i(h_1)K_i(h_2) + \frac{h_1^2}{nh_2^2}\sum_i K_i^2(h_2) - \left(h\widehat{f(h_1)} - h\widehat{f(h_2)}\right)^2 + O_{p^*}\left(\frac{1}{B}\right).$$

Under H_1, we have

$$\frac{1}{n}\sum_i K_i^2(h_1) \xrightarrow{p} p(y_0)K^2(0) \tag{33}$$

$$\frac{h_1}{nh_2}\sum_i K_i(h_1)K_i(h_2) \xrightarrow{p} cp(y_0)K^2(0) \tag{34}$$

,and

$$\frac{h_1^2}{nh_2^2}\sum_i K_i^2(h_2) \xrightarrow{p} c^2 p(y_0)K^2(0) \tag{35}$$

Also, the law of large numbers and Lemma A2 imply

$$h_1\widehat{V}_0 \xrightarrow{p} k_c K(0)p(y_0) \tag{36}$$

and

$$h\widehat{f(h_1)} - h\widehat{f(h_2)} \xrightarrow{p} (1-c)p(y_0)K(0) \tag{37}$$

By using equations (33)-(37), we show

$$\frac{h_1\hat{V}_0}{nh_1^2\widehat{V}^*(\widehat{f_b^*(h_1)} - \widehat{f_b^*(h_2)})} \xrightarrow{p} \frac{k_c}{(1-c)K(0)p(y_0)} \quad \text{under} \quad H_1 \tag{38}$$

Hence, under $H_1 : p(y_0) > 0$, the test statistic $T_{L,B}$ is given by

$$T_{L,B} = \frac{1}{\sqrt{(1-c)K(0)p(y_0)}} T_n + o_p(1)$$

In the proof of Theorem 1, T_L is shown to diverges to the infinity with probability 1 under H_1. Therefore, $T_{L,B}$ also diverges to the infinity with probability 1 under H_1. ■

Proof of Theorem 2.

Firstly, consider the asymptotic distribution of T_S' under H_0. As a first step, let us show equation (10). Let $W_i = 1\left(y_0 - \frac{h_1}{2} \le Y_i \le y_0 + \frac{h_1}{2}\right)$. Then

$$nh_1\widehat{f(h_1)} = \sum_{i=1}^n W_i \sim B\left(n, P_{h_1}(y_0)\right)$$

where $P_{h_1}(y_0)$ is a probability that Y_i is in an interval $[y_0 - h_1/2, y_0 + h_1/2]$. Because of twice continuous differentiability of $f_2(\cdot)$,

$$P_{h_1}(y_0) = h_1 f_2(y_0) + O(h_1^2)$$

, which immediately implies

$$nP_{h_1}(y_0) = nh_1 f_2(y_0) + O(nh_1^2) \to k^* \tag{39}$$

as $n \to \infty$. For a fixed n, the characteristic function of $nh_1\widehat{f(h_1)}$, denoted by $\psi_n(t)$, is given by

$$\psi_n(t) = \left(1 - P_{h_1}(y_0) + P_{h_1}(y_0)e^{it}\right)^n$$

From equation (39) and the definition of e, it follows that

$$\psi_n(t) \to e^{k^*(e^{it}-1)}$$

, which is the characteristic function of a Poisson distribution with mean k^*. Remember that y_0 is a value in the empirical support so that $nh_1\widehat{f(h_1)}$ has a positive integer value. Therefore, it can be shown that

$$nh_1\widehat{f(h_1)} \xrightarrow{d} W^+(k^*)$$

Under given assumptions,

$$\widehat{f(h_2)} \xrightarrow{p} f(y_0) \quad \text{and} \quad \widehat{V} = k_c\widehat{f(h_2)} \xrightarrow{p} f(y_0)$$

Therefore, under $H_0 : p(y_0) = 0$,

$$T \xrightarrow{d} \frac{W^+(k^*) - k^*}{\sqrt{k^*}}$$

Now suppose $H_1 : p(y_0) > 0$. Remember that

$$h_j\widehat{f(h_j)} \xrightarrow{p} p(y_0) \quad \text{for} \quad j = 1, 2.$$

Rewrite T_s' as

$$T_S' = \sqrt{n}\sqrt{\frac{h_2}{h_1}}\left(\frac{h_1\widehat{f(h_1)}}{\sqrt{h_2\widehat{f(h_2)}}} - \frac{h_1}{h_2}\sqrt{h_2\widehat{f(h_2)}}\right) \tag{40}$$

diverges to infinity as $n \to \infty$. ■

Proof of Lemma1.

