Higher-Order Asymptotic Properties of Kernel Density Estimator with Plug-In Bandwidth

Imai, Shunsuke; Nishiyama, Yoshihiko

KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.1076

“Higher-Order Asymptotic Properties of Kernel Density Estimator with Plug-In Bandwidth”

Shunsuke Imai, Yoshihiko Nishiyama

March 2022

KYOTO UNIVERSITY

KYOTO, JAPAN

Higher-Order Asymptotic Properties of Kernel Density Estimator with Plug-In Bandwidth

Shunsuke Imai* Yoshihiko Nishiyama†

March, 2022

Abstract

This study investigates the effect of bandwidth selection via plug-in method on the asymptotic structure of the nonparametric kernel density estimator. We find that the plug-in method has no effect on the asymptotic structure of the estimator up to the order of $O\{(nh_0)^{-1/2}\} = O(n^{-L/(2L+1)})$ for a bandwidth h_0 and any kernel order L. We also provide the valid Edgeworth expansion up to the order of $O\{(nh_0)^{-1}\}$ and find that the plug-in method starts to have an effect from on the term whose convergence rate is $O\{(nh_0)^{-1/2}h_0\} = O(n^{-(L+1)/(2L+1)})$. In other words, we derive the exact convergence rate of the deviation between the distribution functions of the estimator with a deterministic bandwidth and with the plug-in bandwidth. Monte Carlo experiments are conducted to see whether our approximation improves previous results.

JEL Classification: C14

Key words: nonparametric statistics, kernel density estimator, plug-in bandwidth, Edgeworth expansion

1 Introduction

In nonparametric statistics, the target of statistical inference is a function or an infinite dimensional vector f that is not specifically modelled itself. One of the important components of such functions f is the density function because, in statistics and its related fields, there are cases where we are interested in the distribution as a distribution of income or where a target of statistical inference depends on the density function as a conditional expectation function. Although there are different methods for estimating a density function, we focus on the estimator based on the kernel method, namely *kernel density estimator* (KDE), also called Rosenblatt estimator or Rosenblatt-Parzen estimator after their pioneering works (Rosenblatt 1956, and Parzen 1962).

The first-order asymptotic properties of KDE have been studied over a long period of time and it has been proven that, under certain conditions, KDE has pointwise consistency and asymptotic normality (see e.g. Parzen (1962), as well as the monographs by Li and Racine (2007) or Wasserman (2006)). As we will review in Section 2, the rate of convergence of KDE is slower than the parametric rate, and furthermore, becomes slower as the dimension increases. This property is called the curse of dimensionality. We can understand this is the cost of using local data to avoid misspecification. Hall (1991) has clarified the higher-order asymptotic properties of the estimator in both

*Faculty of Economics, Kyoto University, JAPAN, imai.shunsuke.57n@st.kyoto-u.ac.jp

†Institute of Economic Research, Kyoto University, JAPAN, nishiyama@kier.kyoto-u.ac.jp

non-Studentised and Studentised cases. The asymptotic expansion of KDE is no longer a series of $n^{-1/2}$ as parametric estimators, but a series of $(nh)^{-1/2}$, even in non-Studentised case; it is more complicated series in Studentised case, where n and h are the sample size and bandwidth, respectively.

Bandwidth h specifies the flexibility of statistic models and is adjusted between the bias and variance trade-offs in the sense that creating flexible models and consequently decreasing the bias results in increasing variance while creating non-flexible models and consequently decreasing the variance results in increasing bias. It is well known that the performance of the kernel-based estimators depends greatly on the bandwidth, not on the kernel function. By defining a loss function, one can compute the theoretically optimal bandwidth h_0 that minimises the function. The mean integrated squared error (MISE) is the most commonly used global loss measure. However, in practice, such a bandwidth is typically infeasible. Therefore, one has to choose the bandwidth in a data-driven way. Among the many bandwidth selection methods, two famous ones are cross-validation and plug-in method. In this paper, we focus on the latter.

It is natural to ask whether the choice of bandwidth affects the asymptotic structure of the estimator. Ichimura (2000) and Li and Li (2010) have considered the asymptotic distribution of kernel-based non/semiparametric estimators with data-driven bandwidth. They argue that, under certain conditions, the bandwidth selection has no effect on the first-order asymptotic structure of the estimators. Hall and Kang (2001) showed that the bandwidth selection by the plug-in method also has no effect on the asymptotic structure of KDE up to the order of $O\{(nh_0)^{-1/2}\} = O(n^{\frac{-L}{2L+1}})$.

Our contributions are threefold. First, we provide the Edgeworth expansion of KDE with global plug-in bandwidth up to the order of $O\{(nh_0)^{-1}\} = O(n^{\frac{-2L}{2L+1}})$ and show that the bandwidth selection by the plug-in method starts to have an effect from on the term whose convergence rate is $O\{(nh_0)^{-1/2}h_0\} = O(n^{\frac{-(L+1)}{2L+1}})$. Second, we generalise Theorem 3.2 of Hall and Kang (2001), which states that bandwidth selection via the global plug-in method has no effect on the asymptotic structure of KDE up to the order of $O\{(nh_0)^{-1/2}\} = O(n^{\frac{-L}{2L+1}})$. Their results limit the order of kernel functions $K(u)$ and $H(u)$ to $L = 2, L_p = 6$ respectively, but we show that they are valid for general orders L and L_p as well. Finally, we explore Edgeworth expansion of KDE with deterministic bandwidth in more detail than Hall (1991). We show that Edgeworth expansion of standardized KDE with deterministic bandwidth has the term of order $O\{(nh_0)^{-1/2}\} = O(n^{\frac{-L}{2L+1}})$ right after the term $\Phi(z)$ with a gap between them, but after that the terms decrease at the rate of $O(h_0) = O(n^{\frac{-1}{2L+1}})$. The proof of our main theorem owes much to Nishiyama and Robinson (2000). They have established the valid Edgeworth expansion for the semiparametric density-weighted averaged derivatives estimator of the single index model, which has an exact second-order U-statistic form. Although the higher-order asymptotic structure of U-statistics had been studied before Nishiyama and Robinson (2000), the estimator is different from *standard* U-statistics in that it is U-statistics whose kernel depends on the sample size n through the bandwidth. Since KDE with plug-in bandwidth can also be approximated by a sum of first- and second-order U-statistics whose kernel depend on the sample size n through the bandwidth (see (17) in Appendix A), we can benefit from their proof.

The remainder of this paper is organized as follows. In the next section, we introduce KDE and review its known properties. Section 3 provides the main results, namely the Edgeworth expansion of the estimator with global plug-in bandwidth. In section 4, we compare our results with previous works by Monte Carlo studies. Section 5 concludes and discusses future research directions.

2 Review of the Estimator's Properties

2.1 Estimator and Its First Order Properties

Assumption 1. *Let $\{X_i\}_{i=1}^n$ be a random sample with an absolutely continuous distribution with Lebesgue density f.*

First, we introduce nonparametric KDE $\hat{f}$ for unknown density f. Estimator $\hat{f}$ at a point x with a bandwidth h is defined as follows:

$$\hat{f}_h(x) \equiv \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{X_i - x}{h}\Big),$$

where K is a kernel function, and we say that K is a L-th order kernel, for a positive integer L, if

$$\int u^l K(u)du = \begin{cases} 1 & (l = 0) \\ 0 & (1 \leqq l \leqq L-1) \\ C \neq 0, < \infty & (l = L). \end{cases}$$

Assumption 2. *In a neighbourhood of x, f is L times continuously differentiable and its first L derivatives are bounded.*

Assumption 3. *Kernel function K is a bounded, even function with a compact support, of order $L \geqq 2$ and $\int K(u)du = 1$.*

Assumption 4. *x is an interior point in the support of X.*

Assumption 5. $h \to 0, \quad nh \to \infty \quad as \quad n \to \infty$

KDE has pointwise consistency and asymptotic normality for an interior point in the support of X. Although it also converges uniformly for an interior point in the support of X, we only review pointwise properties because we investigate the pointwise higher-order asymptotics of KDE with global plug-in bandwidth. Under Assumptions 1–3, we can expand mean squared error (MSE) of $\hat{f}_h(x)$ as follows:

$$MSE[\hat{f}(x)] \equiv \mathbb{E}[\{\hat{f}(x) - f(x)\}^2] = \Big(C_L f^{(L)}(x)h^L\Big)^2 + \frac{R(K)f(x)}{nh} + o\{h^{2L} + (nh)^{-1}\}, \tag{1}$$

where $R(K) = \int K(u)^2 du, C_L = \frac{1}{L!}\int u^L K(u)du$. Therefore, Markov's inequality, Assumptions 1–5, and (1) imply pointwise consistency $\hat{f}_h(x) \xrightarrow{p} f(x)$. Moreover, we can show that KDE has asymptotic normality as follows by applying Lindberg-Feller's central limit theorem:

$$\sqrt{nh}\Big(\hat{f}_h(x) - \mathbb{E}\hat{f}_h(x)\Big) \xrightarrow{d} N\Big(0, R(K)f(x)\Big).$$

Remark 1. *Since $\mathbb{E}[\hat{f}_h(x)] \approx f(x) + C_L f^{(L)}(x)h^L$, the statistics centred by $f(x)$ asymptotically follows a zero-mean normal distribution if $nh^{2L+1} \to 0$ holds. However, the theoretically optimal bandwidth does not satisfy this condition, as we will discuss later. Therefore, we consider the statistics centred by $\mathbb{E}[\hat{f}_h(x)]$, not $f(x)$. For recent studies on asymptotic bias of KDE, see, for example, Hall and Horowitz (2013) and Calonico et al. (2018). For other nonparametric estimators, recent related studies are Armstrong and Kolesár (2018), Calonico et al. (2014), Calonico et al. (2020, 2022) and Schennach (2018).*

2.2 Plug-In Method

Bandwidth h is a parameter that analysts need to choose in advance. One of the criteria for bandwidth selection is the mean integrated squared error (MISE):

$$MISE(h) = \int \mathbb{E}[\{\hat{f}_h(x) - f(x)\}^2]dx.$$

The theoretically optimal bandwidth is the one that minimises MISE and, from the MISE expansion, this bandwidth is defined as follows:

$$h_0 = \left(\frac{R(K)}{2LC_L^2 I_L} \right)^{\frac{1}{2L+1}} n^{-\frac{1}{2L+1}},$$

where $I_L = \int f^{(L)}(x)^2 dx$. Although h_0 would perform best, it is infeasible because I_L is unknown, so one has to select the bandwidth from the available data. We examine the effect of a certain plug-in method on the distribution of the estimator.

