The Korean Economic Review
Volume 35, Number 1, Winter 2019, 205-248.

Tests for Detecting Probability Mass Points*

Byung-hill Jun** · Hosin Song***

The objective of this paper is developing test statistics to detect the presence of mass points when data are possibly generated by a mixture of a continuous and a discrete distribution. To serve our purpose we propose two versions of the probability mass point (PMP) test. We derive the limiting distributions of the PMP test statistics under the null and alternative hypothesis by exploiting the asymptotic difference between two kernel density estimators with different bandwidths. Specifically, the proposed PMP test statistic is shown to converge to either the standard normal distribution or a linear transformation of a positive Poisson distribution at a non-mass point depending on bandwidths choice, while it diverges to infinity at a mass point. Numerical experiments are conducted to demonstrate the validity of our proposed tests. Korean taxpayers' bunching behavior is considered as an empirical application.

JEL Classification: C14, C18, C46, C52
Keywords: Bunching, Kernel Density Estimator, Probability Mass Point Test

I. Introduction

Empirical researchers are often confronted with variables massed at certain values. Examples of such variables include subjective probability (Bruine de Bruin et al., 2000, 2002), firm's earnings (Burgstahler and Dichev, 1997), income (Saez, 2010;

Received: Feb. 12, 2018. Revised: Aug. 8, 2018. Accepted: Oct. 26, 2018.

* We are grateful for the helpful comments from two anonymous referees and seminar participants at the Asian Meeting of the Econometric Society, WEAI 89th Annual Conference, the 2016 China Meeting of the Econometric Society, the 2017 Autumn Meeting of JAAE, and the 2018 Annual Meeting of the Korean Econometric Society. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A5A8018660) and by the Hankuk University of Foreign Studies Research Fund.

** First Author, Associate Professor, Department of International Economics and Law, Hankuk University of Foreign Studies, 107 Imun-ro, Dongdaemun-gu, Seoul, 02450, Korea. E-mail: bjun@hufs.ac.kr.

*** Corresponding Author, Associate Professor, Department of Economics, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul, 03760, Korea. E-mail:hsong@ewha.ac.kr.

Chetty et al., 2011), job tenure (Ureta, 1992), household expenditure (Pudney, 2008), working hours (Otterbach and Sousa-Poza, 2010), age at retirement (Burtless and Moffitt, 1984), and neonatal mortality (Arulampalam et al., 2017). The recognition of the presence of mass points promotes the use of bunching estimation among economists. The bunching estimation exploits the fact that the magnitude of mass points reveals useful information in identifying the structural parameters. For example, Saez (2010) showed that in a static labor supply model without any frictions, the size of massed observations at a kink point is determined by the elasticity of earnings and the magnitude in the change in the marginal tax rate. Saez (2010) applied the idea to the administrative tax return data in estimating the elasticity of earnings. Many empirical works, especially in the literature of tax policies, adopted the approach in Saez (2010), and thus attempted to measure the bunching at a given location in estimating a structural parameter (e.g., Chetty et al., 2011; Bastani and Selin, 2014; Kleven et al., 2014; Almunia and Lopez-Rodriguez, 2018).

There are several arguments trying to explain these mass points (e.g., personal characteristics associated with rounding or recollection errors (Budd and Guinnane, 1991; Bruine de Bruin et al., 2000, 2002), the nature of a survey question (Holbrook et al., 2014), and the optimization behavior associated with an incentive (Burgstahler and Dichev, 1997; Saez, 2010; Chetty et al., 2011). Regardless of the disagreement about the sources generating the mass points, it has been widely acknowledged that neglecting the presence of mass points leads to the failure of conventional statistical inference in many situations (Heitjan and Rubin, 1991). Many studies addressed the issues arising from the prevalence of mass points in detail (Petoussi et al., 1997; Pudney, 2008; Bar and Lillard, 2012; Crawford et al., 2015; Barreca et al., 2016; Groβ and Rendtel, 2016; Zinn and Würbach, 2016; Arulampalam et al., 2017). They showed that some popular estimates including the maximum likelihood estimate, the dynamic generalized method of moment estimate, the regression-discontinuity estimate, and the kernel density estimate are likely to suffer from biases unless the stochastic process generating mass points is well dealt with in estimating the models.

Our work shares the concern related to the mass points in 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 the 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) reported that some choices by individuals or firms are likely to cluster around some specific points because of their economic incentives. 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.

The purpose of this study is to propose a test to detect the probability mass points when they are present among non-mass observations. Specifically, we attempt to develop a test for a hypothesis that there exists a probability mass at a given location. We presume that mass points are generated by a mixture of a continuous and a discrete distribution. Hence, a parameter capturing the probability of a mass point should be introduced when implementing a test for a single point. Under the null hypothesis of no probability mass at a given value, this probability is zero, and thus it is located on the boundary of the parameter space. As a result, standard regularity conditions for the log-likelihood (LR) test do not hold as in earlier works testing a mixture distributional assumption (Chernoff and Lander, 1995; Gassiat and Keribin, 2000; Cho and White, 2007, 2010; Cho and Han, 2009). In such a situation, the LR test statistic does not follow the standard asymptotic distributional results. One advantage of our proposed PMP test does not have to deal with such potential technical difficulty. The PMP test is easy to conduct since it can be built on two kernel density estimators and exhibits simple limiting distributional properties. Specifically, this test exploits the asymptotic difference between two kernel density estimators with different bandwidths. 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 with Takeuchi (2004), our PMP tests are more general since the PMP tests incorporate a general kernel density estimator, while Takeuchi (2004) considers histograms with a fixed bin.

The paper is organized as follows. In Section II, we introduce a mixture model to represent the situation where mass points are present among non-mass points. Assumptions for the PMP tests are also stated in Section II. In Section III, we propose the PMP tests and show their asymptotic properties. Section IV discusses the local power properties of the PMP tests. In Section V, we conduct some numerical experiments to evaluate the performance of our proposed tests. In Section VI, we apply the PMP tests to the Korean wage earners' expenditure data to investigate the presence of a bunching behavior. Section VII has concluding remarks. Proofs and other supplementary explanations are presented in the Appendix.

II. Assumptions

Let $Y_1, \cdots, Y_n$ 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 $0 \le p \le 1$*,* $F_1(y) = \sum_{j=1}^{J} f_1(d_j)I(y \ge d_j)$*,* $f_1(y)$ *is a probability mass function (pmf) with support* $D = \{d_1,\ldots,d_J\}$*,* $J < \infty$*, 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* $D \subset \mathrm{int}(C)$ *where* $\mathrm{int}(C)$ *is the interior of* C*.*

Assumption 1 models the presence of some mass points among continuous observations. Restrictions for a continuous pdf $f_2(y)$ in Assumption 1 are standard in many econometric studies. These restrictions are used to derive the asymptotic properties of the PMP test statistics in Section 3. The mass points are assumed to be present on the interior of C. This fact complicates the identification of the set of mass points D.

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

Example 1. Bruine 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 $C = \{y \in \mathbb{R} : 0 \le y \le 1\}$.

Example 2. Zinde-Walsh (2010) discusses a bunching in earnings when anticipated lump-sum transfers are made to workers with earnings less than a threshold d_1. In this example, Y is after-transfer earnings, $D = \{d_1\}$, and $C = \{y \in \mathbb{R} : y \ge 0\}$.

Example 3. 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 $C = \{y \in \mathbb{R} : y \ge 0\}$.

Example 4. Zinn and Würbach (2016) consider the digit preference patterns when individuals report their incomes. Specifically, individuals are assumed to tend to report multiples of 100, 500, and 1000 instead of their true incomes. Y is the reported income, $D = \{d = kd_0 : d_0 = 100, 500, 1000; k = 0,1,2,\ldots\}$ and $C = \{y \in \mathbb{R} : y \ge 0\}$. Similar distributions are employed in describing the reported current job starting year (Ureta, 1992) and the reported days between an infant's birth and death (Arulampalam et al., 2017).

Although the presence of mass points has been mostly regarded as a statistical problem, some economic research investigates the effects of mass points on conventional estimates. Via Monte Carlo simulations, Torelli and Trivellato (1993) and Wolff and Augustin (2003) show that the maximum likelihood estimation is likely to generate biased estimates in duration models if mass points generating process is not take into account in constructing a likelihood function. In a panel analysis of household's energy consumption, Pudney (2008) shows that the neglect of mass points leads to substantial biases in some popular dynamic panel estimators. A recent study by Barreca et al. (2016) reports that the regression-discontinuity estimator is inconsistent when mass points exist in the distribution of a running variable, which determines the treatment status.

Motivated by this line of studies, our primary interest lies in testing whether there is probability mass at a certain point. If the test results indicate the presence of probability mass points, then a researcher might want to remedy associated failures of conventional estimates or recover the underlying distribution of a true variable depending on his or her own interest.[1]

Throughout the paper, we denote a kernel density estimator at y_0 with a bandwidth h_1 by $\hat{f}(y_0;h_1)$. To define the PMP tests, let us state the assumptions for a kernel $K(\cdot)$ and a bandwidth h.

Assumption 2. *The kernel* $K(\cdot)$ *satisfies the following conditions:*
(i) $\int K(u)du = 1$, $\int uK(u)du = 0$, *and* $\int u^2 K(u)du < \infty$,
(ii) $K(\cdot)$ *is uniformly bounded,*
(iii) $\lim_{|u|\to\infty} |u|K(u) = 0$,
(iv) $K(\cdot)$ *is symmetric,*
(v) $\int |K(s)|^{2+\delta} ds < \infty$ *for some* $\delta > 0$,
(vi) $K(\cdot)$ *is Lipschitz-continuous. In other words, there exists a finite constant* d^* *satisfying*

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

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

Assumptions 2 and 3 are standard in the literature of kernel density estimation.

[1] For example, as mentioned in Barreca et al. (2016) the mass points can be regarded as the results of measurement errors (e.g., the digit preference, recollection errors, and rounding errors). In such a case, the distribution of a true variable with no measurement error can be an object of interest.

The Lipschitz-continuity condition in Assumption 2(vi) is required to evaluate the variance of $\hat{f}(y_0;h_1)-\hat{f}(y_0;h_2)$. As mentioned by Newey and West (1994), a wide class of kernels satisfy the Lipschitz-continuity condition so that it is not so restrictive.

A standard kernel density estimator $\hat{f}(y_0;h)$ defined as $\hat{f}(y_0;h)=\frac{1}{nh}\sum_{i=1}^{n}K(\frac{y_0-Y_i}{h})$ is consistent when $p=0$. 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.

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

and

$$\begin{aligned}V[\hat{f}(y_0;h)]=&\ p(y_0)\left(\frac{1}{nh^2}(1-p(y_0))K^2(0)+o\left(\frac{1}{nh^2}\right)\right)\\&+(1-p)\left(\frac{1}{nh}\kappa_V+o\left(\frac{1}{nh}\right)\right),\end{aligned} \tag{3}$$

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 $f_2^{(2)}$ is the second derivative of $f_2(\cdot)$.

Recognize that when $p=0$, equations (2) and (3) are reduced to the well-known results for the mean and variance of a kernel density estimator in literature. However, when $p>0$, a kernel estimator at a mass point y_0 has exploding terms in the asymptotic bias and variance. As a result,

$$\hat{f}(y_0;h)\xrightarrow{p}\infty \quad \text{for} \quad y_0\in D \tag{4}$$

and

$$\hat{f}(y_0;h)\xrightarrow{p}(1-p)f_2(y_0) \quad \text{for} \quad y_0\in C. \tag{5}$$

III. PMP Test

Let y_0 be the point where we want to test the presence of positive probability mass. The null and alternative hypotheses are stated as follows:

$$H_0 : p(y_0) = 0 \quad \text{vs.} \quad H_1 : p(y_0) > 0 .$$

Under the null hypothesis, the parameter $p(y_0)$ is located on the boundary. In this situation, the LR test statistic does not have a nice limiting distribution. Instead its limiting distribution is described by a functional of a Gaussian process (Andrews (2001)). The boundary parameter problem received much attention in earlier works testing the hypothesis of a mixture distribution (Chernoff and Lander, 1995; Gassiat and Keribin, 2000; Cho and White, 2007, 2010; Cho and Han, 2009). As in these works, one can use the LR test by establishing the limiting distribution of the LR test statistic in the given model.