This result immediately follows from Theorem 3 and the consistency of $\widehat{f(h_2)}$. ■

Proof of Theorem 3.

Part 1. consistency of a procedure using T_L: To show the consistency of $\widehat{D}$, it is enough to prove that

$$P(y \in \widehat{D} \mid y \in C) \longrightarrow 0 \quad \text{as} \quad n \to \infty \tag{41}$$

and

$$P(y \in \widehat{D} \mid y \in D) \longrightarrow 1 \quad \text{as} \quad n \to \infty \tag{42}$$

For (41), consider the probability of an event of false rejection, R_F, that is, at least one of the remaining hypotheses is rejected. Note that the probability of an event R_F is the family-wise error rejection (FWER) rate. The suggested procedure is the Holm procedure so that the FWER rate is at most α_n. (41) is shown to be true by recognizing $\alpha_n = 1/\sqrt{\nu} = o(1)$.

Now consider the probability in (41). Let A_F be an event that at least one of $H_0^{(1)}, \ldots, H_0^{(|D|)}$ is not rejected. For (42), we attempt to show that the probability of A_F converges to 0 as n increases to the infinity. It can be easily shown that

$$P(A_F) \leq P(R_F) + P(A_F \cap R_F^c)$$

where R_F^c is the complement of R_F, that is, the set of no false rejection. The first term is the FWER so that it converges to 0 under given conditions. So let us focus on the second probability, $P(A_F \cap R_F^c)$. In the set $A_F \cap R_F^c$, a hypothesis associated with a non-mass point in C should have a p-value greater than $\hat{p}_{(j^*)}$. If there is a hypothesis associated with a non-mass point in C and its p-value, $\hat{p}_k$, is not greater than $\hat{p}_{(j^*)}$, this hypothesis would be rejected under the given testing procedure. This fact contradicts the definition of a set $A_F \cap R_F^c$. As a result, we have $\hat{p}_{(j)} = \hat{p}_{(i_j)}$, $j = 1, \ldots, j^*$ where $\hat{p}_{(i_j)}$ is the j-th smallest p-value among the hypotheses associated with the elements in D. In view of the relationship between D and $\widehat{D}$, $A_F \cap R_F^c$ is regarded as a set including all outcomes generating

$$\widehat{D} \subset D \quad \text{and} \quad \widehat{D} \neq D$$

There are two possible cases: (1) $\nu \leq J$ and (2) $j^* \leq J$. The first case happens when we do not test sufficient number of hypotheses. By construction of ν, ν increases to ∞ as $n \to \infty$ so that $\nu \geq J$ for sufficiently large n. Therefore, the probability of the first event converges to 0 as $n \to \infty$. So let us assume that $\nu \geq J$. Recognize that

$$\hat{p}_{(i_{j^*+1})} \geq \hat{p}_{(j^*+1)} \geq \frac{\alpha_n}{\nu - j^*}$$

when $\kappa_\nu \geq J$. Without loss of generality, let the first J values in the empirical support be D. That is,

$$D = \{y_{s,1}, \cdots, y_{s,J}\}$$

Then the probability of $A_F \cap R_F^c$ is written as follows.