Several plug-in methods have been proposed so far (see e.g. Hall, Sheather, Jones and Marron 1991, Sheather and Jones 1991). In this paper, we adopt a simple plug-in method that estimates I_L directly and nonparametrically using the estimator proposed by Hall and Marron (1987). Their estimator, $\hat{I}_L$ for I_L, is given as follows:

$$\hat{I}_L = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b^{-(2L+1)} H^{(2L)} \left(\frac{X_i - X_j}{b} \right) \equiv \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{I}_{Lij},$$

where b is a bandwidth for estimation of I_L, different from h (called *pilot bandwidth*), and H is a kernel function of order L_p.

Proposition 2.1 provides the expansion of the plug-in bandwidth (defined as $\hat{h}$) and plays an essential role in the derivation of the asymptotic expansion of KDE with the plug-in bandwidth. We assume additional conditions for Proposition 2.1:

Assumption 6. *In a neighbourhood of x, f is $(2L+L_p)$-times continuously differentiable and its first $(2L+L_p)$ derivatives are bounded.*

Assumption 7. *Kernel function H is a bounded, even function with compact support, of order $L_p \geqq 2$, $(2L)$-times continuously differentiable and for all integers k such that $1 \leqq k \leqq 2L-1$, $\lim_{u\to\pm\infty} |H^{(k)}(u)| \to 0$.*

Assumption 6 gives regularity conditions on the smoothness of the estimand, which implies Assumption 2. Assumption 7 is on the kernel function H for the estimation of I_L, and the condition at the infinity of u is necessary for the integration by parts in the expanding process of $(\hat{h}-h_0)/h_0$. These assumptions can be interpreted as a generalization of assumption (A_{gpi}) of Hall and Kang (2001) to K of order L and H of order L_p.

Proposition 2.1 (Expansion of Plug-In Bandwidth)**.** *Under Assumptions 1, 3, 4, 6, and 7, we can expand $(\hat{h}-h_0)/h_0$ as follows:*

$$\frac{\hat{h}-h_0}{h_0} = \frac{-C_{PI}}{n} \sum_{i=1}^{n} \left\{ f^{(2L)}(X_i) - \mathbb{E} f^{(2L)}(X_i) \right\} + O_p(n^{-1/2} b^{L_p}) + O_p(n^{-1} b^{-(4L-1)/2}) \tag{2}$$

where

$$C_{PI} = \frac{2}{2L+1} I_L^{-1}.$$

The proof is in Appendix A.1.

Remark 2. *Both the first and second terms on the right-hand side of (2) reflect the projection term of the Hoeffding-decomposition of $\hat{I}_L$. The third term reflects that the quadratic term of the decomposed $\hat{I}_L$ converges at the latest $O_p(n^{-1} b^{-(4L-1)/2})$ because of the standard property of U-statistics (see the proof of Lemma 3.1 in Powell, Stock, and Stoker 1989).*

Remark 3. *We can make the second term of (2) as small as we like up to the order of $O(n^{-3/2})$ by letting kernel order L_p be large enough. This is not unrealistic statement; for example, when one uses a second order kernel function, adopting a fourth order kernel function is sufficient to make the effect of the second order term negligible (see Online supplement S2.5 for details).*

Remark 4. *Since MSE optimal rate of b is $O_p(n^{\frac{-2}{8L+2L_p+1}})$ from Hall and Marron (1987), for example, when one choose the pilot bandwidth via rule of thumb (see silverman 1986), the convergence rate of the third term is $O_p(n^{-1}b^{-(4L-1)/2}) = O_p(n^{-1/2}n^{-L_p/(8L+2L_p+1)})$ at the latest. This implies we can also make the third term of (2) as small as we like up to the order of $O(n^{-1})$ by letting kernel order L_p be large enough. Noting, unlike the convergence rate of the second term, the bound on the third term $O_p(n^{-1}b^{-(4L-1)/2})$ is possibly not sharp, we can not immediately identify how much L_p is sufficient to make the effect of pilot bandwidth negligible without deriving the Edgeworth expansion with pilot bandwidth.*

2.3 Review of Previous Studies

Theorem 2.1 of Hall (1991) established the Edgeworth expansion for KDE with a deterministic bandwidth, which we replicate in Proposition 2.2. Let $S_h(x)$ be Standardised version of KDE with a bandwidth h:

$$S_h(x) \equiv \frac{\sqrt{nh}\{\hat{f}_h(x) - \mathbb{E}\hat{f}_h(x)\}}{\mu_{20}(h)^{1/2}},$$

where

$$\mu_{kl}(h) \equiv h^{-1}\mathbb{E}\left[\left\{K\left(\frac{X_i - x}{h}\right) - \mathbb{E}K\left(\frac{X_i - x}{h}\right)\right\}^k \left\{K\left(\frac{X_i - x}{h}\right)^l - \mathbb{E}K\left(\frac{X_i - x}{h}\right)^l\right\}\right].$$

Remark 5. *Although l of μ_{kl} is always $l = 0$ in this paper, we follow Hall (1991) and use this notation. We need this notation to study the asymptotic structure of the studentised estimator. Since the main purpose of this paper is to investigate the pure effect of bandwidth selection, the studentised KDE does not appear, but for practical purposes we have to consider the effects of studentisation, debias, and bandwidth selection all at the same time, and this notation is intended for that future.*

Assumption 8. $h \to 0, \quad nh/\log n \to \infty \quad as \quad n \to \infty$

Assumption 9 (**Cramér Condition**)**.** *For a sufficiently small h:*

$$\sup_{t\in\mathbb{R}} \left| \int_{-\infty}^{\infty} \exp\{itK(u)\} f(x - uh) du \right| < 1.$$

Remark 6. *Assumption 9 is a high-level condition. Lemma 4.1 in Hall (1991) shows that primitive condition (2.1) in Hall (1991) implies Assumption 9. Moreover, Assumption 9 is weaker than the Cramér condition in Lemma 4.1 of Hall (1991). This is because we only deal with Standardised case, while Hall (1991) also deals with the studentised case.*

Remark 7. *Assumption 9 rules out the uniform kernel, but many kernels which are practically used will satisfy this condition. However, as stated in Hall (1991), one can also derive the Edgeworth expansion in the case of the uniform kernel by routine methods for lattice-valued random variables.*

Proposition 2.2 (Hall 1991, Expansion with a Deterministic Bandwidth)**.** *Under Assumptions 1, 3, 4, 8, and 9, the following expansions are valid:*

$$\sup_{z\in\mathbb{R}} \left| \mathbb{P}(S_h(x) \leqq z) - \Phi(z) - \phi(z)\left[(nh)^{-1/2} p_1(z)\right]\right| = o\{(nh)^{-1/2}\}$$

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}(S_h(x)\leqq z)-\Phi(z)-\phi(z)\left[(nh)^{-1/2}p_1(z)+(nh)^{-1}p_2(z)\right]\right|=o\{(nh)^{-1}\},\tag{3}$$

where $\Phi(z)$ and $\phi(z)$ are the distribution and density functions at z of a standard normal random variable, respectively, and:

$$p_1(z)=-\frac{1}{6}\mu_{20}(h)^{-3/2}\mu_{30}(h)(z^2-1),$$
$$p_2(z)=-\frac{1}{24}\mu_{20}(h)^{-2}\mu_{40}(h)(z^3-3z)-\frac{1}{72}\mu_{20}(h)^{-3}\mu_{30}^2(z^5-10z^3+15z).$$

See Hall (1991) for the proof.

These results are the Edgeworth expansion of KDE up to the order of $O(\{(nh)^{-1/2}\})$ and $O(\{(nh)^{-1}\})$, respectively. However, bandwidth in his results is still deterministic. In this paper, we study KDE with data-driven bandwidth $\hat{f}_{\hat{h}}$ at a point x. The next proposition decomposes the $\hat{f}_{\hat{h}}$ into terms that include the effect of bandwidth selection and one that does not.

Assumption 10. *Kernel function K is a bounded, even function with a compact support, of order $L\geqq 2$ and is twice continuously differentiable.*

Proposition 2.3 (Expansion of KDE with Data-Driven Bandwidth)**.** *Under Assumptions 1, 4, 5, 6, 7, 8, and 10, expanding $\hat{f}_{\hat{h}}(x)$ around $\hat{h}=h_0$ yields:*

$$\begin{aligned}\hat{f}_{\hat{h}}(x)&\equiv\frac{1}{n\hat{h}}\sum_{i=1}^{n}K\left(\frac{X_i-x}{\hat{h}}\right)\\&=\hat{f}_{h_0}(x)-\left(\frac{\hat{h}-h_0}{h_0}\right)\Gamma_{KDE_1}+\frac{1}{2}\left(\frac{\hat{h}-h_0}{h_0}\right)^2\Gamma_{KDE_2}+o_p\left(\left(\frac{\hat{h}-h_0}{h_0}\right)^2\Gamma_{KDE_2}\right),\end{aligned}\tag{4}$$

where

$$\Gamma_{KDE_1}\equiv\frac{1}{nh_0}\sum_{i=1}^{n}\left\{K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)\right\},$$
$$\Gamma_{KDE_2}\equiv\frac{1}{nh_0}\sum_{i=1}^{n}\left\{2K\left(\frac{X_i-x}{h_0}\right)+4K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K''\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)^2\right\}.$$

Let $S_{PI}(x)$ be Standardised version of KDE with global plug-in bandwidth and define $\mu_{kl}=\mu_{kl}(h_0)$. Noting that expanding $\hat{h}^{1/2}$ around $\hat{h}=h_0$ yields $\hat{h}^{1/2}=h_0^{1/2}+\frac{1}{2}h_0^{-1/2}(\hat{h}-h_0)+O_p\{(\hat{h}-h_0)^2h_0^{-3/2}\}$, we have,

$$\begin{aligned}S_{PI}(x)&\equiv\frac{\sqrt{n\hat{h}}\{\hat{f}_{\hat{h}}(x)-\mathbb{E}\hat{f}_{h_0}(x)\}}{\mu_{20}^{1/2}}\\&=S_{h_0}(x)-\frac{\sqrt{nh_0}\left(\frac{\hat{h}-h_0}{h_0}\right)\Gamma_{KDE_1}-\frac{\sqrt{nh_0}}{2}\left(\frac{\hat{h}-h_0}{h_0}\right)^2\Gamma_{KDE_2}}{\mu_{20}^{1/2}}\\&\quad+\frac{1}{2}S_{h_0}(x)\left(\frac{\hat{h}-h_0}{h_0}\right)-\frac{\sqrt{nh_0}\left(\frac{\hat{h}-h_0}{h_0}\right)^2\Gamma_{KDE_1}}{2\mu_{20}^{1/2}}+s.o.\end{aligned}\tag{5}$$

Assumption 11. $\lim_{u\to\pm\infty}|K(u)u|\to 0$

The following theorem generalizes the kernel orders of Theorem 3.2 in Hall and Kang (2001). Their theorem specifically sets the order of the kernels to be $L=2$ and $L_p=6$, and we prove that it holds for general kernel orders L and L_p.