To overcome the technical difficulty in deriving the asymptotic distribution results for the LR test statistic, we propose an alternative test by adopting the idea of Burgstahler and Dichev (1997) and Takeuchi (2004). 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 and present the PMP tests. The PMP tests use the asymptotic properties of the difference between $\hat{f}(y_0;h_1)$ and $\hat{f}(y_0;h_2)$, $h_1 \neq h_2$. More precisely, the PMP test statistic is given by

$$T = \Delta_n (\hat{f}(y_0;h_1) - \hat{f}(y_0;h_2)) ,$$

where Δ_n is a normalizing factor.

Let $h_1 \propto n^{-\alpha_1}$ and $h_2 \propto n^{-\alpha_2}$ where $\alpha_1, \alpha_2 > 0$. Without loss of generality, assume $h_1 < h_2$. A bandwidth h_2 satisfies Assumption 3 while a bandwidth h_1 is allowed to violate Assumption 3. As a result, $\hat{f}(y_0;h_1)$ is not necessarily consistent to $f_2(y_0)$. In the following parts of this section, we show that the distribution of the PMP test statistic T depends on whether or not $\hat{f}(y_0;h_1)$ is consistent. Exploiting these distributional results, we present two versions of the PMP tests T_L and T_S.

3.1. PMP Test with $\alpha_2 \leq \alpha_1 < 1$

In this subsection, we consider the case where $\alpha_1 < 1$ such that h_1 also satisfies Assumption 3 and thus, $\hat{f}(y_0;h_1)$ is a consistent estimator of $f_2(y_0)$. To characterize the relative magnitude of h_1 and h_2, define c as the limit of the ratio of two bandwidths, that is, $c = \lim h_1 / h_2$. Given that $h_1 < h_2$, there are two possible cases. The first case is that h_1 and h_2 have the same order of magnitude ($\alpha_1 = \alpha_2$) with $h_1 = ch_2$, $c \in (0,1)$. The other case happens when $\alpha_1 > \alpha_2$.[2] In that case, h_1 decays to 0 faster than h_2 so that $c = 0$.

Hereafter, suppress y_0 in $\hat{f}(y_0;h)$ so that $\widehat{f(h_1)}$ and $\widehat{f(h_2)}$ denote two estimators $\hat{f}(y_0;h_1)$ and $\hat{f}(y_0;h_2)$. The PMP test statistic T_L when $c > 0$ is formally stated as follows.

$$T_L = \frac{\sqrt{nh_1}(\widehat{f(h_1)} - \widehat{f(h_2)})}{\sqrt{\hat{V}_0}},$$

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

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

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

Theorem 1. *Suppose that Assumptions 1 and 2 are satisfied. In addition, bandwidths* h_1 *and* h_2 *satisfy Assumption 3 and* $nh_2^5 \to 0$ *as* $n \to \infty$. *Then, for any* y_0, $T_L \xrightarrow{d} N(0,1)$ *under* $H_0 : p(y_0) = 0$. *Under* $H_1 : p(y_0) > 0$, T_L *tends to infinity with probability 1.*

In addition to the conventional regularity conditions for a bandwidth, Theorem 1 requires $nh_2^5 \to 0$. Then it follows that $\frac{1}{5} < \alpha_2 \le \alpha_1 < 1$. Under H_0, the order of the asymptotic bias and variance are h_2^2 and $1/(nh_1)$, respectively. The restrictions on α_1 and α_2 guarantee that neglecting the asymptotic bias in T_L has no effect on the limiting distribution as long as the first-order is concerned. Thus, Theorem 1 can be modified as

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

A close look at the proof of Theorem 1 shows that the asymptotic normality of T_L under H_0 is related to the asymptotic distributional property of a kernel density estimator in a standard setting. This fact is more clearly revealed when $\alpha_1 > \alpha_2$. In such a case, the convergence rate of $\widehat{f(h_2)}$ is faster than that of $\widehat{f(h_1)}$

[2] We are grateful to an anonymous referee who suggested the extension of T_L to this general case.

under $H_0 : p(y_0) = 0$. A direct manipulation using the faster convergence rate of $\widehat{f(h_2)}$ and $c = 0$ provides

$$T_L = \frac{\sqrt{nh_1}(\widehat{f(h_1)} - f(y_0))}{\sqrt{f(y_0)\int K^2(s)ds}} + o_p(1) \text{ under } H_0 : p(y_0) = 0 .$$

Recognize that the first term in the right-hand side is the standardized kernel density estimator, whose asymptotic distribution is standard normal in a standard setting.

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 with center y_0 and radius $h/2$. More specifically, the test statistic of Burgstahler and Dichev (1997) and Takeuchi (2004) is given by

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

A direct calculation provides

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

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 also be shown that

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

and

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

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_L does not converge to the standard normal distribution without undersmoothing even under H_0. Thus, conventional critical values fail to generate the desired rejection probabilities under H_0.

The effect of c is not immediately revealed in the limiting distribution of T_L. However, a look at the proof of Theorem 1 hints at the channel through which the choice of c influences the size and power properties of T_L in a finite sample. For given n and bandwidths, the size property partly depends on the magnitude of the bias, which is proportional to $\frac{1}{\sqrt{k_c}}(1-1/c^2)$ under H_0. Similarly, when n and bandwidths are given, the power increases with the magnitude of the bias, which is proportional to $\frac{1}{\sqrt{k_c}}(1-c)$ under H_1.

3.2. PMP Test with $\alpha_2 < \alpha_1 = 1$

Theorem 1 requires that bandwidths h_1 and h_2 satisfy Assumption 3 so that the resulting kernel density estimators are consistent. Now h_1 is allowed to violate Assumption 3. Specifically, we consider a special case where one bandwidth h_1 is given by $h_1 = c_0 / n$ for a positive constant c_0 while the other bandwidth h_2 satisfies Assumption 3.

For this special case we use a uniform kernel $K_u(u) = I(|u| \le 1/2)$, where $I(\cdot)$ is the indicator function. Denote the uniform kernel density estimator by $\widehat{f_u(h_j)}$, $j = 1,2$.

$$\widehat{f_u(h_j)} = \frac{1}{nh_j}\sum_{i=1}^{n} K_u\left(\frac{y_0 - Y_i}{h_j}\right), \quad j = 1,2.$$

A uniform kernel $K_u(u)$ satisfies Assumption 2. Moreover, k_c is 1 so that $\hat{V}_{u,0} = k_c \widehat{f_u(h_2)}$ is $\widehat{f_u(h_2)}$.

In this case, only $\widehat{f_u(h_2)}$ is a consistent estimator when $p(y_0) = 0$. When $p(y_0) = 0$, on the empirical support an inconsistent estimator of $\widehat{f_u(h_1)}$ is shown to have the following limiting distribution:

$$\widehat{f_u(h_1)} \xrightarrow{d} \frac{1}{c_0} W^+(k^*) \quad \text{for} \quad k^* = c_0(1-p) f_2(y_0), \tag{6}$$

where $W^+(k^*)$ is a positive Poisson distribution with a parameter k^*[3]. Slutsky's theorem together with these results imply

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

$$\frac{\sqrt{nh_1}\,(\widehat{f(h_1)}-\widehat{f(h_2)})}{\sqrt{\hat{V}_0}} \xrightarrow{d} \frac{W^{+}(k^{*})-k^{*}}{\sqrt{k^{*}}} \tag{7}$$

under H_0. It can be also shown that under $H_1: p(y_0)>0$, the statistic in the left-hand side in Equation (7) tends to diverge to infinity with probability 1.

Despite these limiting properties, the statistic in the left-hand side in Equation (7) cannot be directly used as a test statistic since its limiting distribution under H_0 is not pivotal. To deal with this issue, we modify the statistic Equation (7) to propose another version of the PMP test statistic. Define

$$T_S = I(\widehat{f_u(h_2)} > d_0)\times\frac{\sqrt{nh_1}\,(\widehat{f_u(h_1)}-\widehat{f_u(h_2)})}{\sqrt{\hat{V}_{u,0}}},$$

where d_0 is a finite positive constant and $I(\cdot)$ is the indicator function. Also define $\overline{t}_\alpha$ as

$$\overline{t}_\alpha = \sup_{s\in[c_0 d_0,\infty)} t_\alpha(s),$$

where $t_\alpha(s)$, $s>0$, is the $(1-\alpha)$-th quantile of $(W^{+}(s)-s)/\sqrt{s}$. Theorem 2 states that the test rejecting $H_0: p(y_0)=0$ against $H_1: p(y_0)>0$ when $T_S>\overline{t}_\alpha$ has an asymptotic size at most α and it is asymptotically consistent.

Theorem 2. *Suppose that Assumption 1 is satisfied. Also assume that* $h_1 = c_0/n$ *for a positive constant* c_0 *and* h_2 *satisfies Assumption 3. Then*

$$\lim_{n\to\infty}\Pr(T_S>\overline{t}_\alpha \mid p(y_0)=0)\le\alpha \tag{8}$$

and

$$\lim_{n\to\infty}\Pr(T_S>\overline{t}_\alpha \mid p(y_0)>0)\to 1. \tag{9}$$

To implement this test, we need a constant d_0. To understand the roles of d_0 and the indicator function in T_S, set d_0 to be 0. Then the test statistic T_S is reduced to the statistic in Equation (6). Let q_{k^*} be the probability that $W^{+}(k^{*})$ is 1. That is, $q_{k^*}=\Pr(W^{+}(k^{*})=1)$. The $(1-\alpha)$-th quantile of $W^{+}(k^{*})$ is 1 for any $\alpha\ge 1-q_{k^*}$. As a result, we have

$$t_\alpha(k^*) = \frac{1-k^*}{\sqrt{k^*}} \quad \text{for} \quad \alpha \geq 1 - q_{k^*},$$

which increases to infinity as $k^* \to 0$. Therefore, $\overline{t_\alpha} = \sup_{k^* \in \mathcal{R}^+} t_\alpha(k^*) = \infty$ for all $\alpha \in (0,1)$ when $d_0 = 0$. To resolve this problem, we take the supremum of $t_\alpha(k^*)$ over the set of k^* greater than $c_0 d_0$. This modification generates a finite critical value, which guarantees the required rejection probability (8) for any $k^* \geq c_0 d_0$, that is, $f_2(y_0) \geq d_0/(1-p)$ under $H_0 : p(y_0) = 0$. For $k^* < c_0 d_0$, that is, $f_2(y_0) < d_0/(1-p)$, the null hypothesis is rejected with probability approaching 0, and thus (8) is also satisfied.