$$\begin{aligned}
P(A_F \cap R_F^c) =& P\left(\hat{p}_{(1)} \geq \frac{\alpha_n}{\nu}\right) \\
&+ P\left(\hat{p}_{(1)} < \frac{\alpha_n}{\nu}, \hat{p}_{(2)} \geq \frac{\alpha_n}{\nu - 1}\right) + \cdots \\
&+ P\left(\hat{p}_{(1)} < \frac{\alpha_n}{\nu}, \hat{p}_{(2)} < \frac{\alpha_n}{\nu - 1}, \ldots, \hat{p}_{(J-1)} < \frac{\alpha_n}{\nu - (J-2)}, \hat{p}_{(J)} \geq \frac{\alpha_n}{\nu - (J-1)}\right) \\
\leq& \sum_{j^*=0}^{J-1} P\left(\hat{p}_{(j^*)} \geq \frac{\alpha_n}{\nu - j^*}\right) \\
\leq& J! \sum_{j=1}^{J} P\left(\hat{p}_j \geq \frac{\alpha_n}{\nu}\right) \\
\leq& J! \sum_{j=1}^{J} P\left(T_{j,n} \leq z_{1-\frac{\alpha_n}{\nu}}\right)
\end{aligned}$$

where $T_{j,n}$ is the PMP test statistic T_L of $H_0^{(j)}$, $j = 1, \ldots, J$. According to the proof of Theorem 2, for any $j = 1, \ldots, J$,

$$T_{j,n} = T_n'' + \sqrt{n} c_j + o_p(1)$$

where $T''_n \to_d N(0,1)$ and $c_j > 0$. So the last equation is approximately

$$J! \sum_{j=1}^{J} \Phi\left(z_{1-\frac{\alpha_n}{\nu}} - \sqrt{n}c_j\right)$$

where $c_j = (1-c)\sqrt{K(0)p(y'_j)/k_c}$. Let $c_* = \min_{j=1,\ldots,J} c_j$. The last term is bounded by

$$JJ!\Phi\left(z_{1-\frac{\alpha_n}{\nu}} - \sqrt{n}c_*\right) \tag{43}$$

where $c_* = \min_j c_j$. For any $c_* \in (0,\infty)$, there exists $N_{\nu'}$ satisfying

$$\sup_{c_* \geq 1/\nu} P(A_F) \leq JJ!\frac{1}{\upsilon^{3/2}} \quad \text{for any} \quad n \geq N_{\nu'} = N_{\max(\nu, 1/c_*)}.$$

Therefore, for all $c^* \in (0,\infty)$,

$$P(A_F) \to 0 \quad \text{as } n \to \infty$$

so that (42) is shown to complete the proof of the first part.

Part 2. consistency of a procedure using T_S: Despite the use of T_S, the probability of false rejection $P(R_F)$ is still α_n because of the property of the Holm procedure. By construction, $\nu \to \infty$ as n increases, and thus $P(R_F)$ converges to 0. The remaining part is showing that the probability of false acceptance $P(A_F)$ converges to 0. Due to the same reason in part 1, it is enough to show $P(A_F \cap R_F^c) \to 0$. Follow the same steps in part 1 to have

$$P(A_F \cap R_F^c) \leq J! \sum_{j=1}^{J} P\left(T_{j,n} \leq t'_{\frac{\alpha_n}{\nu},0}\right) \tag{44}$$

where $T_{j,n}$ is the PMP test statistic T_S of $H_0^{(j)}, j = 1,\ldots,J$ in Corollary 1. Denote the corresponding statistic T'_S in Theorem 3 by $T'_{j,n}$. When $p(y_{s,j}) > 0$,

$$\frac{1}{\sqrt{n^2 h_2}} T'_{j,n} \xrightarrow{p} \sqrt{\frac{1}{c_0} p(y_{s,j})}$$

because of (40). For any k, we can write the event $\{W(k) > n\}$ as an event $\{X_1 + \ldots + X_n < n\}$ where $\lambda(k)$ is a Poisson random variable with parameter k and X_i, $i = 1,\ldots,n$ are *i.i.d.*

exponential random variable with a parameter k. Applying the Cramer's theorem to the sum of X_i provides the following bound for the right tail probability.