Theorem 2.4 (Second Order Equivalence)**.** *Under Assumptions 1, 4, 6, 7, 9, 10, and 11, the following expansion is valid:*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}(S_{PI}(x)\leqq z)-\Phi(z)-\phi(z)\left[(nh_0)^{-1/2}p_1(z)\right]\right|=o\{(nh_0)^{-1/2}\}. \tag{6}$$

See Appendix A.2 for the proof. Comparing this result with the first half of Proposition 2.2, we see that the bandwidth selection via global plug-in method has no effect on the asymptotic structure of KDE up to the order of $O\{(nh_0)^{-1/2}\}$.

3 Main Results

As stated in Theorem 3.2 in Hall and Kang (2001) or our Theorem 2.4, bandwidth selection via the global plug-in method has no effect on the asymptotic properties of KDE up to the order of $O\{(nh_0)^{-1/2}\}=O(n^{-L/(2L+1)})$. Section 3.1 provides a valid Edgeworth expansion for KDE with plug-in bandwidth up to the order of $O\{(nh_0)^{-1}\}=O(n^{-2L/(2L+1)})$ in Theorem 3.1. This expansion possesses a form comparable with that in Hall (1991). In Section 3.2, we rewrite the expansions in Proposition 2.2 and Theorem 3.1 to derive the expansions only in terms of n and the n-independent coefficient functions without h_0 in Corollary 3.2 and 3.3. Using these results, we scrutinize the higher-order difference between the theoretical and plug-in bandwidth in Section 3.3. We realize that the global plug-in bandwidth selection starts to have an impact from on the order of $O\{(nh_0)^{-1/2}h\}=O(h_0^{L+1})$, which is stated in Theorem 3.4. Section 3.4 provides a comprehensive example by considering the special case of $L=2$.

3.1 Edgeworth Expansion for KDE with Global Plug-In Bandwidth up to the order of $O\{(nh_0)^{-1}\}$

We introduce the following assumption:

Assumption 12. *For* $1\leqq k\leqq L-1$, $\lim_{u\to\pm\infty}|K(u)u^k|\to 0$ *and* $\lim_{u\to\pm\infty}|K'(u)u^2|\to 0$.

We have the following theorem which is proved in Appendix A.3

Theorem 3.1 (Main Result)**.** *Under Assumptions 1, 4, 6, 7, 9, 10 and 12, the following expansion is valid:*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}(S_{PI}(x)\leqq z)-\Phi(z)\right.$$
$$\left.-\phi(z)\left[(nh_0)^{-1/2}p_1(z)+\sum_{l=0}^{L-1}h_0^{L+l+1}p_{3,l}(z)+n^{-1/2}h_0^{1/2}p_4(z)+(nh_0)^{-1}p_2(z)\right]\right|=o\{(nh_0)^{-1}\}, \tag{7}$$

where

$$p_{3,l}(z)=-C_{PI}C_{\Gamma,l}(x)\rho_{11}\mu_{20}^{-1}z,$$
$$p_4(z)=-C_{PI}\rho_{11}\xi_{11}\mu_{20}^{-3/2}(z^2-1)+\frac{1}{2}C_{PI}\rho_{11}\mu_{20}^{-1/2}z^2,$$
$$C_{\Gamma,l}(x)=-\left(\int u^{L+l}K(u)du\right)\frac{f^{(L+l)}(x)}{(L+l-1)!},$$

$$\xi_{kl}=h_0^{-\mathbb{1}(\{k\geqq1\}\,\cup\,\{l\geqq1\})}\mathbb{E}\left[\left\{K\left(\frac{X_i-x}{h_0}\right)-\mathbb{E}K\left(\frac{X_i-x}{h_0}\right)\right\}^k\right.$$

$$\times\left\{K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)-\mathbb{E}[K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)]\right\}^l\Bigg],$$

$$\rho_{kl}=h_0^{-\mathbb{1}(k\geqq 1)}\mathbb{E}\left[\left\{K\left(\frac{X_i-x}{h_0}\right)-\mathbb{E}K\left(\frac{X_i-x}{h_0}\right)\right\}^k\left\{f^{(2L)}(X_i)-\mathbb{E}f^{(2L)}(X_i)\right\}^l\right].$$

3.2 Edgeworth Expansions in Powers of $n^{-1/(2L+1)}$

Note that h_0 satisfies Assumption 8. Comparing (3) with $h=h_0$ and (7), we see that $h_0^{L+l+1}p_{3,l}(z)$ and $n^{-1/2}h_0^{1/2}p_4(z)$ reflect the effect of bandwidth selection via global plug-in methods. However, the results in Proposition 2.2, Theorem 2.4, and Theorem 3.1 are still insufficient for identifying the exact difference because $\mu_{kl},\rho_{kl},\xi_{kl}$, accordingly $p_1(z),p_2(z),p_{3,l}(z)$ and $p_4(z)$ in the expansions depend on h_0 and, consequently, the relationship between the terms in the expansions is not clear.

For $S_{h_0}(x)$, we have to expand $p_1(z)$ and $p_2(z)$ in terms of only n, without h. Expanding $p_2(z)$ is easy because only its leading term affects the Edgeworth expansion up to the order of $O\{(nh_0)^{-1}\}$. For $p_1(z)$, recalling that $p_1(z)=-\frac{1}{6}\mu_{20}^{-3/2}\mu_{30}(z^2-1)$, we have to expand $\mu_{30}\mu_{20}^{-3/2}$ up to the term whose convergence rate is $O(h_0^L)$. Letting $\kappa_{st}\equiv\int u^sK(u)^t du$ and, from straightforward computation, we can expand μ_{20},μ_{30} as follows respectively:

$$\mu_{20}=\kappa_{02}f(x)-f(x)^2h_0+\sum_{l=2}^{L}\kappa_{l2}\frac{f^{(l)}(x)}{l!}h_0^l+o(h_0^L), \tag{8}$$

$$\begin{aligned}\mu_{30}=&\kappa_{03}f(x)-3\kappa_{02}f(x)^2h_0+\left\{\kappa_{23}\frac{f^{(2)}(x)}{2!}+2f(x)^3\right\}h_0^2\\&+\sum_{l=3}^{L}\left\{\kappa_{l3}\frac{f^{(l)}(x)}{l!}-3\kappa_{l-1,2}\frac{f^{(l-1)}(x)}{(l-1)!}f(x)\right\}h_0^l+o(h_0^L).\end{aligned} \tag{9}$$

For notational simplicity, we rewrite μ_{20},μ_{30} as a series of h_0:

$$\mu_{20}\equiv\sum_{l=0}^{L}m_{2,l}(x)h_0^l+o(h_0^L),\qquad \mu_{30}\equiv\sum_{l=0}^{L}m_{3,l}(x)h_0^l+o(h_0^L).$$

Then, expanding $\mu_{30}\mu_{20}^{-3/2}$ yields:

$$\begin{aligned}\mu_{30}\mu_{20}^{-3/2}=&\left\{\sum_{l=0}^{L}m_{3,l}(x)h_0^l+o(h_0^L)\right\}\left\{\sum_{j=0}^{L}m_{2,j}(x)h_0^j+o(h_0^L)\right\}^{-3/2}\\=&\left\{\sum_{l=0}^{L}m_{3,l}(x)h_0^l+o(h_0^L)\right\}\\&\times\Bigg[\{m_{2,0}(x)\}^{-3/2}-\frac{3}{2}\{m_{2,0}(x)\}^{-5/2}\left(\sum_{j=1}^{L}m_{2,j}(x)h_0^j\right)\\&+\frac{15}{4\times 2!}\{m_{2,0}(x)\}^{-7/2}\left(\sum_{l=1}^{L}m_{2,l}(x)h_0^l\right)^2\\&-\frac{105}{8\times 3!}\{m_{2,0}(x)\}^{-9/2}\left(\sum_{j=1}^{L}m_{2,j}(x)h_0^j\right)^3\\&\vdots\end{aligned}$$

$$
\begin{aligned}
&\qquad + \frac{(-1)^L(2L+1)!!}{2^L L!}\{m_{2,0}(x)\}^{\frac{-(2L+3)}{2}} m_{2,L} h_0^L + o(h_0^L)\Bigg] \\
&= \left\{\sum_{l=0}^{L} m_{3,l}(x) h_0^l + o(h_0^L)\right\} \\
&\quad \times \left[\sum_{k=0}^{L} \frac{(-1)^k(2k+1)!!}{2^k k!}\{m_{2,0}(x)\}^{\frac{-(2k+3)}{2}} \left(\sum_{j=1}^{L} m_{2,j} h_0^j\right)^k + o(h_0^L)\right] \\
&= \sum_{l=0}^{L}\sum_{k=0}^{l} \frac{(-1)^k(2k+1)!!}{2^k k!}\Big\{m_{2,0}(x)\Big\}^{\frac{-(2k+3)}{2}} \\
&\quad \times \sum_{k \leqq i_1+\cdots+i_k+l \leqq L} \!\!\cdots \sum m_{3,l}(x) m_{2,i_1}(x) \cdots m_{2,i_k}(x) h_0^{i_1+\cdots i_k+l} + o(h_0^L).
\end{aligned}
$$

We define $\gamma_{1,0}(x), \gamma_{1,1}(x), \gamma_{2,1,0}(x)$ and $\gamma_{2,2,0}(x)$ as follows:

$$
\begin{aligned}
\gamma_{1,0}(x) &\equiv \frac{-1}{6}\kappa_{02}^{-3/2}\kappa_{03} f(x), \\
\gamma_{1,1}(x) &\equiv \frac{1}{2}\Big(\kappa_{02}^{-1/2} f(x)^{1/2} - \kappa_{02}^{-5/2}\kappa_{03} f(x)^{1/2}/2\Big), \\
\gamma_{2,1,0}(x) &\equiv \frac{-1}{24}\kappa_{02}^{-2}\kappa_{04} f(x)^{-1}, \\
\gamma_{2,2,0}(x) &\equiv \frac{-1}{72}\kappa_{02}^{-3}\kappa_{03}^{2} f(x)^{-1}.
\end{aligned}
$$

From the above results, we obtain the following corollary.