IV. Local Power of the PMP Test

4.1. PMP Test with $\alpha_2 \leq \alpha_1 < 1$

For the PMP test T_L, we consider a local alternative hypothesis $H_{1,\delta} : p(y_0) = \delta / \sqrt{n/h_1}$, $\delta > 0$. The probability of y_0 in this asymptotics converges to 0 at a rate of $n^{-0.5-\alpha_1/2}$. It follows from Equations (2) and (3) that for $j = 1,2$,

$$E[\widehat{f(h_j)}] = (1-p)\left(f(y_0) + \frac{h_j^2}{2} f_2^{(2)}(y_0) \int s^2 K(s) ds \right) + o(h_j^2) \tag{10}$$

$$V[\widehat{f(h_j)}] = \frac{1}{nh_j}(1-p) f_2(y_0) \int K^2(s) ds + o\left(\frac{1}{nh_j} \right) \tag{11}$$

under $H_{1,\delta}$. As a result, we have

$$\widehat{f(h_j)} \xrightarrow{p} (1-p) f_2(y_0), \quad j = 1,2. \tag{12}$$

The PMP test statistic T_L can be written as

$$\begin{aligned} T_L &= T_{L,1} + \Delta_{L,\delta,n} \\ &:= \frac{\sqrt{nh_1}(\widehat{f(h_1)} - \widehat{f(h_2)} - E[\widehat{f(h_1)} - \widehat{f(h_2)}])}{\sqrt{\hat{V}_0}} + \frac{\sqrt{nh_1}}{\sqrt{\hat{V}_0}}(E[\widehat{f(h_1)}] - E[\widehat{f(h_2)}]). \end{aligned}$$

The first term $T_{L,1}$ converges to a standard normal distribution under $H_{1,\delta}$.

Using Equations (2), (3), and (12), the second term is shown to converge to a constant $\Delta_{L,\delta}$ given by

$$\Delta_{L,\delta} = \delta(1-c)\sqrt{\frac{K(0)}{k_c(1-p)f_2(y_0)}}.$$

Recall that the PMP test with nominal size α rejects $H_1 : p(y_0) > 0$ when $T_L > z_\alpha$, $0 < \alpha < 1$ where z_α is the $(1-\alpha)$-th quantile of a standard normal distribution. Then the power of T_L against $H_{1,\delta}$ is

$$\begin{aligned}\Pr(T_L > z_\alpha) &= \Pr(T_{L,1} > z_\alpha - \Delta_{L,\delta,n}) \\ &\approx 1 - \Phi(z_\alpha) + \phi(z_\alpha)\Delta_{L,\delta},\end{aligned} \tag{13}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the distribution function and density function of a standard normal distribution, respectively. The second line is obtained by using the above-mentioned results: $Z_{1,L}$ converges to a standard normal distribution and $\Delta_{L,\delta,n} \xrightarrow{p} \Delta_{L,\delta}$ under $H_{1,\delta}$.

This result shows that the PMP test T_L has nontrivial power against the sequences of $p(y_0)$ shrinking to 0 at a rate of $n^{-0.5-\alpha_1/2}$. Recognize that the nontrivial power crucially hinges on the limiting behavior of $\Delta_{L,\delta,n}$. If $\Delta_{L,\delta,n}$ approaches 0 in the limit so that $p(y_0)$ converges to 0 at a rate faster than $n^{-0.5-\alpha_1/2}$, the limiting distribution of T_L under H_1 is the same as the one under H_0. Thus, the asymptotic power is simply α.

4.2. PMP Test with $\alpha_2 < \alpha_1 = 1$

Here we consider a local alternative hypothesis $H_{1,\delta} : p(y_0) = \delta / n$, $\delta > 0$. Recall that we consider a local alternative hypothesis $H_{1,\delta} : p(y_0) = \delta / n^{0.5+\alpha/2}$. Since α_1 for T_S is 1, this specific alternative hypothesis corresponds to the local alternative hypothesis for T_L.

Similar with Equation (6), under $H_{1,\delta}$

$$\widehat{f_u(h_1)} \xrightarrow{d} \frac{1}{c_0} W^+(k^{**}) \ \text{ for } \ k^{**} = \delta + c_0(1-p)f_2(y_0) \tag{14}$$

on the empirical support. This result together with Equation (12) implies

$$\frac{\sqrt{nh_1}\,(\widehat{f_u(h_1)} - \widehat{f_u(h_2)})}{\sqrt{\hat{V}_{u,0}}} \xrightarrow{d} \frac{W^+(k^{**}) - k^*}{\sqrt{k^*}}.$$

As long as d_0 is sufficiently small such that $(1-p)f_2(y_0) > d_0$, Equation (12) provides that $\widehat{f_u(h_2)} > d_0$ with probability approaching 1. The power of T_S against the local alternative $H_{1,\delta}$ is

$$\begin{aligned}\Pr(T_s > \overline{t}_\alpha) &\approx \Pr(W^+(k^{**}) > k^* + \sqrt{k^*}\overline{t}_\alpha) \\ &= \Pr\left(\frac{W^+\left(k^{**}\right) - k^{**}}{\sqrt{k^{**}}} > \sqrt{\frac{k^*}{k^{**}}}\overline{t}_\alpha - \frac{k^{**} - k^*}{\sqrt{k^{**}}}\right).\end{aligned} \tag{15}$$

Unlike the local power of T_S stated in Equation (13), the approximate rejection probability in Equation (15) is not simply obtained since the critical value $\overline{t}_\alpha$ is selected to guarantee the asymptotic conservativeness of the PMP test T_S in the presence of a nuisance parameter k^*. Therefore, instead of deriving a more specific expression of the power in Equation (15), let us compare it with the rejection probability under H_0. From the definitions of k^* and k^{**}, $k^{**} - k^* = \delta > 0$. As a result,

$$\sqrt{\frac{k^*}{k^{**}}}\overline{t}_\alpha - \frac{k^{**} - k^*}{\sqrt{k^{**}}} < \overline{t}_\alpha .$$

Thus, we have for any δ and k^*,

$$\Pr(T_s > \overline{t}_\alpha \mid H_{1,\delta}, k^*) > \Pr(T_s > \overline{t}_\alpha \mid H_0, k^*),$$

which holds with probability approaching 1. With the last inequality, it can be asserted that the PMP test T_S has nontrivail power against the local alternatives with $p(y_0)$ shrinking to 0 at a rate of n^{-1} and that the power increases with δ.

V. Simulation

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

[Table 1] Actual rejection probabilities of T_L and T_S (p =0.2,0.1,0.05,0)

	T_L						T_S					
y_0	0	0	1	1	0.5	0.5	0	0	1	1	0.5	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	0.003	0.027
n=1000	1.000	1.000	1.000	1.000	0.009	0.041	1.000	1.000	1.000	1.000	0.002	0.024
n=2000	1.000	1.000	1.000	1.000	0.009	0.040	1.000	1.000	1.000	1.000	0.000	0.015
						$p=0.1$						
n=500	0.761	0.925	0.886	0.965	0.008	0.054	1.000	1.000	1.000	1.000	0.000	0.005
n=1000	0.990	1.000	0.997	1.000	0.007	0.042	1.000	1.000	1.000	1.000	0.000	0.001
n=2000	1.000	1.000	1.000	1.000	0.007	0.045	1.000	1.000	1.000	1.000	0.003	0.007
						$p=0.05$						
n=500	0.242	0.515	0.367	0.653	0.006	0.052	0.980	1.000	0.990	1.000	0.000	0.005
n=1000	0.610	0.827	0.778	0.910	0.008	0.045	1.000	1.000	1.000	1.000	0.002	0.010
n=2000	0.962	0.993	0.987	0.999	0.008	0.054	1.000	1.000	1.000	1.000	0.001	0.005
						$p=0$						
n=500	0.009	0.048	0.009	0.055	0.008	0.053	0.000	0.005	0.000	0.010	0.000	0.005
n=1000	0.009	0.040	0.010	0.048	0.007	0.041	0.002	0.008	0.001	0.021	0.001	0.009
n=2000	0.008	0.051	0.005	0.046	0.008	0.048	0.001	0.008	0.001	0.023	0.000	0.004

Note: Numbers in each cell indicate the actual rejection probabilities for the PMP tests T_L and T_S in 1000 replications. A uniform kernel $K(u)=I(|u|\le 1/2)$, $h_1=0.5n^{-0.4}$, and $h_2=n^{-0.4}$ are used for T_L. $h_1=n^{-1}$, $h_2=n^{-0.4}$, and $d_0=0.3$ are used for T_S.

† α is a nominal significance level.

Table 1 displays the actual rejection probabilities of two versions of PMP tests at mass points 0 and 1, and a non-mass point 0.5.[4] For T_L, we choose $h_1=0.5n^{-0.4}$ and $h_2=n^{-0.4}$. For T_S, we set $h_1=1/n$, $h_2=n^{-0.2}$, and $d_0=0.3$.[5] As expected, the rejection probabilities of T_L and T_S at mass points turn out to increase with p. In both PMP tests, when p is sufficiently large ($p=0.2$), the rejection probabilities at mass points are close to 1 for all n. However, two PMP tests have noticeably different rejection probabilities at mass points when p is small ($p=0.05$, 0.1). The powers of T_L display an increasing pattern with p while the rejection probabilities of T_S still remain close to 1. The rejection probabilities at mass points also increase with the sample size, and thus have values close to 1 for all positive values of p when the sample size is sufficiently large ($n=2000$). In all, both PMP tests show similar power properties although T_S has greater power than T_L. In contrast to the power properties, different features are observed in the empirical sizes of the two PMP tests. The empirical rejection probabilities of T_L at

[4] Because of the symmetry of the distribution, the rejection probabilities show similar patterns at -1 and 1. Thus, we do not report the results for another mass point -1.

[5] The critical values $\bar{t}_\alpha$ at the 1% and 5% significance levels are 4.93 and 3.13, respectively.

non-mass points (either $p=0$ or $y_0=0.5$ with $p>0$) turn out to be close to nominal sizes while those of T_S remain much smaller than the nominal sizes for all sample sizes n.

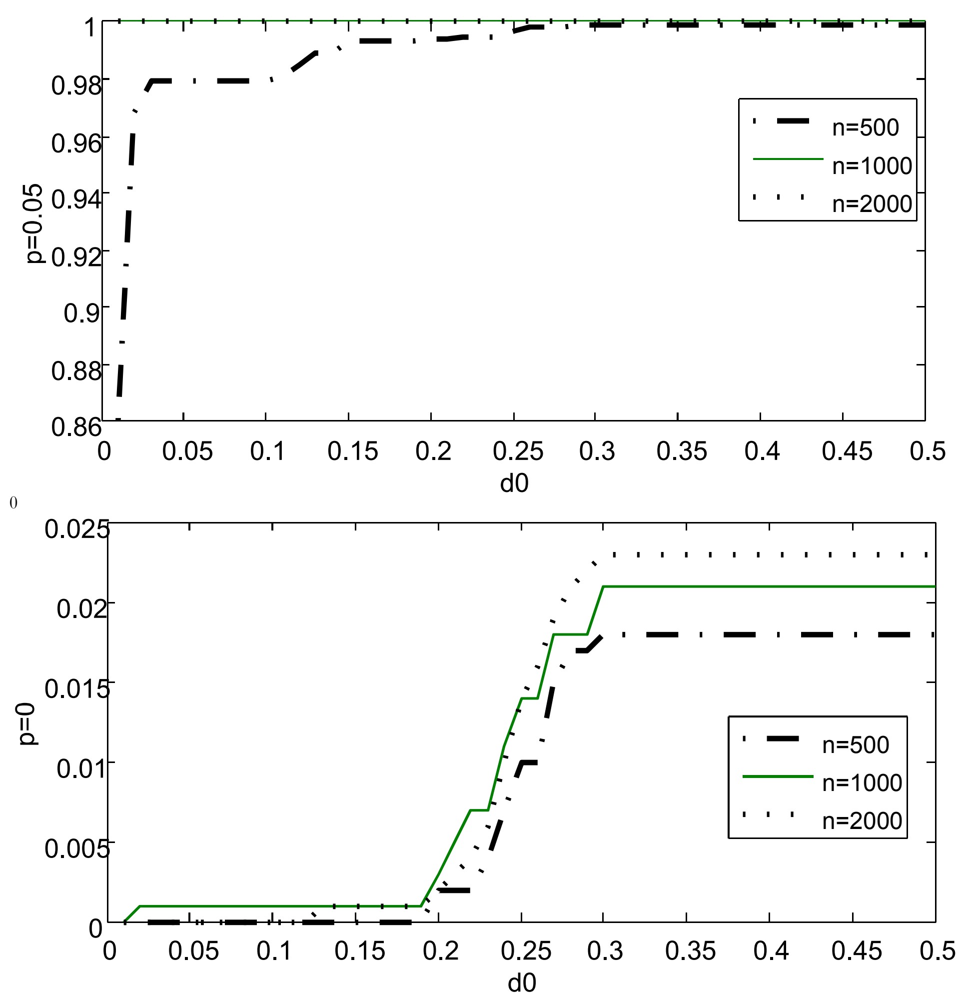
[Table 2] Actual RP's of T_L and T_S in the neighborhood of a mass point 0 ($\alpha=0.05$)