$$P\left(W(k) > n\right) = O\left(e^{-nm_0}\right)$$

for a positive constant m_0. This fact implies for any $p(y_{s,j}) \geq 1/\sqrt[4]{n^2 h_2}$

$$P\left(W^{+}(k) > \sqrt{n^2 h_2 p(y_{s,j})/c_0}\right) \leq m_2 e^{-\sqrt[4]{n^2 h_2} m_3}$$

for some constants m_2 and m_3. Let $k_j^* = c_0 f_2(y_{s,j})$. Then under $H_0^{(j)}$,

$$\begin{aligned} P\left(T_{j,n} \geq \sqrt{n^2 h_2 p(y_{s,j})/c_0}\right) &\leq P\left(T'_{j,n} \geq \sqrt{n^2 h_2 p(y_{s,j})/c_0}\right) \\ &\leq P\left(W^{+}(k_j^*) \geq \sqrt{k_j^*}\sqrt{n^2 h_2 p(y_{s,j})/c_0}\right) \\ &\leq m_2 e^{-\sqrt[4]{n^2 h_2}\sqrt{k_j^*} m_3} \end{aligned}$$

Note that $m_2 e^{-\sqrt[4]{n^2 h_2} m_3} < \alpha_n/\nu = \nu^{-3/2}$ for any $k_j^* > d_1$ if n is sufficiently large. Therefore,

$$\sqrt{n^2 h_2 p(y_{s,j})/c_0} \geq t'_{\frac{\alpha_n}{\nu},0} \quad \text{for} \quad p(y_{s,j}) \geq 1/\sqrt[4]{n^2 h_2}$$

when n is sufficiently large. As a result, (44) also converges to 0 for any $p(y_{s,j}) \geq 1/\sqrt[4]{n^2 h_2}$. Now let $c_* = \min c'_j$ where $c'_j = p(y_{s,j})$. Then, for any $c_* \in (0, \infty)$ and $\varepsilon > 0$, there exists $N_{\nu'}$ satisfying

$$\sup_{c_* \geq 1/\sqrt[4]{n^2 h_2}} P(A_F) < \varepsilon \quad \text{for any} \quad n \geq N_{\nu'}.$$

Therefore, for all $c^* \in (0, \infty)$,

$$P(A_F) \to 0 \quad \text{as } n \to \infty$$

so that the proof of the part 2 is completed. ■

References

Andrews, D. W. (1986), “Complete Consistency: A Testing Analogue of Estimator Consistency”, *Review of Economic Studies*, 53, 263-269

Arulampalam, W., V. Corradi, and D. Gutknecht (2014), “Modelling Heaped Duration Data: An Application to Neonatal Motality ”, *IZA discussion paper series*

Bar, H. Y., and D. R. Lillard (2012), “Accounting for heaping in retrospective reported event data - A mixture-model approach ”, *Statistics in Medicine*, 31, 3347-3365

Barreca, A., J. M. Lindo, and G. R. Waddell (2016),“Heaping-induced bias in regression-discontinuity designs”, *Economic Inquiry*, 54, 268-293

Bruine de Bruin, W., B. Fischhoff, S. G. Millstein, and B. L. Halpern-Felsher (2000), “Verbal and Numerical Expressions of Probability: It’s a Fifty-Fifty Chance ”, *Organization Behavior and Human Decision Processes*, 81. 115-131

Bruine de Bruin, W., P. S. Fischbeck, N. A. Stiber, and B. Fischhoff (2002), “What number is “Fifty-Fifty”?: Redistributing Excessive 50% Responses in Elicited Probabilities ”, *Risk Analysis*, 22, 713-723

Budd, J.W. and T. Guinnane (1991) “Age-Misreporting, Age-Heaping, and the 1908 Old Age Pension Act in Ireland ”, *Population Studies*, 45, 497-518

Burgstahler, D. and I. Dichev (1997), “Earnings management to avoid earnings decreases and losses ”, *Journal of Accounting and Economics*, 24, 99-126

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011), “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records”, *The Quarterly Journal of Economics*, 126, 749-804.

Crawford, F. W., R. E. Weiss, and M. A. Suchard (2015), “Sex, lies and self-reported counts: Bayesian mixture models for heaping in longitudinal count data via birth-death processes”, *The Annals of Applied Statistics*, 9, 572-596.