Corollary 3.2 (Expansion of Hall (1991) in powers of $n^{-1/(2L+1)}$)**.** *Under Assumptions 1, 3, 4, and 9:*

$$
\sup_{z\in\mathbb{R}} \left| \mathbb{P}(S_{h_0}(x) \leqq z) - \Phi(z) - \phi(z) \sum_{j=0}^{L} a_j(z,x) n^{\frac{-(L+j)}{2L+1}} \right| = o\{(nh_0)^{-1}\},
$$

where the definitions of $a_j(z,x)$ are given as follow for $2 \leqq q \leqq L-1$:

$$
\begin{aligned}
a_0(z,x) &= \gamma_{1,0}(x)(z^2-1), \\
a_1(z,x) &= \gamma_{1,1}(x)(z^2-1), \\
a_q(z,x) &= \sum_{l=0}^{L}\sum_{k=0}^{l} \frac{(-1)^k(2k+1)!!}{2^k k!}\Big\{m_{2,0}(x)\Big\}^{\frac{-(2k+3)}{2}} \\
&\qquad \times \sum_{i_1+\cdots+i_k+l=q} \!\!\cdots \sum m_{3,l}(x) m_{2,i_1}(x) \cdots m_{2,i_k}(x) h_0^{i_1+\cdots i_k+l}(z^2-1), \\
a_L(z,x) &= \gamma_{2,0,1}(x)(z^3-3z) + \gamma_{2,0,2}(x)(z^5-10z+15) \\
&\quad + \sum_{l=0}^{L}\sum_{k=0}^{l} \frac{(-1)^k(2k+1)!!}{2^k k!}\Big\{m_{2,0}(x)\Big\}^{\frac{-(2k+3)}{2}} \\
&\qquad \times \sum_{i_1+\cdots+i_k+l=L} \!\!\cdots \sum m_{3,l}(x) m_{2,i_1}(x) \cdots m_{2,i_k}(x) h_0^{i_1+\cdots i_k+l}(z^2-1).
\end{aligned}
$$

Remark 8. *Note that a_0 and a_1 are special cases of a_q, but we explicitly write these terms for comparison of this result with the next corollary.*

Remark 9. *From this corollary, we can identify that, the Edgeworth expansion of the standardized KDE with deterministic bandwidth has the term of order $O\{(nh_0)^{-1/2}\} = O(n^{\frac{-L}{2L+1}})$ right after term $\Phi(z)$, with a gap between them, but the subsequent terms decrease at the rate of $O(h_0) = O(n^{\frac{-1}{2L+1}})$, which is not clear in Hall (1991).*

Next, for (7), we also have to expand $p_{3,l}(z)$ and $p_4(z)$. Although we do not provide the details here, one can use the similar process to for $p_1(z)$. We define:

$$\begin{aligned}
\tau_l &\equiv \int u^l\{K(u)K'(u)u + K(u)^2\}du,\\
\mathcal{L}(x) &\equiv f^{(2L)}(x) - \mathbb{E}[f^{(2L)}(x)],\\
\gamma_{3,1,0}(x) &\equiv -C_{PI}C_{\Gamma,0}(x)\kappa_{02}^{-1}\mathcal{L}(x),\\
\gamma_{3,1,1}(x) &\equiv C_{PI}C_{\Gamma,0}(x)\kappa_{02}^{-2}\mathcal{L}(x)f(x)\\
\gamma_{4,1,0}(x) &\equiv -C_{PI}\kappa_{02}^{-3/2}\tau_0\mathcal{L}(x)f^{1/2}(x),\\
\gamma_{4,1,1}(x) &\equiv \frac{3}{2}C_{PI}\kappa_{02}^{-5/2}\tau_0\mathcal{L}(x)f(x)^{3/2}.\\
\gamma_{4,2,0}(x) &\equiv \frac{1}{2}C_{PI}\kappa_{02}^{-1/2}\mathcal{L}(x)f(x)^{1/2}\\
\gamma_{4,2,1}(x) &\equiv \frac{-1}{4}C_{PI}\kappa_{02}^{-3/2}\mathcal{L}(x)f(x)^{3/2}
\end{aligned}$$

Then, we have next corollary.

Corollary 3.3 (Main Theorem in powers of $n^{-1/(2L+1)}$)**.** *Under the same assumptions as in Theorem 3.1:*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}(S_{PI}(x) \leqq z) - \Phi(z) - \phi(z)\sum_{j=0}^{L} b_j(z,x)n^{\frac{-(L+j)}{2L+1}}\right| = o\{(nh_0)^{-1}\},$$

where

$$\begin{aligned}
b_0(z,x) &= a_0(z,x),\\
b_1(z,x) &= a_1(z,x) + \gamma_{3,1,0}(x)z + \gamma_{4,1,0}(x)(z^2-1) + \gamma_{4,2,0}(x)z^2.\\
b_2(z,x) &= a_2(z,x) + \gamma_{3,1,1}(x)z + \gamma_{4,1,1}(x)(z^2-1) + \gamma_{4,2,1}(x)z^2.
\end{aligned}$$

We do not provide here the definitions of $b_j(z,x)$, $j = 3,...$ because they are too lengthy and tedious, but they can be straightforwardly obtained.

Hall and Kang (2001) and Theorem 2.4 state that the global plug-in method has no effect on the terms up to whose convergence rates are $O\{(nh_0)^{-1/2}\}$; in other words, $b_0(z,x)$ does not include the effect of bandwidth selection in view of Corollary 3.3. Comparing $a_1(z,x)$ and $b_1(z,x)$, the bandwidth selection via the global plug-in method starts to have an effect on the term with the order of $O\{(nh_0)^{-1/2}h_0\} = O(n^{\frac{-(L+1)}{2L+1}})$. The deviation between $b_0(z,x)$ (the smallest term not affected by bandwidth selection) and $b_1(z,x)$ (the largest term affected by bandwidth selection) is only $O(h_0) = O(n^{-1/(2L+1)})$.

Remark 10. *Although we omit $b_3(z,x)$ and the subsequent terms, we can show that these terms are also affected by bandwidth selection via the global plug-in method in the same way as the process of deriving Corollary 3.2. However, the most important point is that the influence of the bandwidth selection via the global plug-in method starts to appear at $b_1(z,x)$.*

3.3 Difference between $S_{h_0}(x)$ and $S_{PI}(x)$

From Corollaries 3.2 and 3.3, we can easily deduce the following theorem, which states the exact order of the difference between $S_{h_0}(x)$ and $S_{PI}(x)$. See Appendix A.4 for the proof.

Theorem 3.4 (Exact Evaluation of the Deviation)**.** *Under the same assumptions as in Theorem 3.1:*

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}(S_{h_0}(x)\leqq z)-\mathbb{P}(S_{PI}(x)\leqq z)-\phi(z)\left[\{\gamma_{3,1,0}(x)z+\gamma_{4,0}(x)(z^2-1)\}n^{\frac{-(L+1)}{2L+1}}\right]\right|=O(n^{\frac{-(L+1)}{2L+1}}),$$

and the order is exact.

This theorem implies that:

$$\sup_{z\in\mathbb{R}}|\mathbb{P}(S_{h_0}(x)\leqq z)-\mathbb{P}(S_{PI}(x)\leqq z)|=O(n^{\frac{-(L+1)}{2L+1}}).$$

We can only claim that this deviation is $o\{(nh_0)^{-1/2}\}=o(n^{-L/(2L+1)})$ from Theorem 3.2 in Hall and Kang (2001) and our Theorem 2.4, whereas Theorem 3.4 gives a stronger result, stating that the convergence rate is exactly $O\{(nh_0)^{-1/2}h_0\}=O(n^{-(L+1)/(2L+1)})$.

Remark 11. *The larger is the kernel order L we use, the slower the convergence rate of the approximation in Theorems 2.4, 3.1, and 3.4 will be. This is because we centralize at $\mathbb{E}\hat{f}_{h_0}(x)$. However, as stated in Section 5, one of the final goals will be to examine the effect of bandwidth selection and 'debias' simultaneously (we are in the process of working on it), and it is not clear that the second-order kernel $L=2$ is optimal.*

3.4 Special Case

Since the previous results are a difficult to interpret because of their generality, we consider a special case of $L=2$. Here, we also provide the details of the expansions of $p_{3,l}(z)$ and $p_4(z)$ as well as that of $p_1(z)$.