	$y_0=0.0001$	$y_0=0.001$	$y_0=0.01$	$y_0=0.1$	h_1	h_2
Panel I. Rejection Probabilities of T_L						
$p=0.2$						
$n=500$	1.000	1.000	0.999	0.052	0.042	0.083
$n=1000$	1.000	1.000	1.000	0.045	0.032	0.063
$n=2000$	1.000	1.000	1.000	0.046	0.024	0.048
$p=0.1$						
$n=500$	0.923	0.924	0.926	0.053	0.042	0.083
$n=1000$	1.000	0.999	1.000	0.041	0.032	0.063
$n=2000$	1.000	1.000	1.000	0.050	0.024	0.048
$p=0.05$						
$n=500$	0.518	0.520	0.500	0.053	0.042	0.083
$n=1000$	0.828	0.836	0.838	0.044	0.032	0.063
$n=2000$	0.994	0.992	0.988	0.045	0.024	0.048
Panel II. Rejection Probabilities of T_S						
	$y_0=0.0001$	$y_0=0.001$	$y_0=0.01$	$y_0=0.1$	h_1	h_2
$p=0.2$						
$n=500$	1.000	1.000	0.010	0.005	0.002	0.289
$n=1000$	1.000	0.008	0.009	0.005	0.001	0.251
$n=2000$	1.000	0.000	0.000	0.000	0.0005	0.219
$p=0.1$						
$n=500$	1.000	1.000	0.010	0.005	0.002	0.289
$n=1000$	1.000	0.000	0.000	0.000	0.001	0.251
$n=2000$	1.000	0.006	0.007	0.006	0.0005	0.219
$p=0.05$						
$n=500$	1.000	1.000	0.010	0.005	0.002	0.289
$n=1000$	1.000	0.008	0.009	0.005	0.001	0.251
$n=2000$	1.000	0.006	0.012	0.007	0.0005	0.219

Note: Numbers in each cell indicate the actual rejection probabilities for both PMP tests in 1000 replications. For T_L, a uniform kernel, $h_1=0.5n^{-0.4}$, and $h_2=n^{-0.4}$ are employed. For T_S, $h_1=n^{-1}$, $h_2=n^{-0.4}$, and $d_0=0.3$ are employed.

Table 2 presents the size properties of T_L and T_S at non-mass points neighboring to a mass point 0. Particularly, we consider four non-mass points $0.0001, 0.001, 0.01$, and 0.1. The three different non-zero values are used for p ($p=0.05, 0.1, 0.2$). When a non-mass point is sufficiently away from 0, T_S has at least as small false rejection 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 delivers relatively a small number of false detections while T_L generates false detection with a rate close to 1. This finding is explained by the magnitudes of the selected bandwidths relative to the distance from a mass point. As a non-mass point becomes closer to 0, both tests T_S and T_L make more false rejections, and thus suffer from large size distortion. At first glance, a larger sample size seems to aggravate the size distortion problem as shown in the results for T_L at 0.0001. However, this observation misleads the true effect of the sample size n. In fact, the size distortion disappears eventually as the sample size n increases so that the bandwidths h_1 and h_2 become sufficiently smaller than the distance between a

[Figure 1] Illustration of power and size at $y_0 = 1\,(\alpha = 0.05)$



Note: The top panel displays the rejection probability of $H_0 : p(y_0) = 0$ when $p = 0.05$ and $y_0 = 1$. The bottom panel displays the rejection probability of $H_0 : p(y_0) = 0$ when $p = 0$ and $y_0 = 1$.

non-mass point and 0.[6] This fact explains the better performance of T_S at points 0.001, 0.01 and 0.1.

Recall that the critical value $\bar{t}_\alpha$ is dependent on the choice of d_0. In order to have a better understanding of the effect of d_0, we present the rejection probabilities of T_S depending on the change of d_0 in Figure 1. Specifically, the rejection probabilities of T_S at a point $y_0 = 1$ are investigated for two cases: $p = 0.05$ and $p = 0$. Bandwidths h_1 and h_2 are selected as described in Table 1. The upper and lower panels of Figure 1 summarize the profiles of rejection probabilities when $p = 0.05$ and $p = 0$, respectively. Recognize that $y_0 = 1$ is a mass (non-mass, *resp.*) point when $p > 0$ ($p = 0$, *resp.*) so that the upper (lower, *resp.*) panel presents the power (size, *resp.*) properties. Results in the upper panel hint that the power stays close to 1, and thus is not quite sensitive to the choice of d_0 except when sample size is small ($n = 500$). In contrast, the size in the lower panel turns out to be sensitive to the choice of d_0. However, the size does not show much variation when d_0 is sufficiently large. For example, all size curves remain flat for all $d_0 > 0.3$. These findings imply that size and power properties of the PMP test T_S displayed in Tables 1 and 2 are not much varied even when different values of d_0 are used.

[Table 3] Actual rejection probabilities of T_L and T_S under local alternative

	T_L				T_S			
y_0	0	0	1	1	0	0	1	1
α †	0.01	0.05	0.01	0.05	0.01	0.05	0.01	0.05
	$p(y_0) = \sqrt{h_1 / n}$				$p(y_0) = 1/n$			
n=500	0.166	0.392	0.236	0.522	0.039	0.152	0.120	0.326
n=1000	0.173	0.410	0.227	0.517	0.061	0.180	0.121	0.340
n=2000	0.184	0.412	0.246	0.516	0.050	0.169	0.142	0.363

Note: Numbers in each cell indicate the actual rejection probabilities for the PMP test T_L and T_S in 1000 replications. A uniform kernel $K(u) = I(|u| \le 1/2)$, $h_1 = 0.5n^{-0.4}$, and $h_2 = n^{-0.4}$ are used for T_L. $h_1 = n^{-1}$, $h_2 = n^{-0.4}$, $d_0 = 0.3$, and a uniform kernel are used for T_S.

† α is a nominal significance level.

[6] The size problem at a non-mass point y_0 in the neighborhood of a mass point y^* results from the short distance between y_0 and y^*. Recall that in our example, a mass point y^* affects the uniform kernel density estimate at y_0 when $h > 2|y^* - y_0|$. A fall in h associated with an increase in sample size n eventually makes the distance between y^* and y_0 greater than $2h$ so that the effect of the mass point y^* disappears and the size problem improves.

[Table 4] Actual rejection probabilities of the modified HCP test

	$p=0.2$	$p=0.1$	$p=0.05$	$p=0$	$p=0.2$	$p=0.1$	$p=0.05$	$p=0$
				$y_0=0$ ($\alpha=0.05$)				
		$h=h_{rt}$				$h=n^{-0.4}$		
n=500	0.768	0.454	0.241	0.021	0.758	0.441	0.229	0.019
n=1000	0.923	0.624	0.346	0.030	0.920	0.614	0.326	0.033
n=2000	0.981	0.769	0.473	0.036	0.981	0.759	0.464	0.038
				$y_0=1$ ($\alpha=0.05$)				
		$h=h_{rt}$				$h=n^{-0.4}$		
n=500	0.905	0.595	0.354	0.033	0.901	0.582	0.343	0.028
n=1000	0.983	0.767	0.467	0.210	0.978	0.761	0.453	0.020
n=2000	1.000	0.910	0.637	0.031	1.000	0.905	0.623	0.027

Note: Numbers in each cell indicate the actual rejection probabilities for both HCP tests in 1000 replications. For $\widehat{f_u}(0)$, a uniform kernel and two bandwidths ($h_{rt}=1.06\widehat{\sigma_{\hat{u}}}n^{-0.2}$ and $h=n^{-0.4}$) are used.

We conduct numerical experiments to see the local power properties of the PMP tests discussed in section IV. Table 3 shows the empirical rejection probabilities of PMP tests against a sequence of local alternative hypotheses. We maintain the same simulation setup except for $p(y_0)$. $p(y_0)$ is assumed to converge to 0 at a rate of $\sqrt{h_1/n}$.[7] As predicted, the simulation results suggest that both PMP tests have nontrivial power against the given local alternative hypotheses. Moreover, the rejection probabilities are reported to remain stable despite the increase in the sample size n, which is also expected from the result in Section IV.

Han, Cho and Phillips (2011), hereafter HCP (2011), proposed a test for detecting infinite density at the median.[8] Considering the similarity of the hypotheses in HCP (2011) and the PMP tests, we also conduct numerical simulations to compare the performance. Some modifications are made in implementing the HCP test. First, there is no specific quantile regression that we should explicitly consider. Thus, we independently generate one variable X and estimate a quantile regression showing the relationship between the variable of interest Y and a newly generated variable X. By design, the true slope parameter is 0. This knowledge simplifies the construction of the HCP test statistic by enabling us not to rely on a sample-splitting technique used in HCP (2011). Second, the location that we want to test the presence of probability masses is not necessarily the median, while the original version of the HCP test investigates the infinite density at the median. So we generalize the HCP test for any q-th quantile

[7] Specifically, we set $p(y_0)=\sqrt{2}n^{-0.7}$ for T_L and $p(y_0)=n^{-1}$ for T_S by using the bandwidths as before.

[8] We are grateful to an anonymous referee who suggested this idea.

regression, $0 < q < 1$ when using an estimate of q. For details of the simulation, see Appendix B. Table 4 shows the performance of the HCP test at $y_0 = 0$ and 1 with different combinations of p and n. We used the rule-of-thumb bandwidth, $h_{rt} = 1.06\hat{\sigma}n^{-0.2}$, where $\hat{\sigma}$ is the sample standard deviation. We also use another bandwidth $h = n^{-0.4}$ to the purpose of comparing the simulation results with the results in Table 1.[9] The simulation results indicate that the HCP test has desirable size and power properties in the sense that empirical power increases to 1 with sample size and empirical size does not differ much from its nominal size. Comparing these results with the results in Table 1, we find that PMP test performs slightly better than HCP test in the current data generating process.

VI. Empirical application: Korean Wage Earners' Bunching Behavior

In this section we consider two expenditure distributions of Korean wage earners. For the purpose of income tax filing, Korean wage earners can claim tax deductions on certain expenditures. Selected but not complete list of eligible expenditures are medical expenses, educational expenses, private insurance expenses, charitable gifts, and credit card expenditures.

The amount of tax deduction D for each category is typically stipulated as follows.

$$D = \min(k \times \max(E - \underline{E}, 0), k \times (\overline{E} - \underline{E})),$$

where E is the amount of money spent on an eligible category, k is a deduction rate between 0 and 1, and $\underline{E}$ and $\overline{E}$ represent the minimum and maximum of the phase-in range, respectively. According to this rule, tax deduction on each category remains at 0 for $E < \underline{E}$, linearly increases for E, $\underline{E} \leq E < \overline{E}$, and remains at $k(\overline{E} - \underline{E})$ for $E \geq \overline{E}$.

Tax deduction lowers income taxes, and thus it is equivalent to subsidizing eligible expenses. As a result, a wage earner is faced with an effective cost of $1-$ (marginal tax rate $\times$ tax deduction for an extra unit of expenditure) when spending an additional unit of money on an eligible item. In this regard, a wage earner experiences an abrupt change in the effective cost around the endpoints in the phase-in area, $\underline{E}$ and $\overline{E}$. Therefore, excessive numbers of observations are expected to be found at $\underline{E}$ and $\overline{E}$ as in Saez (2010) and Chetty et al. (2011).[10]

---

[9] These two bandwidths satisfy the condition for the bandwidth in HCP (2011): $h \to 0$ and $\sqrt{nh} \to \infty$ as $n \to \infty$.