Hanisch, J. U. (2005), “Rounded responses to income questions Allgemeines”, *Statistisches Archiv*, 89, 39-48

Heitjan, D. F., and D. B. Rubin (1991), "Ignorability and coarse data", *The Annals of Statistics*, 19, 2244-2253

Holbrook, A.L., S. Anand, T.P. Johnson, Y.I. Cho, S. Shavitt, N. Chávez, and S. Weiner (2014), "Response Heaping in Interviewer-Administered Surveys Is it Really a Form of Satisficing?", *Public Opinion Quarterly*, 78, 691-683

Holm, S. (1979), "A simple sequentially rejective multiple multiple test procedure", *Scandinavian Journal of Statistics*, 6, 65-70

McCrary, J. (2008), "Manipulation of the Running Variable in the Regression Discontinuity Design", *Journal of Econometrics*, 142, 698-714

Newey and West (1994), "Automatic lag selection in covariance matrix estimation", *Review of Economic Studies*, 61, 631-653

Otsu, D., Xu, K. and Y. Matsushita (2013), "Estimation and Inference of Discontinuity in Density ", *Journal of Business & Economic Statistics*, 31, 507-524

Otterbach, S. and A. Sousa-Poza (2010), "How Accurate are German Work-Time Data? A Comparison of Time-Diary Reports and Stylized Estimates", *Social Indicators Research*, 97, 325-339

Pagan, A. and A. Ullah (1999), "*Nonparametric Econometrics*", *Cambridge University Press*

Petoussi, K., R. D. Gill, and C. Zeelenberg (1997) "Statistical Analysis of Heaped Duration Data", Vrije University. Working Paper.

Pudney, S. (2008), "Heaping and leaping: Survey Response Behaviour and the Dynamics of Self-Reported Consumption Expenditure", *ISER Working Paper Series*

Saez, E. (2010), "Do Taxpayers Bunch at Kink Points?", *American Economic Journal: Economic Policy*, 2, 180-212

Serfling, R. J. (1980), "*Approximation Theorems of Mathematical Statistics*", *John Wiley and Sons*

Takeuchi, Y. (2004), "On a statistical method to detect discontinuity in the distribution function of reported earnings ", *Mathematics and Computers in Simulation*, 64, 103-111

Ureta, M. (1992), “The Importance of Lifetime Jobs in the U.S. Economy, Revisited ”, *The American Economic Review*, 82, 323-335

Zinde-Walsh, V. (2008), “Kernel estimation when density may not exist ”, *Econometric Theory*, 24, 696-725

Zinn, S. and A. Würbach (2016), “A statistical approach to address the problem of heaping in self-reported income data ”, *Journal of Applied Statistics*, 43, 682-703

Table 1: Actual rejection probabilities of KS tests: KS_n $(p = 0.2, 0.1, 0.05, 0)^{\dagger}$

	$p = 0.2$		$p = 0.1$		$p = 0.05$		$p = 0$	
$\alpha^{\ddagger}$	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$n = 500$	0.357	0.867	0.000	0.000	0.000	0.000	0.000	0.000
$n = 1000$	1.000	1.000	0.002	0.023	0.000	0.000	0.000	0.000
$n = 2000$	1.000	1.000	0.149	0.722	0.000	0.000	0.000	0.000

† Numbers in each cell indicate the actual rejection probabilities for the KS test in 1000 replications.

†† α is a nominal significance level.

Table 2: Actual rejection probabilities (RP) of PMP tests: T_L and T_S ($p = 0.2, 0.1, 0.05, 0$)[†]