First, we have to expand $p_1(z)$ and $p_2(z)$. From (8) and (9), we can expand $p_1(z)$ as follows (see Appendix C):

$$\begin{aligned}
p_1(z)&=\frac{-1}{6}\mu_{30}\mu_{20}^{-3/2}(z^2-1)\\
&=\frac{-1}{6}\Bigg[\kappa_{02}^{-3/2}\kappa_{03}f(x)-3\left\{\frac{f(x)^{1/2}}{\kappa_{02}^{1/2}}-\frac{\kappa_{03}f(x)^{1/2}}{2\kappa_{02}^{5/2}}\right\}h_0\\
&\qquad+\Bigg\{\frac{-3}{4}\{\kappa_{02}f(x)\}^{-5/2}\kappa_{03}\kappa_{23}f^{(2)}(x)f(x)-3\{\kappa_{02}f(x)\}^{-5/2}\kappa_{03}f(x)^4\\
&\qquad\quad+\frac{15}{8}\{\kappa_{02}f(x)\}^{-7/2}\kappa_{03}f(x)^5+\frac{9}{2}\kappa_{02}^{-3/2}f(x)^{3/2}\Bigg\}h_0^2\Bigg](z^2-1)+o(h_0^2),
\end{aligned}$$

and since for $p_2(z)$ we only need the leading term; a straightforward computation yields:

$$p_2(z)=\frac{-1}{24}\kappa_{02}^{-2}\kappa_{04}f(x)^{-1}(z^3-3z)-\frac{1}{72}\kappa_{02}^{-3}\kappa_{03}^2f(x)^{-1}(z^5-10z^3+15)+o(1).$$

From the above results, in the special case $L=2$, expansion (3) is as follows:

$$\begin{aligned}
&\sup_{z\in\mathbb{R}}\left|\mathbb{P}(S(x)\leqq z)-\Phi(z)-\phi(z)\left[a_0(z,x)n^{-2/5}+a_1(z,x)n^{-3/5}+a_2(z,x)n^{-4/5}\right]\right|\\
&=o\{(nh_0)^{-1}\},
\end{aligned}$$

where

$$\begin{aligned}
a_0(z,x)&=\gamma_{1,0}(x)(z^2-1),\\
a_1(z,x)&=\gamma_{1,1}(x)(z^2-1),\\
a_2(z,x)&=\gamma_{1,2}(x)(z^2-1)+\gamma_{2,1,0}(x)(z^3-3z)+\gamma_{2,2,0}(x)(z^5-10z^3+15).
\end{aligned}$$

Next, we have to expand $p_{3,0}(z), p_{3,1}(z)$ and $p_4(z)$. From a straightforward computation, noting $\tau_1 = 0$ from the properties of the odd function, we can expand ρ_{11} and ξ_{11} as follows:

$$\rho_{11} = \mathcal{L}(x)f(x) + O(h_0^L),$$
$$\xi_{11} = \tau_0 f(x) + \tau_1 f^{(1)}(x)h_0 + o(h_0) = \tau_0 + o(h_0).$$

These imply:

$$\begin{aligned} p_{3,0}(z) &= -C_{PI}C_{\Gamma,0}(x)\rho_{11}\mu_{20}^{-1}z \\ &= -C_{PI}C_{\Gamma,0}(x)\kappa_{02}^{-1}\mathcal{L}(x)z + C_{PI}C_{\Gamma,0}(x)\kappa_{02}^{-2}\mathcal{L}(x)f(x)zh_0 + o(h_0). \end{aligned}$$

See Appendix C for the second equality. Noting that $C_{\Gamma,1}(x) = 0$ from the properties of the odd function:

$$p_{3,1}(z) = -C_{PI}C_{\Gamma,1}(x)\rho_{11}\mu_{20}^{-1}z = 0,$$

and, as shown in Appendix C:

$$\begin{aligned} p_4(z) &= -C_{PI}\rho_{11}\xi_{11}\mu_{20}^{-3/2}(z^2-1) + \frac{1}{2}C_{PI}\rho_{11}\mu_{20}^{-1/2}z^2 \\ &= -C_{PI}\kappa_{02}^{-3/2}\tau_0\mathcal{L}(x)f(x)^{1/2}(z^2-1) + \frac{3}{2}C_{PI}\kappa_{02}^{-5/2}\tau_0\mathcal{L}(x)f(x)^{3/2}(z^2-1)h_0 + o(h_0) \\ &\quad + \frac{1}{2}C_{PI}\kappa_{02}^{-1/2}\mathcal{L}(x)f(x)^{1/2}z^2 - \frac{1}{4}C_{PI}\kappa_{02}^{-3/2}\mathcal{L}(x)f(x)^{3/2}z^2h_0 + o(h_0). \end{aligned}$$

From the above results, in the special case of $L = 2$, the expansion (7) is as follows.

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}(S_{PI}(x) \leqq z) - \Phi(z) - \phi(z)\left[b_0(z,x)n^{-2/5} + b_1(z,x)n^{-3/5} + b_2(z,x)n^{-4/5}\right]\right| = o\{(nh_0)^{-1}\},$$

where the definitions of $b_0(z,x)$, $b_1(z,x)$ are given as follow.

$$\begin{aligned} b_0(z,x) &= a_0(z,x) \\ b_1(z,x) &= a_1(z,x) + \gamma_{3,1,0}(x)z + \gamma_{4,1,0}(x)(z^2-1) + \gamma_{4,2,0}(x)z^2 \\ b_2(z,x) &= a_2(z,x) + \gamma_{3,1,1}(x)z + \gamma_{4,1,1}(x)(z^2-1) + \gamma_{4,2,1}(x)z^2 \end{aligned}$$

4 Simulation Study

In order to examine the higher order improvements by the Edgeworth expansions, We compare the coverage accuracies of the normal approximation, the expansion with deterministic bandwidth, and the expansion with plug-in bandwidth. The data generating process is an exponential distribution whose density is $f(x) = xe^{-x}$. We use the following kernel functions:

$$K(u) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}, \quad H(u) = \frac{1}{8\sqrt{2\pi}}(u^4 - 10u^2 + 15)e^{\frac{-x^2}{2}},$$

namely $L = 2$ and $L_p = 6$.

Let z_α, w_α and w_α^{PI} be the 100α%-quantile point of normal distribution, Cornish-Fisher expansion of KDE with deterministic bandwidth, and Cornish-Fisher expansion of the KDE with plug-in bandwidth, respectively. In this experiment, we set $\alpha = 0.95$. We construct the following intervals for 10000 iterations and count the number of

Table 1: $x = 1.0, b = 0.9$

	$n = 50$	$n = 100$	n=400
$N(0,1)$	**0.961**	**0.963**	**0.972**
Hall (1991)	0.963	0.966	0.974
Our Result	0.962	0.965	0.973

Table 2: $x = 1.0, b =$MSE optimal

	$n = 50$	$n = 100$	$n = 400$
N(0,1)	**0.955**	**0.958**	**0.968**
Hall (1991)	0.957	0.960	0.970
Our Result	0.956	**0.958**	0.969

intervals that include $\mathbb{E}\hat{f}_h(x)$. We evaluate the performance of each approximation by the closeness of the number of such intervals divided by 10000 (iteration) to 0.9500:

$$I_N = \left[\hat{f}_{\hat{h}_{PI}} - \frac{z_\alpha \mu_{20}^{1/2}}{\sqrt{n\hat{h}_{PI}}}, \infty\right], I_D = \left[\hat{f}_{\hat{h}_{PI}} - \frac{w_\alpha \mu_{20}}{\sqrt{n\hat{h}_{PI}}}, \infty\right], I_{PI} = \left[\hat{f}_{\hat{h}_{PI}} - \frac{w_\alpha^{PI} \mu_{20}}{\sqrt{n\hat{h}_{PI}}}, \infty\right].$$

We choose $x = 0.5, 1.0$, and 5.0 as the evaluation points of f. Considering the shape of $f(x) = xe^{-x}$, we can regard $x = 1.0$ as a representative for data-rich points, $x = 0.5$ for points with moderate amount of data, and $x = 5.0$ for points with poor data. The experiment is conducted with pilot bandwidth $b = 0.6, 0.9$ and b_0, the MSE-optimal one, and for sample sizes $n = 50, 100$, and 400. The MSE-optimal pilot bandwidth is defined as follows:

$$b_0 \equiv \left(\frac{(8L+1)\left\{\int f(x)^2 dx\right\}^2}{L_p(L_p!)^{-2}\left\{\int u^{L_p} H(u)du\right\}^2 \left\{\int f(x)^{2L}(x) f^{2L+L_p}(x)dx\right\}^2} \right) n^{-2/(8L+2L_p+1)}.$$

In this simulation setting $b_0 = (0.8113, 0.7735, 0.7029)$ for sample sizes $n = (50, 100, 400)$, respectively. Tables 1-9 show the results of Monte Carlo simulation, where the bold figures indicate the most accurate approximation in closeness to 0.950.

Tables 1–3 give coverage probabilities at $x = 1.0$ with pilot bandwidths of 0.9, b_0, and 0.6 respectively. All three approximations work well because all the figures are mostly close to the nominal value of 0.950. The normal approximation might look slightly better with relatively large and MSE-optimal pilot bandwidths, though the differences are extremely small. For relatively small pilot bandwidth, it is difficult to say which approximation is best.

Results at $x = 0.5$ are reported in Tables 4–6 for pilot bandwidths of 0.9, b_0, and 0.6 respectively. When the sample size is relatively large, our result seems the best. However, for a small sample size, Hall (1991) dominates our results. In any case the differences are quite small.

At $x = 5.0$ (see Tables 7–9), for all sample sizes and for all pilot bandwidths, Hall (1991) and our result are virtually the same and outperform the normal approximation, which gives poor performances.

In summary, at a point with abundant data, the normal approximation provides a good approximation for relatively large sample size. At a point with a moderate amount of data, our approximation provides the best performance with relatively large sample size, while the other approximation provides better performance with small sample size. At a point with poor data, the difference of Hall (1991) and our result is marginal.