[10] Some taxpayers might fail for recognizing the exact tax deduction schedule since it is determined

[Table 5] Summary statistics of the credit card expenditure and taxpayers' demographic variables: 2007-2015

Year	mean	standard deviation	min	max
credit card expenditure	1,501.3	1,190.3	0.0	15,993.0
age	39.8	9.5	18	78
marriage dummy	0.80	0.40	0	1
male dummy	0.72	0.45	0	1

Note: These summary statistics are obtained from a sample of 12,589 labor income taxpayers in the NaSTaB survey reporting credit card expenditure between 2007 and 2015. The unit of credit card expenditure is 10,000 KRW. The marriage dummy is one if the taxpayer is married, and zero otherwise. The male dummy is one if the taxpayer is male, and zero otherwise.

[Table 6] Summary statistics of private insurance and taxpayers' demographic variables: 2007-2015

Year	mean	standard deviation	min	max
private insurance	227.5	190.5	2.0	6,535.0
age	41.1	9.8	18	77
marriage dummy	0.80	0.40	0	1
male dummy	0.70	0.46	0	1

Note: These summary statistics are computed from a sample of 20,831 labor income taxpayers in the NaSTaB survey reporting the expenditure on private insurance between 2007 and 2015. The marriage dummy is one if the taxpayer is married, and zero otherwise. The male dummy is one if the taxpayer is male, and zero otherwise. The unit of private insurance is 10,000 KRW.

Motivated by this line of idea, we investigate the presence of mass points in two tax deductible expenditure distributions of Korean wage earners: credit card expenditures[11] and private insurance expenses. To this purpose, we use the National Survey of Tax and Benefit (NaSTaB) from survey years 2008-2016 for our analysis.[12] We restrict the samples to individuals who report positive wage incomes

by their annual incomes of a tax year. If a majority of taxpayers commit an error, little bunching is observed at $\underline{E}$ and $\overline{E}$. We expect that such an error is not common for wage earners since there are little unexpected variation in wage incomes, and thus they can guess annual incomes with high precision.

[11] In 1999, Korean government introduced credit card expenditure tax deduction with an aim to better keep track of business incomes, and thus decrease the size of the underground economy. A wage earner can claim a fraction of his/her expenditures paid through credit card system, debit card system, and cash receipt system. Overseas expenditure, educational expenses, insurance expenses, utility charges, rents, charitable donation, and some other categories of expenses are excluded in measuring credit card expenditures. Such an incentive system makes an individual to use credit/debit cards or request a cash receipt, which generates the transaction record accessible to the National Tax Services (NTS).

[12] The NaSTaB is a panel survey of Korean households, which aims to find the distributions of

and the expenditure for a tax deductible item. Individuals with business incomes are dropped since they are treated differently in tax deductions. All remaining data are pooled. As a result, we construct two samples. The first sample consists of 12,589 wage earners reporting positive wages and credit card expenses during this period. The summary statistics of this sample is presented in Table 5. The average age in this sample is 39.8, and the shares of the married and males are 80% and 72%, respectively. On average, an individual in this sample reports approximately 15 million KRW as credit card expenditure. The second sample includes 20,831 wage earners reporting positive wages and expenditures for the private insurance against disease, accident and death between 2007-2015. As seen in Table 6, the demographic composition of this sample is similar to that of the first sample. The average age is 41.1, and the shares of married individuals and males are 80% and 70%, respectively. The average expenditure for private insurance is reported to be approximately 2.3 million KRW.

[Table 7] Tax deduction schedule for credit card expenditure: Assessment year 2007-2015 (unit: 10,000 KRW)

Year	lower end ($\underline{E}$)	upper end ($\overline{E}$)	deduction rate (k)
2007-2009	$0.2Y$	$\min\{0.2Y+2500,\ 1.2Y\}$	0.2
2010-2012	$0.25Y$	$\min\{0.25Y+1500,\ 1.25Y\}$	0.2
2013-2015	$0.25Y$	$\min\{0.25Y+2000,\ 19/12Y\}$	0.15

Note: The Korean tax law has required the upper end $\overline{E}$ to depend on the labor earnings of an employee Y. The upper end $\overline{E}$ is computed by using the tax schedule and labor earnings. A higher deduction rate has been applied to the amount of expenditures paid by a debit card since 2010. For those years, we assume that all expenditures are paid by credit cards in computing the upper end $\overline{E}$.

Table 7 presents the tax deduction schedule for credit card expenditures.[13] Between 2007 and 2009, the lower end ($\underline{E}$) was set at the 20% of labor income and the upper end of the upper end ($\overline{E}$) was set at the minimum of (i) 15 million KRW plus 20% of labor income and (ii) 120% of labor income. The deduction rate k was 20%. In 2010, the lower and upper bound increased to 25% of labor income and the minimum of (i) 125% of labor income and (ii) 15 million KRW plus 25% of

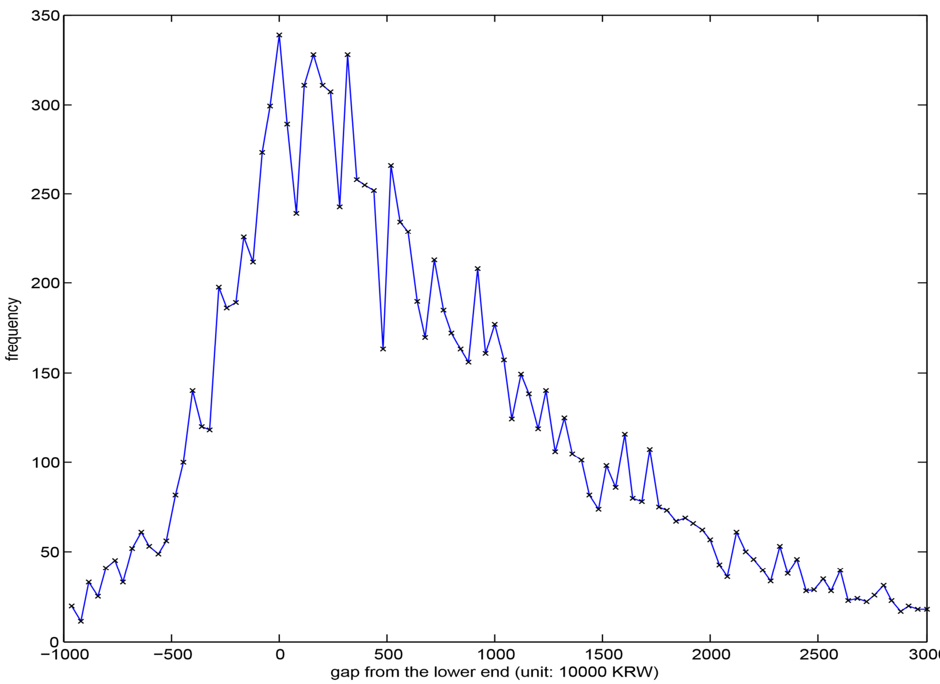
taxes and benefits across households and their members. To serve this goal, the NaSTaB collects information on last year's tax burdens and benefits of each household and its members. Therefore, our sample of the NaSTaB 2008-2016 includes the expenditures for the period between 2007 and 2015.

[13] Korean tax law stipulates the maximum tax deduction for the credit card expenditure instead of the upper end $\overline{E}$ of the phase-in area. For example, in 2008 the maximum tax deduction for the credit card expenditure is stated as a minimum of 5 million KRW and the 20% of labor income. Using the relationship between the maximum tax deduction and the upper end $\overline{E}$, we can compute $\overline{E}$ with given deduction rate k and lower end $\underline{E}$.

labor income. In 2013, the deduction rate was lowered to 15%. The upper bound was also lowered to the minimum of (i) 158.3% of labor income and (ii) 20 million KRW plus 25% of labor income.

Figure 2 presents the distribution of credit card expenditure of 12,589 wage earners in the sample. The horizontal axis indicates the difference between the actual expenditure and the lower bound in the phase-in area, that is, $E-\underline{E}$. We divide the empirical support into bins with a size of 1 million KRW and plot the frequency in each bin at the mid point. As expected, the highest frequency is observed at 0. This finding implies the possibility that individuals' optimizing behavior generates an excessive observation at the lower bound $\underline{E}$. To confirm the presence of probability masses, we conduct the PMP tests with $h_1 = 0.5n^{-0.4}\hat{\sigma}$ and $h_2 = n^{-0.4}\hat{\sigma}$ for T_L, and $h_1 = 1/n\hat{\sigma}$, $h_2 = n^{-0.2}\hat{\sigma}$, and $d_0 = 0.3$ for T_S, where $\hat{\sigma}$ is the sample standard deviation[14]. Table 9 presents the PMP test statistics. The

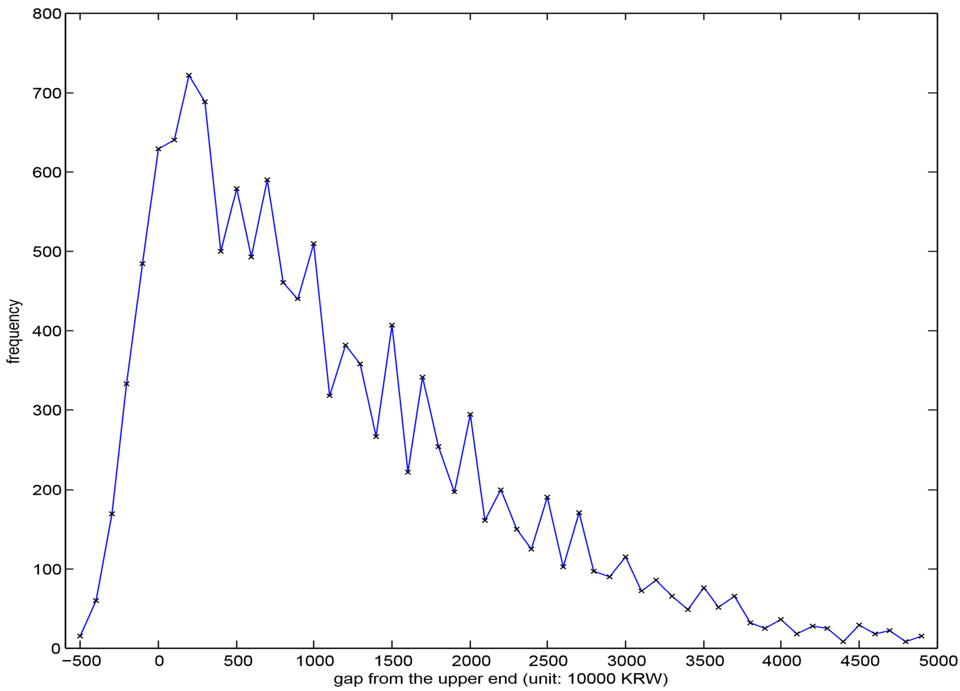
[Figure 2] The distribution of credit card expenditure from the lower end $\underline{E}$



Note: This graph shows the credit card expenditure distribution of 12,589 wage earners in the NaSTaB survey reporting positive wage and credit card expenditure between 2007 and 2015. The horizontal axis indicates the difference between the expenditure and the lower end, that is, $E-\underline{E}$. The differences are divided into bins with size 1 million KRW. The frequency in each bin is plotted at the mid point.

[14] The same procedure is used in implementing the PMP tests for other bunching points in this section.

[Figure 3] The distribution of credit card expenditure from the upper end $\bar{E}$



Note: This graph shows the credit card expenditure distribution of 12,589 wage earners in the NaSTaB survey reporting positive wage and credit card expenditure between 2007 and 2015. The horizontal axis indicates the difference between the expenditure and the upper end. The differences are divided into bins with size 1 million KRW. The frequency in each bin is plotted at the mid point.