	$T_L(0)$		$T_L(1)$		$T_L(0.5)$		$T_S(0)$		$T_S(1)$		$T_S(0.5)$	
α[††]	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
$p = 0.2$												
$n = 500$	0.997	1.000	1.000	1.000	0.005	0.053	1.000	1.000	1.000	1.000	0.000	0.000
$n = 1000$	1.000	1.000	1.000	1.000	0.009	0.041	1.000	1.000	1.000	1.000	0.000	0.000
$n = 2000$	1.000	1.000	1.000	1.000	0.009	0.040	1.000	1.000	1.000	1.000	0.000	0.000
$p = 0.1$												
$n = 500$	0.761	0.925	0.868	0.965	0.008	0.053	0.997	1.000	0.998	1.000	0.000	0.000
$n = 1000$	0.990	1.000	0.997	1.000	0.007	0.041	1.000	1.000	1.000	1.000	0.001	0.001
$n = 2000$	1.000	1.000	1.000	1.000	0.007	0.050	1.000	1.000	1.000	1.000	0.000	0.000
$p = 0.05$												
$n = 500$	0.242	0.515	0.367	0.653	0.006	0.053	0.698	0.934	0.859	0.979	0.000	0.000
$n = 1000$	0.610	0.827	0.778	0.910	0.008	0.044	1.000	1.000	1.000	1.000	0.001	0.001
$n = 2000$	0.962	0.993	0.987	0.999	0.008	0.045	1.000	1.000	1.000	1.000	0.000	0.000
$p = 0$												
$n = 500$	0.009	0.048	0.009	0.055	0.008	0.052	0.000	0.000	0.000	0.000	0.001	0.001
$n = 1000$	0.009	0.040	0.010	0.048	0.007	0.042	0.001	0.001	0.001	0.001	0.001	0.001
$n = 2000$	0.008	0.051	0.005	0.046	0.008	0.052	0.000	0.000	0.000	0.000	0.000	0.000

[†] Numbers in each cell indicate the actual rejection probabilities for the PMP test T_L and T_S in 1000 replications. A uniform kernel $K(u) = I(|u| \leq 1/2)$ and $h_1 = 0.5n^{-2/5}$ and $h_2 = n^{-2/5}$ are used for T_L. A uniform kernel $K(u) = I(|u| \leq 1/2)$ and $h_1 = n^{-1}$ and $h_2 = n^{-2/5}$ are used for T_S.

[††] α is a nominal significance level.

Table 3: Actual RP's of bootstrap PMP tests: $T_{L,B}$ †

	$T_{L,B}(0)$		$T_{L,B}(1)$		$T_{L,B}(0.5)$	
$\alpha^{\ddagger}$	0.01	0.05	0.01	0.05	0.01	0.05
			$p = 0.2$			
$n = 500$	0.997	1.000	1.000	1.000	0.005	0.054
$n = 1000$	1.000	1.000	1.000	1.000	0.012	0.041
$n = 2000$	1.000	1.000	1.000	1.000	0.008	0.041
			$p = 0.1$			
$n = 500$	0.758	0.924	0.888	0.967	0.007	0.053
$n = 1000$	0.989	1.000	0.998	1.000	0.007	0.041
$n = 2000$	1.000	1.000	1.000	1.000	0.006	0.049
			$p = 0.05$			
$n = 500$	0.246	0.518	0.379	0.656	0.007	0.055
$n = 1000$	0.612	0.827	0.780	0.911	0.009	0.047
$n = 2000$	0.963	0.994	0.987	0.999	0.008	0.048
			$p = 0$			
$n = 500$	0.010	0.048	0.009	0.057	0.009	0.052
$n = 1000$	0.009	0.044	0.009	0.048	0.008	0.048
$n = 2000$	0.010	0.052	0.004	0.043	0.010	0.055

† Numbers in each cell indicate the actual rejection probabilities for the bootstrapped PMP test $T_{L,B}$ in 1000 replications and 1000 bootstrap resamples. A uniform kernel $K(u) = I(|u| \leq 1/2)$ and $h_1 = 0.5n^{-2/5}$ and $h_2 = n^{-2/5}$ are used.

†† α is a nominal significance level.