Table 3: $x = 1.0, b = 0.6$

	$n = 50$	$n = 100$	$n = 400$
$N(0,1)$	0.946	0.949	**0.959**
Hall (1991)	**0.949**	0.952	0.960
Our Result	0.947	**0.950**	0.960

Table 4: $x = 0.5, b = 0.9$

	$n = 50$	$n = 100$	$n = 400$
$N(0,1)$	**0.949**	0.966	0.981
Hall (1991)	0.954	0.968	0.981
Our Result	0.938	**0.962**	**0.980**

Table 5: $x = 0.5, b =$MSE optimal

	$n = 50$	$n = 100$	$n = 400$
$N(0,1)$	0.941	0.961	0.976
Hall (1991)	**0.946**	0.963	0.976
Our Result	0.932	**0.956**	**0.975**

Table 6: $x = 0.5, pb = 0.6$

	$n = 50$	$n = 100$	$n = 400$
$N(0,1)$	0.925	0.940	0.955
Hall (1991)	**0.931**	**0.943**	0.957
Our Result	0.914	0.935	**0.954**

Table 7: $x = 5.0, b = 0.9$

	$n = 50$	$n = 100$	$n = 400$
$N(0,1)$	0.922	0.927	0.938
Hall (1991)	**0.942**	**0.942**	**0.947**
Our Result	0.941	**0.942**	**0.947**

Table 8: $x = 5.0, b =$MSE optimal

	$n = 50$	$n = 100$	$n = 400$
$N(0,1)$	0.924	0.926	0.939
Hall (1991)	**0.945**	**0.942**	**0.948**
Our Result	**0.945**	0.941	0.947

Table 9: $x = 5.0, b = 0.6$

	$n = 50$	$n = 100$	$n = 400$
$N(0,1)$	0.932	0.927	0.940
Hall (1991)	**0.949**	**0.943**	**0.948**
Our Result	0.948	0.942	**0.948**

5 Discussion and Conclusions

This study investigated the higher-order asymptotic properties of KDE with global plug-in bandwidth. As a first contribution, we provided the Edgeworth expansion of KDE with global plug-in bandwidth up to the order of $O\{(nh_0)^{-1}\} = O(n^{\frac{-2L}{2L+1}})$ and show that the bandwidth selection by the global plug-in method starts to have an effect from on the term whose convergence rate is $O\{(nh_0)^{-1/2}h_0\} = O(n^{\frac{-(L+1)}{2L+1}})$. As the second contribution, we generalized Theorem 3.2 in Hall and Kang (2001), which states that bandwidth selection via the global plug-in method has no effect on the asymptotic structure of KDE up to the order of $O\{(nh_0)^{-1/2}\} = O(n^{\frac{-L}{2L+1}})$. Their results limited the order of kernel functions $K(u)$ and $H(u)$ to $L = 2, L_p = 6$ respectively, but we show that they are valid for general orders L and L_p as well. The third contribution is that we explored the Edgeworth expansion of KDE with deterministic bandwidth in more detail than Hall (1991). The Edgeworth expansion of the Standardized KDE with deterministic bandwidth has the term of order $O\{(nh_0)^{-1/2}\} = O(n^{\frac{-L}{2L+1}})$ right after term $\Phi(z)$, with a gap between them, but after that, the terms decrease at the rate of $O(h_0) = O(n^{\frac{-1}{2L+1}})$.

As stated in Remark 1, centring at $\mathbb{E}\hat{f}_h(x)$ leaves asymptotic bias under standard conditions. Two standard methods to deal with asymptotic bias (*debias*) are 'undersmoothing' and 'explicit bias reduction'. The former refers to choosing the bandwidth satisfying $\sqrt{nh}h^L \to 0$ and the latter directly estimates and removes the bias term. Hall (1992) examined the effect of undersmoothing and explicit bias reduction on the asymptotic structure via the Edgeworth expansion up to the order of $O\{(nh)^{-1}\}$ and stated that undersmoothing provides better coverage than explicit bias correction does. After that, Calonico, Cattaneo and Farrell (2018) have proposed alternative bias correction method and show thier method is comparable with undersmoothing by Edgeworth expansion up to the order of $O\{(nh)^{-1}\}$. However the bandwidth in their expansion is still deterministic. We can interpret that Hall and Kang (2001), this paper, and Hall (1992), Calonico, Cattaneo, and Farrell (2018) studied these effects separately, that is, the pure effect of bandwidth selection and the pure effect of debias respectively. A goal for future research will be to investigate the effect of bandwidth selection and debias simultaneously, on which we are working at the moment.

Among the recent topics in which the density estimator plays an important role is the manipulation test of regression discontinuity designs (RDD). Cattaneo, Jansson, and Ma (2020) proposed a local polynomial density estimator for adaptability at or near the boundary points. We expect that the asymptotic structure of their estimator with the corresponding plug-in bandwidth has a similar structure to that of the KDE provided in this paper.

One of the other possible extensions of this work is considering the pilot bandwidth, which are eliminated by using a large L_p in this paper, as a general n-dependent sequence b_n, and to study its influence on the asymptotic structure. Another extension, which we are in the process of working on, is investigating the effect of cross-validation on the asymptotic structure.

Acknowledgement

This work was supported by JSPS KAKENHI, Grant Number 19H01473, and the Joint Usage and Research Project of Institute of Economic Research, Kyoto University. The authors are grateful to Takahide Yanagi for the useful discussions. We would also Like to thank Daisuke Kurisu, Takuya Ishihara, and the participants at the Kansai Econometric Meeting held at Tohoku University in January 2022 for their helpful comments.

References

[1] Armstrong, T. B., Kolesár, M. (2018). Optimal Inference in a Class of Regression Models. *Econometrica, 86*, 655–83.

[2] Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference. *Journal of American Statistical Association, 113*, 767–79.

[3] Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs. *Econometrics Journal, 23*, 192–210.

[4] Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2022). Coverage Error Optimal Confidence Intervals for Local Polynomial Regression. *Bernoulli, forthcoming*.

[5] Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica, 82*, 2295–326.

[6] Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple Local Polynomial Density Estimators. *Journal of American Statistical Association, 115*, 1449–55.

[7] Hall, P. (1991). Edgeworth Expansions for Nonparametric Density Estimators, with Applications. *Statistics, 22*, 215–32.

[8] Hall, P. (1992). Effect of Bias Estimation on Coverage Accuracy of Bootstrap Confidence Intervals for a Probability Density. *Annals of Statistics, 20*, 675–94.

[9] Hall, P., and Horowitz, J. L. (2013). A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions. *The Annals of Statistics, 41*, 1892–1921.

[10] Hall, P., and Kang, K. H. (2001). Bootstrapping Nonparametric Density Estimators with Empirically Chosen Bandwidth. *The Annals of Statistics, 29*, 1443–68.

[11] Hall, P., and Marron, J. S. (1987). Estimation of Integrated Squared Density Derivatives. *Statistics and Probability Letters, 6*, 109–15.

[12] Hall, P., Sheather, S. J., Jones, M. C., and Marron, J. S. (1991). On Optimal Data-Based Bandwidth Selection in Kernel Density Estimation. *Biometrika, 78*, 263–70.

[13] Ichimura, H. (2000). Asymptotic Distribution of Nonparametric and Semiparametric Estimators with Data Dependent Smoothing Parameters. *Unpublished manuscript*.

[14] Li, D. and Li, Q. (2010). Nonparametric/Semiparametric Estimation and Testing of Econometric Models with Data Dependent Smoothing Parameters. *Journal of Econometrics, 157*, 179–90.

[15] Li, Q. and Racine, J. (2007). *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

[16] Nishiyama, Y., and Robinson, P. M. (2000). Edgeworth Expansions for Semiparametric Averaged Derivatives. *Econometrica, 68*, 931–79.

[17] Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics, 33*, 1065–76.

[18] Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric Estimation of Index Coefficients, *Econometrica, 57*, 1403–30.

[19] Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics, 27*, 832–37.

[20] Schennach, S. M. (2020). A Bias Bound Approach to Nonparametric Inference. *The Review of Economic Studies, 87*, 2439–72

[21] Sheather, S. J., and Jones, M. C. (1991). 'A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation', *Journal of Royal Statistical Society Series B, 53*, 683–90.

[22] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Vol. 26, CRC press

[23] Wasserman, L. (2006). *All of Nonparametric Statistics*, Springer.

A Proofs of Results

A.1 Proof of Proposition 2.1

Proof. Recall

$$\hat{I}_L = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} b^{-(2L+1)} H^{(2L)}\left(\frac{X_i - X_j}{b}\right) \equiv \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \hat{I}_{Lij}.$$

Since $\hat{I}_L$ has a U-statistic form, we can use Hoeffding-Decomposition,

$$\hat{I}_L = \mathbb{E}\hat{I}_{Lij} + \frac{2}{n} \sum_{i=1}^{n} \left\{\hat{I}_{Li} - \mathbb{E}\hat{I}_{Lij}\right\} + \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\{\hat{I}_{Lij} - \hat{I}_{Li} - \hat{I}_{Lj} + \mathbb{E}\hat{I}_{Lij}\right\}, \tag{10}$$

where $\hat{I}_{Li} = \mathbb{E}[\hat{I}_{Lij}|X_i]$. In order to examine $\hat{I}_L$, we have to compute $\mathbb{E}\hat{I}_{Lij}$ and $\hat{I}_{Li}$.

$$\begin{aligned}
\hat{I}_{Li} = \mathbb{E}\left[\hat{I}_{Lij}|X_i\right] &= \int \frac{1}{b^{2L+1}} H^{(2L)}\left(\frac{X_i - x}{b}\right) f(x)dx \\
&= \int \frac{1}{b^{2L}} H^{(2L)}(u) f(X_i + ub)du \\
&= \int H(u) f^{(2L)}(X_i + ub)du \\
&= f^{(2L)}(X_i) + \frac{b^{L_p}}{L_p!}\left(\int u^{L_p} H(u)du\right) f^{(2L+L_p)}(X_i) + o_p(b^{L_p}),
\end{aligned} \tag{11}$$

where the third equality follows from integration by part and Assumptions 6 and 7 and the fourth equality follows from the expansion of $f^{(2L)}(X_i + ub)$ around $b = 0$ and Assumptions 6, 7. This implies

$$\mathbb{E}\hat{I}_{Lij} = \mathbb{E}[f^{(2L)}(X_1)] + \frac{b^{L_p}}{L_p!}\left(\int u^{L_p} H(u)du\right)\mathbb{E}[f^{(2L+L_p)}(X_1)] + o(b^{L_p}) \tag{12}$$

From integration by parts and Assumption 6, the first term of the right-hand side of (12) is

$$\mathbb{E}[f^{(2L)}(X_1)] = \int f^{(2L)}(x) f(x)dx = \int f^{(L)}(x)^2 dx = I_L, \tag{13}$$

Inserting (11), (12) and (13) into (10), we have

$$\hat{I}_L = I_L + \frac{2}{n}\sum_{i=1}^{n}\{f^{(2L)}(X_i) - \mathbb{E}f^{(2L)}(X_i)\}$$
$$+ \frac{2}{n}\left(\int u^{L_p}H(u)du\right)\frac{b^{L_p}}{L_p!}\sum_{i=1}^{n}\{f^{(2L+L_p)}(X_i) - \mathbb{E}f^{(2L+L_p)}(X_i)\}$$
$$+ o_p(n^{-1/2}b^{L_p}) + O_p(n^{-1}b^{-(4L-1)/2}). \tag{14}$$