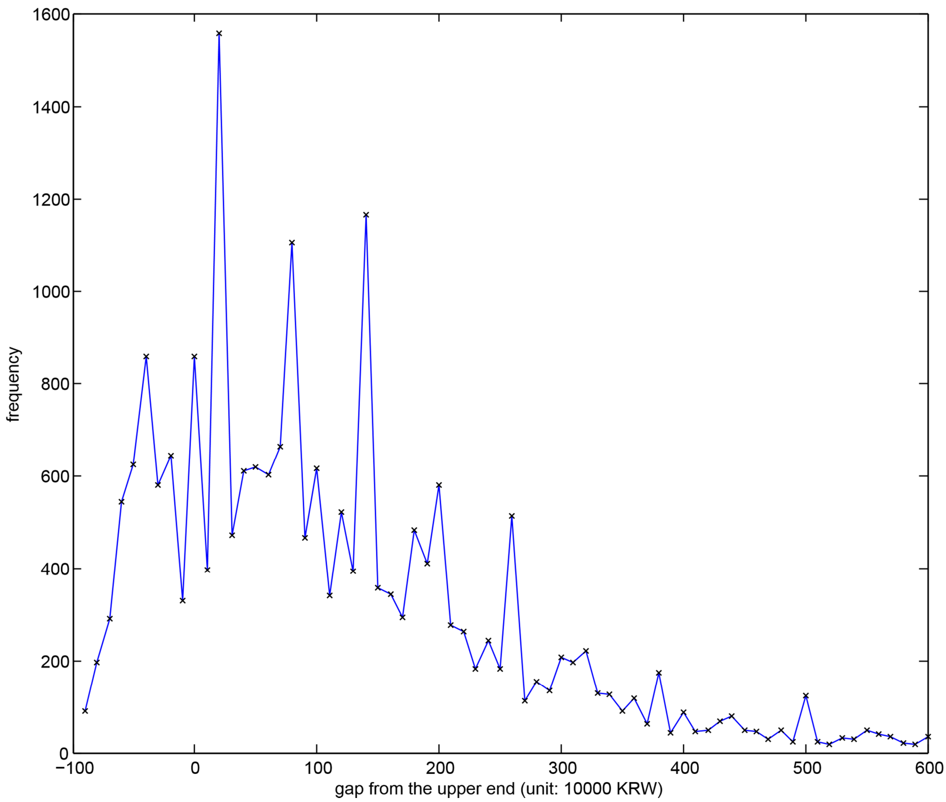
test statistics at $\underline{E}$ turn out to be $T_L = 6.35$ and $T_S = 178.70$. Hence, the hypothesis of no point masses at $\underline{E}$ is rejected at the 1% significance level. We also conduct the PMP tests for the hypothesis of no point masses at the upper end $\bar{E}$. As shown in Figure 3, the bunching at the upper end $\bar{E}$ is not as much striking as the one at the lower end $\underline{E}$. Despite the fact, the PMP tests reject the null hypothesis of no point masses at $\bar{E}$ at the 1% significance level, and thus imply that there is a bunching of positive probability masses at $\bar{E}$.

As another example, we consider the distribution of wage earners' expenditures for the private insurance against disease, accident and death. Table 8 presents the tax deduction schedule for private insurance expenditures. During the entire years of analysis, the lower and upper end remained at 0 and 1 million KRW, respectively. Figure 4 displays the distribution of the private insurance expenditures of 20,831 wage earners in the sample between 2007 and 2015. In this example we investigate only the upper end $\bar{E}$ since the lower end $\underline{E}$ is located at the boundary of the support of the private insurance expenditure distribution. A hike is observed at $\bar{E}$.

[Table 8] Tax deduction schedule for private expenditure: Assessment year 2007-2015

Year	lower end ($\underline{E}$)	upper end ($\overline{E}$)	deduction rate (k)
2007 - 2015	0	1 mil. KRW	1

[Figure 4] The distribution of private insurance expenditure from the upper end $\overline{E}$



Note: This graph shows the private insurance expenditure distribution of 20,831 wage earners in the NaSTaB survey reporting positive wage and the expenditure on private insurance between 2007 and 2015. The horizontal axis indicates the difference between the expenditure and the upper end. The differences are divided into bins with size 100 thousand KRW. The frequency in each bin is plotted at the mid point.

The PMP test results at $\overline{E}$ are presented in the lower panel of Table 9. The PMP test statistics are $T_L = 20.26$ and $T_S = 654.07$. Thus, both PMP tests reject the null hypothesis of no probability masses at $\overline{E}$ at the 1% significance level, which leads to the conclusion that there is a bunching of observations at the upper end $\overline{E}$.

[Table 9] The PMP tests of Korean wage earners' expenditure distributions

Panel I. Credit card expenditure		
	T_L	T_S
H_0 : No point masses at $\underline{E}$	6.35	178.70
H_0 : No point masses at $\bar{E}$	6.76	265.50
Panel II. Private Insurance expenditure		
	T_L	T_S
H_0 : No point masses at $\bar{E}$	20.26	654.07

Note: Numbers in each cell present the PMP test statistics of the null hypothesis of no probability mass at a specified point. $\hat{\sigma}$ is the sample standard deviation of the normalized gap. Bandwidths ($h_1 = 0.5n^{-0.4}\hat{\sigma}$ and $h_2 = n^{-0.4}\hat{\sigma}$) are used for T_L. Bandwiths ($h_1 = n^{-1}\hat{\sigma}$, $h_2 = n^{-0.4}\hat{\sigma}$) and $d_0 = 0.3$ are chosen for T_S. A uniform kernel are commonly used for both PMP tests. The critical values for T_S at the 1% and 5% significance levels are 4.93 and 3.13, respectively. All PMP tests reject the null hypothesis at the 1% significance level.

VII. Concluding Remarks

In this study, we proposed two PMP test statistics to detect the presence of mass points among non-mass points. We derived the limiting distributions of the proposed test statistics under the null and alternative hypothesis by exploiting the asymptotic difference between two kernel density estimators using different bandwidths. Specifically, the proposed PMP test statistic is shown to converge to either the standard normal distribution or a linear transformation of a positive Poisson distribution at a non-mass point depending on bandwidths choice while it diverges to the infinity at a mass point. The consistency and size properties of the PMP tests are immediate from the limiting distributions. Numerical experiments are conducted to confirm the theoretical properties of the PMP tests.

We apply the PMP tests to see if there is a bunching behavior among taxpayers facing kinked effective tax rate due to the deduction schedule. In general, it is expected that PMP tests can be useful as a pretest when researchers are doing some kernel-based nonparametric analyses. For example, they 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 the 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).

An open question for future research is to detect all mass points in the entire support. When mass points are assumed to result from a certain stochastic transformation from a true continuous variable, a researcher might want to have the set of all mass points to recover the underlying distribution of a true variable. The consistent detection of all existing mass points requires a multiple hypotheses testing procedure which detects all mass points as well as does not falsely detect non-mass points with probability approaching 1. Such a procedure can be developed by applying the idea of a complete consistent test in Andrews (1986) and a multiple hypotheses procedure, for example, the Holm procedure in Holm (1979) to our PMP tests.

Appendix

A. Proofs

Lemma A1. *Under Assumptions 1-3, the mean and variance of* $\widehat{f(h)}$ *are (2) and (3).*

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

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

Recognize that

$$\frac{1}{h}K\left(\frac{y_0 - d_j}{h}\right) = \frac{1}{h}K(0) \quad \text{for} \quad y_0 = d_j , \tag{16}$$

$$\frac{1}{h}K\left(\frac{y_0 - d_j}{h}\right) \to 0 \quad \text{for} \quad y_0 \neq d_j . \tag{17}$$

Plug (16) and (17) into the previous expression for the mean of $\widehat{f(h)}$ to have

$$\begin{aligned} E[\widehat{f(h)}] &= p(y_0)\frac{1}{h}K(0) + \left(\sum_{d_j \in D, d_j \neq y_0} p(d_j)\right) o(1) + (1-p)\int \frac{1}{h}K\left(\frac{y_0 - y}{h}\right) f_2(y)dy \\ &= p(y_0)\frac{1}{h}K(0) + \left(\sum_{d_j \in D, d_j \neq y_0} p(d_j)\right) o(1) \\ &\quad + (1-p)\left(f_2(y) + \frac{1}{2} f_2^{(2)}(y_0)\int s^2 K(s)ds + o(h^2)\right), \end{aligned}$$

where $p(d_j) = pf_1(d_j)$ and the last equality results from the well-known result for a kernel density estimator without the presence of mass points. Then, Equation (2) immediately follows from the last equation.

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

$$\begin{aligned} V[\widehat{f(h)}] &= \frac{1}{n}V\left[\frac{1}{h}K\left(\frac{y_0 - Y_i}{h}\right)\right] \\ &= \frac{1}{n}\left(\frac{1}{h^2}E\left[K^2\left(\frac{y_0 - Y_i}{h}\right)\right] - \left(E\left[\frac{1}{h}K\left(\frac{y_0 - Y_i}{h}\right)\right]\right)^2\right). \end{aligned} \tag{18}$$

Firstly,

$$\frac{1}{h^2}E\left[K^2\left(\frac{y_0-Y_i}{h}\right)\right]=\frac{1}{h^2}\sum_{y_j\in D}K^2\left(\frac{y_j-y_0}{h}\right)p(y_j)+\frac{1-p}{h^2}\int K^2\left(\frac{y_0-y}{h}\right)f_2(y)dy$$
$$=\frac{1}{h^2}K^2(0)p(y_0)+\left(\sum_{d_j\in D,d_j\neq y_0}p(d_j)\right)o(1)$$
$$+\frac{1-p}{h}\Big(f_2(y_0)\int K^2(s)ds+o(1)\Big). \tag{19}$$

Plug (2) and (19) into (18). Then

$$V[\widehat{f(h)}]=\frac{1}{nh^2}K^2(0)p(y_0)(1-p(y_0))+\frac{1}{nh}(1-p)f_2(y_0)\int K^2(s)ds$$
$$-\frac{1}{n}\left(\frac{1}{h}K(0)p(y_0)+\big(1-p\big)f_2(y_0)+o(1)\right)^2+p(y_0)o\left(\frac{1}{nh^2}\right)$$
$$+(1-p)o\left(\frac{1}{nh}\right)$$
$$=\frac{1}{nh^2}K^2(0)p(y_0)(1-p(y_0))+\frac{1}{nh}(1-p)f_2(y_0)\int K^2(s)ds$$
$$-\frac{1}{nh}2p(y_0)(1-p)f_2(y_0)K(0)+p(y_0)o\left(\frac{1}{nh^2}\right)+(1-p)o\left(\frac{1}{nh}\right)$$
$$=\frac{1}{nh^2}p(y_0)\Big((1-p(y_0))K^2(0)-2h(1-p)f_2(y_0)K(0)\Big)$$
$$+\frac{1}{nh}(1-p)f_2(y_0)\int K^2(s)ds+p(y_0)o\left(\frac{1}{nh^2}\right)+(1-p)o\left(\frac{1}{nh}\right)$$
$$=\frac{1}{nh^2}p(y_0)\Big((1-p(y_0))K^2(0)+o(1)\Big)+\frac{1}{nh}(1-p)f_2(y_0)\int K^2(s)ds$$
$$+p(y_0)o\left(\frac{1}{nh^2}\right)+(1-p)o\left(\frac{1}{nh}\right)$$
$$=\frac{1}{nh^2}p(y_0)(1-p(y_0))K^2(0)+\frac{1}{nh}(1-p)f_2(y_0)\int K^2(s)ds$$
$$+o\left(p(y_0)\frac{1}{nh^2}+\frac{1}{nh}\right). \qquad \blacksquare$$

To prove Theorem 1, we investigate the mean and variance of the difference of the two kernel density estimators in Lemmas A2 and A3.