Table 4: Actual RP's of T_L and T_S in neighborhood of a mass point $\{0\}$

	$T_L(0.00001)$	$T_L(0.0001)$	$T_L(0.001)$	$T_L(0.01)$	$T_L(0.1)$	h_1	h_2
			$p = 0.2$				
$n = 500$	1.000	1.000	1.000	0.999	0.052	0.042	0.083
$n = 1000$	1.000	1.000	1.000	1.000	0.045	0.032	0.063
$n = 2000$	1.000	1.000	1.000	1.000	0.046	0.024	0.048
			$p = 0.1$				
$n = 500$	0.925	0.923	0.924	0.926	0.053	0.042	0.083
$n = 1000$	1.000	1.000	0.999	1.000	0.041	0.032	0.063
$n = 2000$	1.000	1.000	1.000	1.000	0.050	0.024	0.048
			$p = 0.05$				
$n = 500$	0.516	0.518	0.520	0.500	0.053	0.042	0.083
$n = 1000$	0.827	0.828	0.836	0.838	0.044	0.032	0.063
$n = 2000$	0.993	0.994	0.992	0.988	0.045	0.024	0.048

	$T_S(0.00001)$	$T_S(0.0001)$	$T_S(0.001)$	$T_S(0.01)$	$T_S(0.1)$	h_1	h_2
			$p = 0.2$				
$n = 500$	1.000	1.000	1.000	0.000	0.000	0.002	0.289
$n = 1000$	1.000	1.000	0.000	0.000	0.000	0.001	0.251
$n = 2000$	1.000	1.000	0.000	0.000	0.000	0.0005	0.219
			$p = 0.1$				
$n = 500$	0.997	0.997	0.997	0.000	0.000	0.002	0.289
$n = 1000$	1.000	1.000	0.000	0.000	0.000	0.001	0.251
$n = 2000$	1.000	1.000	0.000	0.000	0.000	0.0005	0.219
			$p = 0.05$				
$n = 500$	0.698	0.699	0.695	0.000	0.000	0.002	0.289
$n = 1000$	1.000	1.000	0.000	0.000	0.000	0.001	0.251
$n = 2000$	1.000	1.000	0.000	0.000	0.000	0.0005	0.219

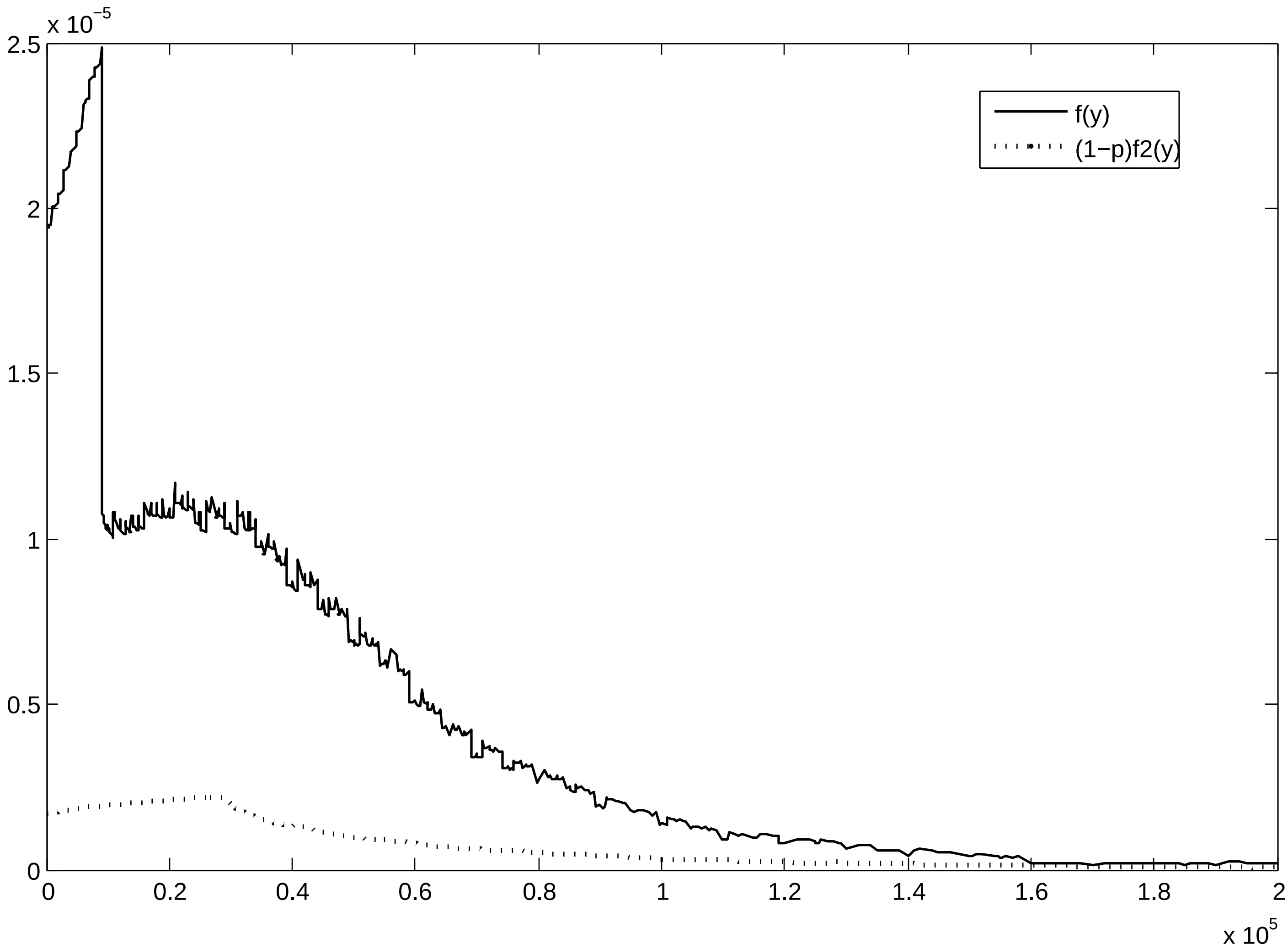
† Numbers in each cell indicate the actual rejection probabilities for the PMP test T_L and T_S in 1000 replications. For T_L, a uniform kernel, $h_1 = 0.5n^{-2/5}$, and $h_2 = n^{-2/5}$ are employed. For T_S, a uniform kernel, $h_1 = n^{-1}$, and $h_2 = n^{-2/5}$ are employed. The nominal size α is 0.05.