Recall that Plug-In bandwidth is defined as follows,

$$\hat{h} = \left(\frac{R(K)}{2LC_L^2\hat{I}_L}\right)^{\frac{1}{2L+1}} n^{-\frac{1}{2L+1}}. \tag{15}$$

We evaluate the difference between $\hat{h}$ and h_0 using (14).

$$\hat{I}_L^{\frac{-1}{2L+1}} = I_L^{\frac{-1}{2L+1}} - \frac{1}{2L+1}I_L^{\frac{-1}{2L+1}-1}\left[\frac{2}{n}\sum_{i=1}^{n}\{f^{(2L)}(X_i) - \mathbb{E}f^{(2L)}(X_i)\}\right.$$
$$+ \frac{2}{n}\left(\int u^{L_p}H(u)du\right)\frac{b^{L_p}}{L_p!}\sum_{i=1}^{n}\{f^{(2L+L_p)}(X_i) - \mathbb{E}f^{(2L+L_p)}(X_i)\} + o_p(n^{-1/2}b^{L_p}) \tag{16}$$
$$\left. + \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\{\hat{I}_{Lij} - \hat{I}_{Li} - \hat{I}_{Lj} + \mathbb{E}\hat{I}_{Lij}\right\}\right].$$

Inserting this expansion into (15) yields

$$\hat{h} = h_0 - \frac{h_0}{2L+1}I_L^{-1}\left[\frac{2}{n}\sum_{i=1}^{n}\{f^{(2L)}(X_i) - \mathbb{E}f^{(2L)}(X_i)\}\right.$$
$$+ \frac{2}{n}\left(\int u^{L_p}H(u)du\right)\frac{b^{L_p}}{L_p!}\sum_{i=1}^{n}\{f^{(2L+L_p)}(X_i) - \mathbb{E}f^{(2L+L_p)}(X_i)\} + o_p(n^{-1/2}b^{L_p})$$
$$\left. + \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\{\hat{I}_{Lij} - \hat{I}_{Li} - \hat{I}_{Lj} + \mathbb{E}\hat{I}_{Lij}\right\}\right].$$

This implies

$$\frac{\hat{h} - h_0}{h_0} = -\frac{1}{2L+1}I_L^{-1}\left[\frac{2}{n}\sum_{i=1}^{n}\{f^{(2L)}(X_i) - \mathbb{E}f^{(2L)}(X_i)\}\right.$$
$$+ \frac{2}{n}\left(\int u^{L_p}H(u)du\right)\frac{b^{L_p}}{L_p!}\sum_{i=1}^{n}\{f^{(2L+L_p)}(X_i) - \mathbb{E}f^{(2L+L_p)}(X_i)\} + o_p(n^{-1/2}b^{L_p})$$
$$\left. + \binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\{\hat{I}_{Lij} - \hat{I}_{Li} - \hat{I}_{Lj} + \mathbb{E}\hat{I}_{Lij}\right\}\right]$$
$$\Longrightarrow \frac{\hat{h} - h_0}{h_0} = \frac{-C_{PI}}{n}\sum_{i=1}^{n}\left\{f^{(2L)}(X_i) - \mathbb{E}f^{(2L)}(X_i)\right\}$$
$$+ \frac{Cb^{L_p}}{n}\sum_{i=1}^{n}\left\{f^{(2L+L_p)}(X_i) - \mathbb{E}f^{(2L+L_p)}(X_i)\right\} + o_p(n^{-1/2}b^{L_p})$$
$$- \frac{C_{PI}}{2}\binom{n}{2}^{-1}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\{\hat{I}_{Lij} - \hat{I}_{Li} - \hat{I}_{Lj} + \mathbb{E}\hat{I}_{Lij}\right\}.$$

□

A.2 Proof of Theorem 2.1

Proof. Since h_0 satisfies Assumption 8, Proposition 2.1 holds with $h = h_0$. In view of (2.5), if the following evaluation is correct,

$$\mathbb{E}\left|\sqrt{nh}\left(\frac{\hat{h}-h_0}{h_0}\right)\Gamma_{KDE_1}\right| = o\{(nh_0)^{-1/2}\}, \quad \mathbb{E}\left|S_{h_0}(x)\left(\frac{\hat{h}-h_0}{h_0}\right)\right| = o\{(nh_0)^{-1/2}\}$$

then bandwidth selection has no effect on the asymptotic structure up to the order of $O\{(nh_0)^{-1/2}\}$. From Cauchy-Schwarz Inequality

$$\mathbb{E}\left|\sqrt{nh_0}\left(\frac{\hat{h}-h_0}{h_0}\right)\Gamma_{KDE_1}\right| \leqq \sqrt{nh_0}\left\{\mathbb{E}\left|\frac{\hat{h}-h_0}{h_0}\right|^2 \mathbb{E}\left|\Gamma_{KDE_1}\right|^2\right\}^{1/2}$$

Straightforward calculation gives

$$\mathbb{E}\left|\frac{\hat{h}-h_0}{h_0}\right|^2 = O(n^{-1})$$

Next, we evaluate $\mathbb{E}|\Gamma_{KDE_1}|^2$.

$$\begin{aligned}
\mathbb{E}\left|\Gamma_{KDE_1}\right|^2 &= \frac{1}{(nh_0)^2}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\{K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)\right\}\left\{K'\left(\frac{X_j-x}{h_0}\right)\left(\frac{X_j-x}{h_0}\right)+K\left(\frac{X_j-x}{h_0}\right)\right\}\right] \\
&\quad + \frac{1}{(nh_0)^2}\mathbb{E}\left[\sum_{i=1}^{n}\left\{K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)\right\}^2\right] \\
&= \frac{1}{h_0^2}\mathbb{E}\left[\left\{K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)\right\}\left\{K'\left(\frac{X_j-x}{h_0}\right)\left(\frac{X_j-x}{h_0}\right)+K\left(\frac{X_j-x}{h_0}\right)\right\}\right]+O\{(nh_0)^{-1}\} \\
&= \frac{1}{h_0^2}\left(\int\left\{K'\left(\frac{z_1-x}{h_0}\right)\left(\frac{z_1-x}{h_0}\right)+K\left(\frac{z_1-x}{h_0}\right)\right\}f(z_1)dz_1\right)^2+O\{(nh_0)^{-1}\} \\
&= \left(\int K'(u)uf(x+uh_0)du+\int K(u)f(x+uh_0)du\right)^2+O\{(nh_0)^{-1}\} \\
&= \left(-\int K(u)f(x+uh_0)du-\int K(u)uf'(x+uh_0)h_0du+\int K(u)f(x+uh_0)du\right)^2+O\{(nh_0)^{-1}\} \\
&= \left(-\int K(u)uf'(x+uh_0)h_0du\right)^2+O\{(nh_0)^{-1}) \\
&= O(h_0^{2L})+O\{(nh_0)^{-1}\}.
\end{aligned}$$

The fifth equality follows from integration by part of the first term and Assumption 11, and the final equality follows from the expansion of $f'(x+uh_0)$ around $h_0 = 0$ and Assumption 6,10. Therefore form Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}\left|\sqrt{nh_0}\left(\frac{\hat{h}-h_0}{h_0}\right)\Gamma_{KDE}\right| &\leqq \sqrt{nh_0}\left\{\mathbb{E}\left|\frac{\hat{h}-h_0}{h_0}\right|^2\mathbb{E}\left|\Gamma_{KDE}\right|^2\right\}^{1/2} \\
&= O(n^{1/2}h_0^{1/2})\Big(O(n^{-1})O(h_0^{2L}+(nh_0)^{-1})\Big)^{1/2} = O(h_0^{L+\frac{1}{2}}+n^{-1/2}) = o\{(nh_0)^{-1/2}\}.
\end{aligned}$$

Similar to above evaluation, Cauchy-Schwarz inequality gives $\mathbb{E}\left|S(x)\left(\frac{\hat{h}-h_0}{h_0}\right)\right| = O(n^{-1/2}) = o\{(nh_0)^{-1/2}\}$. Therefore bandwidth selection via Plug-In Method has no effect on the asymptotic structure up to the order of $O\{(nh_0)^{-1/2}\}$. □

A.3 Proof of Theorem 3.1

Proof. From Proposition 2.3 and Lemma 3, we have,

$$\sqrt{n\hat{h}}\Big(\hat{f}_{\hat{h}}(x)-\mathbb{E}\hat{f}_{h_0}(x)\Big)=\sqrt{nh_0}\Big(\hat{f}_{h_0}(x)-\mathbb{E}\hat{f}_{h_0}(x)\Big)-\sqrt{nh_0}\left(\frac{\hat{h}-h_0}{h_0}\right)\Gamma_{KDE_1}+\frac{1}{2}S_{h_0}(x)\left(\frac{\hat{h}-h_0}{h_0}\right)+o_p\{(nh_0)^{-1}\}.$$

Proposition 2.1 provides the expansion of Plug-In bandwidth,

$$\frac{\hat{h}-h_0}{h_0}=\frac{-C_{PI}}{n}\sum_{i=1}^{n}\Big\{f^{(2L)}(X_i)-\mathbb{E}f^{(2L)}(X_i)\Big\}+\frac{Cb^{L_p}}{n}\sum_{i=1}^{n}\Big\{f^{(2L+L_p)}(X_i)-\mathbb{E}f^{(2L+L_p)}(X_i)\Big\}+o_p(n^{-1/2}b^{L_p})+O_p(n^{-1}).$$

Since we can, as stated in remark 2.3, make the effect of the second term of $(\hat{h}-h_0)/h_0$ negligible, we consider only the effect of the first term of $(\hat{h}-h_0)/h_0$ (see Appendix A.4 for details).
Define

$$\begin{aligned}
S_i &\equiv \mu_{20}^{-1/2}\left(K\left(\frac{X_i-x}{h_0}\right)-\mathbb{E}K\left(\frac{X_i-x}{h_0}\right)\right),\\
\Gamma_i &\equiv K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)-\mathbb{E}\left[K'\left(\frac{X_i-x}{h_0}\right)\left(\frac{X_i-x}{h_0}\right)+K\left(\frac{X_i-x}{h_0}\right)\right],\\
\mathcal{L}_i &\equiv f^{(2L)}(X_i)-\mathbb{E}f^{(2L)}(X_i).
\end{aligned}$$