Lemma A2. *Under Assumptions 1-3 with* $h_1 = ch_2$, $c \in (0,1)$, *the mean and variance of* $\widehat{f(h_1)} - \widehat{f(h_2)}$ *are given by*

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

and

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

where $\mu_1 = (1-c)K(0)p(y_0)$, $V_0 = (1-p)k_c f_2(y_0)$, and $V_1 = (1-c)^2 K^2(0)p(y_0)(1-p(y_0))$.

Proof. By using Equation (2), we have

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

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

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

The first two terms immediately follow from (3). Let $w_{1i} = \frac{1}{h_1}K(\frac{y_i - y_0}{h_1})$ and $w_{2i} = \frac{1}{h_2}K(\frac{y_i - y_0}{h_2})$. The last term is written as

$$\begin{aligned} Cov(\widehat{f(h_1)}, \widehat{f(h_2)}) &= \frac{1}{n}E(w_{1i}w_{2i}) - \frac{1}{n}E(w_{1i})E(w_{2i}) \\ &= \frac{1}{n}E(w_{1i}w_{2i}) - \frac{1}{nh_1h_2}K^2(0)p(y_0)^2 + o\left(p(y_0)\frac{1}{nh_1^2} + \frac{1}{nh_1}\right), \quad (20) \end{aligned}$$

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

$$E(w_1 w_2) = \frac{1}{h_1 h_2}K(0)^2 p(y_0) + (1-p)\frac{1}{h_2}\int K(s)K\left(\frac{h_1}{h_2}s\right) f_2(y_0 - h_1 s)ds + o(1) \quad (21)$$

because of (16) and (17). The second term is decomposed into two parts.

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

First,

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

where y^* lies between y_0 and y_0-h_1s. Second, notice that the Lipschitz-continuity of a kernel $K(\cdot)$ implies

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

for a constant d^*. Therefore, we have

$$\left|\frac{1}{h_2}\int K(s)\left(K\left(\frac{h_1}{h_2}s\right)-K(cs))\right)f(y_0+h_1s)ds\right|\le\frac{d^*}{h_2}\left|\frac{h_1}{h_2}-c\right|\int|s||K(s)|f(y_0+h_1s)ds$$
$$=o\left(\frac{1}{h_2}\right),$$

where the last equality comes from the properties of a kernel $K(\cdot)$ described in the definition. Plug these results into (21) to obtain

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

Plugging this equation into (20) provides

$$Cov(\widehat{f(h_1)},\widehat{f(h_2)}) = \frac{1}{nh_1h_2}K(0)^2 p(y_0)(1-p(y_0))$$
$$+(1-p)\frac{1}{nh_2}f_2(y_0)\int K(s)K(cs)ds + o\left(p(y_0)\frac{1}{nh_1^2}+\frac{1}{nh_1}\right).$$

This result together with (3) imply that

$$V(\widehat{f(h_1)}-\widehat{f(h_2)}) = \frac{1}{nh_1^2}(1-c)^2 K(0)^2 p(y_0)(1-p(y_0))$$
$$+\frac{1}{nh_1}(1-p)f_2(y_0)\Big((1+c)\int K(s)^2 ds - 2c\int K(s)K(cs)ds\Big)$$
$$+o\left(p(y_0)\frac{1}{nh_1^2}+\frac{1}{nh_1}\right).$$ ■

Lemma A3. *Under Assumptions 1-3 with* $h_1 = ch_2$, $c \in (0,1)$,

$$T_L'' = (\widehat{f(ch)}-\widehat{f(h)}-E[\widehat{f(ch)}-\widehat{f(h)}])/\sqrt{V(\widehat{f(ch)}-\widehat{f(h)})} \xrightarrow{d} N(0,1) \tag{22}$$

as $n \to \infty$.

Proof. Write T_L'' as follows.

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

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

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

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

Notice that this result together with a condition

$$\sum_{i=1}^{n} E\,|\,L_{n,i}\,|^{2+\delta} < \infty, \quad \delta > 0 \tag{23}$$

are the conditions of Liapounov's central limit theorem so that (22) is shown to be satisfied. Therefore, let us show that (23) is satisfied.

We have

$$\sum_{i=1}^{n} E\,|\,L_{n,i}\,|^{2+\delta} = \left(\frac{1}{\sqrt{nV(w_{1i}-w_{2i})}}\right)^{2+\delta} \sum_{i=1}^{n} E\,|\,w_{1i}-w_{2i}\,|^{2+\delta}$$
$$\leq n2^{1+\delta}\left(\frac{1}{\sqrt{nV(w_{1i}-w_{2i})}}\right)^{2+\delta} (E\,|\,w_{1i}\,|^{2+\delta} + E\,|\,w_{2i}\,|^{2+\delta}), \qquad (24)$$

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

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

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 (24) is bounded by

$$n2^{3+\delta}\left(\frac{1}{\sqrt{nV(w_{1i}-w_{2i})}}\right)^{2+\delta}$$
$$\times\Big(p(y_0)K(0)^{2+\delta}(h_1^{-2-\delta} + h_2^{-2-\delta}) + (1-p)f_2(y_0)\int |\,K(s)\,|^{2+\delta}\, ds (h_1^{-1-\delta} + h_2^{-1-\delta}) + o(h_1^{-1-\delta})\Big). \qquad (25)$$

First, consider a case where $p(y_0) = 0$. In such a case, (25) is written as

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

which is again bounded by

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

$$= 2^{3+2\delta}\left(\frac{1}{\sqrt{h_1 V(w_{1i}-w_{2i})}}\right)^{2+\delta}(nh_1)^{-\delta/2}\left((1-p)f_2(y_0)\int|K(s)|^{2+\delta}\,ds + o(1)\right)$$
$$+2^{3+2\delta}\left(\frac{1}{\sqrt{h_2 V(w_{1i}-w_{2i})}}\right)^{2+\delta}(nh_2)^{-\delta/2}\left((1-p)f_2(y_0)\int|K(s)|^{2+\delta}\,ds + o(1)\right).$$

Lemma A2 implies $h_j V(w_{1i}-w_{2i}) = nh_j V(\widehat{f(h_1)}-\widehat{f(h_2)}) = O(1)$, $j=1,2$. 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 $p(y_0)=0$.

Second, consider a case where $p(y_0)\neq 0$. (25) is written as

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

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

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

Remember that $nh_j^2 V(\widehat{f(h_1)}-\widehat{f(h_2)}) = h_j^2 V(w_{1i}-w_{2i}) = O(1)$ 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.

(*The limiting distribution of* T_L *under* $H_0 : p(y_0) = 0$):

First, consider the case where $\alpha_1 = \alpha_2 < 1$. The asymptotic standard normality of T_L immediately follows from Lemmas A2 and A3 since under $H_0 : p(y_0) = 0$,

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

where the second equality follows from Lemma A2, and $nh_1 h_2^4 = o(1)$ by assumption.

Second, consider the case where $\alpha_2 < \alpha_1 = 1$. According to Pagan and Ullah (1999),

$$\sqrt{nh_1}(\widehat{f(h_1)} - f_2(y_0)) \xrightarrow{d} N\left(0, f_2(y_0)\int K^2(s)ds\right) \tag{26}$$

under the given conditions. In a similar manner, we have

$$\widehat{f(h_2)} - f_2(y_0) = O_p\left(\frac{1}{\sqrt{nh_2}}\right).$$

Therefore,

$$\begin{aligned} \sqrt{nh_1}(\widehat{f(h_1)} - \widehat{f(h_2)}) &= \sqrt{nh_1}(\widehat{f(h_1)} - f_2(y_0)) + \sqrt{\frac{h_1}{h_2}}\sqrt{nh_2}(\widehat{f(h_2)} - f_2(y_0)) \\ &= \sqrt{nh_1}(\widehat{f(h_1)} - f_2(y_0)) + o_p(1), \end{aligned}$$

which is shown to converge to a normal distribution with mean 0 and variance $f_2(y_0)\int K^2(s)$ by (26). Then T_L is shown to converge to a standard normal distribution by recognizing that $k_c = \int K^2(s)ds$ when $c = 0$.

(*The limiting property of* T_L *under* $H_1 : p(y_0) > 0$):

Recall the facts that $h_j \widehat{f(h_j)} = K(0)p(y_0) + o_p(1)$, $j = 1,2$. These facts imply that under $H_1 : p(y_0) > 0$,

$$T_L = \sqrt{n}\frac{1}{\sqrt{k_c K(0)p(y_0)}}\left(1-\frac{h_1}{h_2}\right)(K(0)p(y_0)+o_p(1)). \tag{27}$$

Since $\lim \frac{h_1}{h_2} = c \in [0,1)$, (27) is stated as

$$T_L = \sqrt{n}\left((1-c)\sqrt{\frac{K(0)p(y_0)}{k_c}} + o_p(1)\right).$$

Therefore, T_L diverges as $n \to \infty$ when $H_1 : p(y_0) > 0$ holds true. ■

Lemma A4. *Suppose that Assumption 1 is satisfied. Also assume that* $h_1 = c_0 / n$ *for a positive constant* c_0 *and* h_2 *satisfies Assumption 3. Let* y_0 *be a point in its empirical support. Then under* $H_0 : p(y_0) = 0$*,*

$$T_S' = \frac{\sqrt{nh_1}(\widehat{f_u(h_1)} - \widehat{f_u(h_2)})}{\sqrt{\hat{V}_{u,0}}} \xrightarrow{d} \frac{W^+(k^*) - k^*}{\sqrt{k^*}}$$

as $n \to \infty$. Under $H_1 : p(y_0) > 0$, T_S' diverges to ∞ with probability 1 as $n \to \infty$.

Proof. Firstly, consider the asymptotic distribution of T_S' under H_0. As a first step, let us show Equation (6). Let $W_i = 1(y_0 - \frac{h_1}{2} \le Y_i \le y_0 + \frac{h_1}{2})$. Then

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

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

$$P_{h_1}(y_0) = h_1(1-p)f_2(y_0) + O(h_1^2).$$

It implies

$$nP_{h_1}(y_0) = nh_1(1-p)f_2(y_0) + O(nh_1^2) \to k^* \tag{28}$$

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

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

From Equation (28) 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 $nh_1\widehat{f_u(h_1)}$ has a positive value. Therefore, it can be shown that

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

Under the given assumptions,

$$\widehat{f_u(h_2)} \xrightarrow{p} (1-p)f_2(y_0) \quad \text{and} \quad \hat{V}_{u,0} = k_c\widehat{f(h_2)} \xrightarrow{p} (1-p)f_2(y_0) .$$

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

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

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

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

Rewrite T_S' as

$$T_S' = \sqrt{n}\sqrt{\frac{h_2}{h_1}}\left(\frac{h_1\widehat{f_u(h_1)}}{\sqrt{h_2\widehat{f_u(h_2)}}} - \frac{h_1}{h_2}\sqrt{h_2\widehat{f_u(h_2)}}\right). \tag{29}$$

Then the last result immediately follows. ■

Proof of Theorem 2.