Table 5: Performance of the detection procedure†

	T_L			T_S		
	CT	FR	TR	CT	FR	TR
		$p = 0.2$				
$n = 500$	0.000	1.000	0.324	0.233	0.767	1.000
$n = 1000$	0.000	1.000	0.461	0.374	0.626	1.000
$n = 2000$	0.000	1.000	0.503	0.468	0.532	1.000
		$p = 0.1$				
$n = 500$	0.000	1.000	0.167	0.208	0.792	1.000
$n = 1000$	0.000	1.000	0.313	0.372	0.628	1.000
$n = 2000$	0.000	1.000	0.494	0.433	0.567	1.000
		$p = 0.05$				
$n = 500$	0.000	1.000	0.017	0.156	0.807	0.803
$n = 1000$	0.000	1.000	0.150	0.332	0.332	0.995
$n = 2000$	0.000	1.000	0.380	0.411	0.589	1.000
		$p = 0$				
$n = 500$	0.000	1.000	-	0.648	0.352	-
$n = 1000$	0.000	1.000	-	0.815	0.185	-
$n = 2000$	0.000	1.000	-	0.949	0.051	-

† Numbers in each cell indicate the fractions of correct detection (CT), false detection of non-mass points (FR), and correct detection of all mass points (TR) in 1000 replications. In a procedure using T_L, a uniform kernel, $h_1 = 0.5n^{-2/5}$, and $h_2 = n^{-2/5}$ are employed. In a procedure using T_S, a uniform kernel, $h_1 = n^{-1}$, and $h_2 = n^{-2/5}$ are employed. The symbol $-$ indicates that TR is not available when $p = 0$.

Figure 1: Kernel density estimates



These curves are the kernel density estimates for 8727 individual's wage in PSID 2013 wave. The solid line is the profile of kernel density estimates using all observations. The dotted line presents the profile of kernel density estimates obtained by removing detected mass points and rescaling.