Suppose that $p(y_0) = 0$. If $f_2(y_0) > d_0$, and thus, $k^* > c_0 d_0$, it follows from Lemma A4 that

$$\begin{aligned}\lim_{n\to\infty}\Pr(T_S > \overline{t}_\alpha \mid p(y_0)=0) &= \Pr\left(\frac{W^+(k^*)-k^*}{\sqrt{k^*}} > \overline{t}_\alpha \mid p(y_0)=0\right)\\ &\le \Pr\left(\frac{W^+(k^*)-k^*}{\sqrt{k^*}} > t_\alpha(k^*) \mid p(y_0)=0\right)\\ &\le \alpha .\end{aligned}$$

If $f_2(y_0) \le d_0$,

$$\lim_{n\to\infty}\Pr(T_S > \overline{t}_\alpha \mid p(y_0)=0) \le \lim_{n\to\infty}\Pr(\widehat{f_u(h_2)} > d_0 \mid p(y_0)=0) = 0 .$$

Therefore, (8) is shown. Suppose $p(y_0) > 0$. Then it is shown that $\widehat{f_u(h_2)}$ diverges to ∞ with probability 1. This fact together with Lemma A4 imply (9). ■

B. The Modified HCP Test

HCP (2011) concern the presence of infinite density at the median. They developed a test by exploiting the asymptotic behaviors of L_1-estimators in Knight (1998). In a least absolute deviation (LAD) estimation of $y_i = \mathbf{x}_i'\boldsymbol{\beta} + u_i$, the idea of the HCP test relies on the fact that under suitable conditions, a statistic

$$B_n^* = 4n\hat{f}_{\hat{u}}(0)^2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\hat{\mathrm{C}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) ,$$

where $\hat{f}_{\hat{u}}(0)$ is a kernel density estimator of the residuals ($\hat{u}_i := y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}(q)$) at 0 and $\hat{\mathbf{C}} = n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i'$. The statistic B_n^* converges in distribution to $\chi^2(p)$ under H_0, where p is the dimension of $\mathbf{x}_i$ and diverges to ∞ with probability 1 under H_1. B_n^* cannot be used as a test statistic since the true parameter value $\boldsymbol{\beta}$ is not available in a general situation. HCP (2011) proposed a test statistic, which does not require the knowledge of $\boldsymbol{\beta}$. Specifically, HCP proposed to use B_n as a test statistic, which is defined as

$$B_n = 4n\hat{f}_{\hat{u}}(0)^2(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)'\hat{\mathrm{C}}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) ,$$

where $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are the LAD estimators of evenly divided subsamples. HCP showed that B_n share the same limiting behaviors with B_n^* under both the null and alternative hypotheses.

As briefly mentioned, the current testing problem is different from the one in HCP (2011). First, our testing problem does not have a specific quantile regression that we should explicitly consider. Let x_i be a random variable independent with y_i. Then, the slope parameter of a quantile regression (and as well as a LAD regression) between y_i and x_i is 0 by assumption. To be more specific, consider a quantile regression

$$y_i = \beta_0(q) + \beta_1(q)x_i + u_i = \mathbf{x}_i'\boldsymbol{\beta}(q) + u_i, \tag{30}$$

where $\boldsymbol{\beta}(q) = (\beta_0(q), \beta_1(q))'$, $\beta_0(q)$ and $\beta_1(q)$ are parameters of the q-th quantile regression, and u_i is an error term. By the independence between y_i and x_i, $\beta_1(q) = 0$. Moreover, $\beta_0(q)$ is the q-th quantile of y_i if we impose an identifying restriction that the q-th quantile of u_i is 0. We know the true parameter value of (a part of) $\boldsymbol{\beta}(q)$, which is contrast to the case in HCP (2011). As a result, we do not need to use the sample-splitting method in HCP (2011). Instead we directly use the statistic B_n^*. Second, y_0 where we want to test the presence of probability masses is not necessarily the median, while the original version of the HCP test investigates the infinite density at the median. Let y_0 be the q_0-th quantile of y. The alternative hypothesis that y has infinite density at y_0 ($H_1 : f_y(y_0) = \infty$) implies the infinite density of u in a quantile regression (30) at 0 ($f_u(0) = \infty$) only when $q = q_0$.[15] As soon as q is available, we can use genearalize the idea of HCP (2011) and the results in Knight (1998) to have B_n^* for q, $0 < q < 1$:

$$B_n^* = nq^{-1}(1-q)^{-1}\hat{f}_{\hat{u}}(0)^2(\hat{\boldsymbol{\beta}}(q) - \boldsymbol{\beta}(q))'\hat{\mathrm{C}}(\hat{\boldsymbol{\beta}}(q) - \boldsymbol{\beta}(q)), \tag{31}$$

which has the same limiting distributions with the median case where $q = 1/2$. In practice, we do not know the exact value of q. Therefore, we use an estimate of q, $\hat{q}$. A technical difficulty in defining $\hat{q}$ occurs if there are ties at y_0. In such a case, multiple values for q can be matched to y_0. To overcome this problem, we define $\hat{q}$ as follows.[16]

$$\hat{q} = \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n} I(y_i \le y_0) + \frac{1}{2}\left(1 + \frac{1}{n}\sum_{i=1}^{n} I(y_i < y_0)\right). \tag{32}$$

With $\hat{q}$ in (32), we obtain $\hat{q}$-th quantile regression estimate:

---

[15] From the definition of the quantile, it readily follows that $\beta_0(q) = y_0$.

[16] Admittedly, there are other ways in defining $\hat{q}$. For example, one can use the empirical cdf as $\hat{q}$.

$$\hat{\boldsymbol{\beta}}(\hat{q}) = \arg\min_{\mathbf{b}} \frac{1}{n}\sum_{i=1}^{n}(1-\hat{q})\,|\,y_i - \mathbf{x}_i'\mathbf{b}\,|\,I(y_i - \mathbf{x}_i'\mathbf{b} \le 0) + \hat{q}\,|\,y_i - \mathbf{x}_i'\mathbf{b}\,|\,I(y_i - \mathbf{x}_i'\mathbf{b} > 0)\,. \quad (33)$$

Similar with Corollary 2 in Knight (1998), the solution to (33) can be shown to satisfy

$$a_n(\hat{\boldsymbol{\beta}}(\hat{q}) - \boldsymbol{\beta}(q)) \xrightarrow{d} \arg\min_{\mathbf{u}=(u_1,u_2)'} -u_2 W + 2\tau(\mathbf{u})\,, \quad (34)$$

where a_n is suitable normalizing factor, $W \sim N(0, q(1-q)\sigma_x^2)$, and $\tau(\mathbf{u})$ is a quadratic term involving u_1 and u_2.[17] Knight (1998) showed that under H_0,

$$\tau(\mathbf{u}) = \frac{\lambda}{2}\mathbf{u}'\mathbf{C}\mathbf{u}$$

where $\lambda = f_u(0)$. Since we draw x_i from a distribution with mean 0, this result is simplified as follows: $\tau(\mathbf{u}) = \frac{\lambda}{2}(u_1^2 + \mu_{2,x}u_2^2)$ where $\mu_{2,x} = E(x_i^2) = \sigma_x^2$. By following the steps in Knight (1998), we can show that the solution to the minimization problem (34) under H_0 is given by

$$(u_1, u_2) = (0, -\lambda^{-1}\mu_{2,x}^{-1}W)$$

and $a_n = \sqrt{n}$. Thus, the slope parameter estimate of the $\hat{q}$-th quantile regression is asymptotically normal under H_0:[18]

$$\sqrt{n}(\hat{\beta}_1(\hat{q}) - \beta_1(q)) \xrightarrow{d} N(0, q(1-q)\mu_{2,x}^{-1}f_u(0)^{-2})\,. \quad (35)$$

The asymptotic property of $\hat{\beta}_1(\hat{q})$ under H_1 also remains the same despite the use of $\hat{q}$. We use the limiting distribution in (35) to obtain the following modified formula for the HCP test $\hat{B}_n(q)$:

$$\hat{B}_n = n\hat{q}^{-1}(1-\hat{q})^{-1}\hat{f}_u(0)^2\beta_1(\hat{q})^2 s_x^2\,,$$

where s_x^2 is the sample variance. Since $s_x^2 \xrightarrow{p} \sigma_x^2$ and $\beta_1(q) = 0$, the modified

[17] The definition of $\tau(\mathbf{u})$ is the same as that used in Knight (1998), which is $\tau(\mathbf{u}) =$ plim $n^{-1}\sum_{i=1}^{n}\Psi_n(\mathbf{u}'\mathbf{x}_i)$, where $\Psi_n(t) = \int_0^t \sqrt{n}(F(s/a_n) - F(0))ds$.

[18] The intercept estimate remains fixed at y_0, that is, $\hat{\beta}_0(\hat{q}) = y_0$. This property results from the fact that y_0 is the $\hat{q}$-th quantile in a sample.

HCP test statistic $\hat{B}_n \xrightarrow{d} \chi^2(1)$ under H_0 and $\hat{B}_n \xrightarrow{p} 0$ under H_1 when $\sqrt{n}h_n \to \infty$ and $h_n \to 0$. Therefore, the modified HCP test would reject $H_0 : f_y(y_0) < \infty$ when $\hat{B}_n > \chi^2_\alpha(1)$, where $\chi^2_\alpha(1)$ is the α-th quantile of the chi-squared distribution with degree freedom of 1.

References

Almunia, M., and D. Lopez-Rodriguez (2018), "Under the Radar: The Effects of Monitoring Firms on Tax Compliance," *American Economic Journal: Economic Policy*, 10, 1–38.

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

________________ (2001), "Complete Consistency: Testing When a Parameter is on the Boundardy of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

Arulampalam, W., V. Corradi, and D. Gutknecht (2017), "Modelling Heaped Duration Data: An Application to Neonatal Motality," *Journal of Econometrics*, 200, 363–377.

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.

Bastani, S., and H. Selin (2014), "Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule," *Journal of Public Economics*, 109, 36–49.

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.

Burtless, G., and R. Moffitt (1984), "The Effect of Social Security Benefits on the Labor Supply of the Aged." In: Aaron, H. J., and G. Burtless (eds), *Retirement and Economic Behavior*, Washington D.C.: Brookings Institution.

Burgstahler, D. and I. Dichev (1997), "Earnings Management to Avoid Earnings Decreases and Losses," *Journal of Accounting and Economics*, 24, 99–126.

Chernoff, H., and E. Lander (1995), "Asymptotic Distribution of the Likelihood Ratio Test That a Mixture of Two Binomials is a Single Binomial," *Journal of Statistical Planning and Inference*, 43, 19–40.

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.

Cho, J., and C. Han (2009), "Testing for the Mixture Hypothesis of Geometric Distributions," *Journal of Economic Theory and Econometrics*, 20, 31–55.

Cho, J., and H. White (2007), "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

Cho, J., and H. White (2010), "Testing for Unobserved Heterogeneity in Exponential and Weibull Duration Models," *Journal of Econometrics*, 157, 458–480.

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.

Gassiat, E., and C. Keribin (2000), "The Likelihood Ratio Test for the Number of Components in a Mixture with Markov Regime," *ESAIM: Probability and Statistics*, 4, 25–52.

Groβ, M., and U. Rendtel (2016), "Kernel Density Estimation for Heaped Data," *Journal of Survey Statistics and Methodology*, 4, 339–361.

Han, C., J. Cho, and P. C. B. Phillips (2011), "Infinite Density at the Median and the Typical Shape of Stock Return Distributions," *Journal of Business & Economic Statistics*, 29, 282–294.

Heitjan, D. F., and D. B. Rubin (1991), "Ignorability and Coarse Data," *The Annals of Statistics*, 19, 2244–2253.

Holbrook, A. L., S. An, 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 Test Procedure," *Scandinavian Journal of Statistics*, 6, 65–70.

Kleven, H., C. Landais, E. Saez, and E. Schultz (2014), "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark," *Scandinavian Journal of Statistics*, 129, 333–378.

Knight, K. (1998), "Limiting Distributions for L_1 Regression Estimators under General Conditions," *Annals of Statistics*, 26, 755–770.

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

Newey, W. K. and K. D. 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.

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.

Torelli, N., and U. Trivellato (1993), "Modelling Inaccuracies in Job-Search Duration Data," *Journal of Econometrics*, 59, 187–211.

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

Wolff, J., and T. Augustin (2003), "Heaping and Its Consequences for Duration Analysis: A Simulation Study," *AStA Advances in Statistical Analysis*, 87, 59–86.

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.