Recalling that $S_{PI}(x)$ is defined as (2.5), we have from Lemma 1,

$$\begin{aligned}
S_{PI}(x) &= \frac{\sqrt{n\hat{h}}\Big(\hat{f}_{\hat{h}}(x)-\mathbb{E}\hat{f}_{h_0}(x)\Big)}{\mu_{20}^{1/2}}\\
&= \frac{\sqrt{nh_0}\Big(\hat{f}_{h_0}(x)-\mathbb{E}\hat{f}_{h_0}(x)\Big)}{\mu_{20}^{1/2}}-\frac{\sqrt{nh_0}\left(\frac{\hat{h}-h_0}{h_0}\right)\mathbb{E}\Gamma_{KDE_1}}{\mu_{20}^{1/2}}\\
&\quad-\frac{\left(\frac{\hat{h}-h_0}{h_0}\right)\sqrt{nh_0}(\Gamma_{KDE_1}-\mathbb{E}\Gamma_{KDE_1})}{\mu_{20}^{1/2}}+\frac{\sqrt{nh_0}\Big(\hat{f}_{h_0}(x)-\mathbb{E}\hat{f}_{h_0}(x)\Big)\left(\frac{\hat{h}-h_0}{h_0}\right)}{2\mu_{20}^{1/2}}+o_p\{(nh_0)^{-1}\}\\
&= \frac{1}{\sqrt{nh_0}}\sum_{i=1}^{n}S_i+\sqrt{nh_0}\sum_{l=0}^{L-1}\frac{C_{PI}C_{\Gamma,l}(x)h_0^{L+l}}{n}\sum_{i=1}^{n}\frac{\mathcal{L}_i}{\mu_{20}^{1/2}}\\
&\quad+\left(\frac{1}{\sqrt{nh_0}}\sum_{i=1}^{n}\frac{\Gamma_i}{\mu_{20}^{1/2}}\right)\left(\frac{C_{PI}}{n}\sum_{i=1}^{n}\mathcal{L}_i\right)-\frac{1}{2}\left(\frac{1}{\sqrt{nh_0}}\sum_{i=1}^{n}S_i\right)\left(\frac{C_{PI}}{n}\sum_{i=1}^{n}\mathcal{L}_i\right)+o_p\{(nh_0)^{-1}\}\\
&= \frac{1}{\sqrt{nh_0}}\sum_{i=1}^{n}S_i+\frac{C_{PI}h_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}}\sum_{i=1}^{n}\mathcal{L}_i\sum_{l=0}^{L-1}C_{\Gamma,l}(x)h_0^{l}\\
&\quad+\frac{C_{PI}}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\Gamma_i\mathcal{L}_j+\frac{C_{PI}}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}\sum_{i=1}^{n}\Gamma_i\mathcal{L}_i-\frac{C_{PI}}{2n^{3/2}h_0^{1/2}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}S_i\mathcal{L}_j-\frac{C_{PI}}{2n^{3/2}h_0^{1/2}}\sum_{i=1}^{n}S_i\mathcal{L}_i+o_p\{(nh_0)^{-1}\}\\
&\equiv S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)+o_p\{(nh_0)^{-1}\}. \qquad (17)
\end{aligned}$$

Define $F_{PI}(z)$ and $\tilde{F}_{PI}(z)$ as follows,

$$\begin{aligned}
F_{PI}(z) &= \mathbb{P}\Big(S_{PI}(x)\leqq z\Big),\\
\tilde{F}_{PI}(z) &= \Phi(z)+\phi(z)\left[(nh_0)^{-1/2}p_1(z)+(nh_0)^{-1}p_2(z)+\sum_{l=0}^{L-1}h_0^{L+l+1}p_{3,l}(z)+n^{-1/2}h_0^{1/2}p_4(z)\right].
\end{aligned}$$

To show the Edgeworth expansion is valid, we have to confirm $\sup_{z\in\mathbb{R}}\left|F_{PI}(z)-\tilde{F}_{PI}(z)\right| = o\{(nh_0)^{-1}\}$. First, we evaluate the remainder term.

$$\begin{aligned}\sup_{z\in\mathbb{R}}\left|F_{PI}(z)-\tilde{F}_{PI}(z)\right| &\leqq \sup_{z\in\mathbb{R}}\left|\mathbb{P}\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\leqq z\Big)-\tilde{F}_{PI}(z)\right| \\ &+\mathbb{P}\left(\left|S_{PI}(x)-\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\Big)\right|\geqq a_n\right)+O(a_n^{-1})\end{aligned}$$

where $a_n = nh_0(\log n)$. Since

$$\left|S_{PI}(x)-\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\Big)\right| = o_p\{(nh_0)^{-1}\},$$

we have

$$\mathbb{P}\left(\left|S_{PI}(x)-\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\Big)\right|\geqq a_n\right) = O\{(nh_0)^{-1}a_n^{-1}\} = o\{(nh_0)^{-1}\}.$$

Obviously, $O(a_n^{-1}) = o\{(nh_0)^{-1}\}$. Then, we only need to evaluate

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\leqq z\Big)-\tilde{F}_{PI}(z)\right|.$$

Define $\chi_{PI}(t)$ and $\tilde{\chi}_{PI}(t)$ as follows,

$$\begin{aligned}\chi_{PI}(t) &\equiv \mathbb{E}\left[exp\left\{it\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\Big)\right\}\right], \\ \tilde{\chi}_{PI}(t) &\equiv \exp\left(\frac{-t^2}{2}\right)\left[\left\{1+\frac{\mu_{30}\mu_{20}^{-3/2}}{6n^{1/2}h_0^{1/2}}(it)^3+\frac{\mu_{40}\mu_{20}^{-2}}{24nh_0}(it)^4+\frac{\mu_{30}^2\mu_{20}^{-3}}{72nh_0}(it)^6\right\}\right. \\ &\qquad \left. +C_{PI}\rho_{11}\mu_{20}^{-1}\left(\sum_{l=0}^{L-1}C_{\Gamma,l}(x)h_0^{L+l+1}\right)(it)^2+C_{PI}\frac{\rho_{11}\xi_{11}\mu_{20}^{-3/2}h_0^{1/2}}{n^{1/2}}(it)^3-C_{PI}\frac{\mu_{20}^{-1/2}\rho_{11}h_0^{1/2}}{2n^{1/2}}\{(it)^3+(it)\}\right].\end{aligned}$$

From Esseen's (1945) smoothing lemma,

$$\begin{aligned}&\sup_{z\in\mathbb{R}}\left|P\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\leqq z\Big)-\tilde{F}_{PI}(z)\right| \\ &\leqq \int_{-n^{\frac{2L}{2L+1}}\log n}^{n^{\frac{2L}{2L+1}}\log n}\left|\frac{\chi_{PI}(t)-\tilde{\chi}_{PI}(t)}{t}\right|dt+O\left(\frac{1}{n^{\frac{2L}{2L+1}}\log n}\right) \\ &\leqq \int_{-p}^{p}\left|\frac{\chi_{PI}(t)-\tilde{\chi}_{PI}(t)}{t}\right|dt+\int_{p\leqq|t|\leqq n^{\frac{2L}{2L+1}}\log n}\left|\frac{\chi_{PI}(t)}{t}\right|dt+\int_{p\leqq|t|\leqq n^{\frac{2L}{2L+1}}\log n}\left|\frac{\tilde{\chi}_{PI}(t)}{t}\right|dt+o\{(nh_0)^{-1}\} \\ &\leqq \int_{-p}^{p}\left|\frac{\chi_{PI}(t)-\tilde{\chi}_{PI}(t)}{t}\right|dt+\int_{p\leqq|t|\leqq n^{\frac{2L}{2L+1}}\log n}\left|\frac{\chi_{PI}(t)}{t}\right|dt+\int_{p\leqq|t|}\left|\frac{\tilde{\chi}_{PI}(t)}{t}\right|dt+o\{(nh_0)^{-1}\} \\ &\equiv (A)+(B)+(C)+o\{(nh_0)^{-1}\} \qquad (18)\end{aligned}$$

where $p = \min\left\{\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}, \log n\right\}$. To prove the validity of the Edgeworth expansion, we show that each term of (18) has the convergence rate $o\{(nh_0)^{-1}\}$.

In order to evaluate (A), we represent $\chi_{PI}(t)$ as $\tilde{\chi}_{PI}(t)$ plus a remainder. From Lemmas 8, 9, 11, 12, and 13,

$$\begin{aligned}\chi_{PI}(t) &= \mathbb{E}\left[e^{it\Big(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x)\Big)}\right] \\ &= \mathbb{E}\left[e^{itS(x)}\Big\{1+it\Lambda_1(x)\Big\}\Big\{1+it\Lambda_2(x)\Big\}\Big\{1+it\Lambda_4(x)\Big\}\Big\{1+it\Lambda_5(x)\Big\}\right] \\ &\qquad +O(t^2\mathbb{E}|\Lambda_1(x)|^2)+O(t^2\mathbb{E}|\Lambda_2(x)|^2)+O(|t|\mathbb{E}|\Lambda_3(x)|)+O(t^2\mathbb{E}|\Lambda_4(x)|^2)+O(t^2\mathbb{E}|\Lambda_5(x)|^2)\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\Big[e^{itS(x)}\Big\{1 + it\Lambda_1(x) + it\Lambda_2(x) + it\Lambda_4(x) + it\Lambda_5(x)\Big\}\Big] \\
&\quad + O(t^2\mathbb{E}|\Lambda_1(x)|^2) + O(t^2\mathbb{E}|\Lambda_2(x)|^2) + O(|t|\mathbb{E}|\Lambda_3(x)|) + O(t^2\mathbb{E}|\Lambda_4(x)|^2) + O(t^2\mathbb{E}|\Lambda_5(x)|^2) \\
&\quad + O(t^2\mathbb{E}|\Lambda_1(x)\Lambda_2(x)|) + O(t^2\mathbb{E}|\Lambda_1(x)\Lambda_4(x)|) + O(t^2\mathbb{E}|\Lambda_1(x)\Lambda_5(x)|) \\
&\quad + O(t^2\mathbb{E}|\Lambda_2(x)\Lambda_4(x)|) + O(t^2\mathbb{E}|\Lambda_2(x)\Lambda_5(x)|) + O(t^2\mathbb{E}|\Lambda_4(x)\Lambda_5(x)|) \\
&\equiv \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)} + \text{(V)} \\
&\quad + O(t^2\mathbb{E}|\Lambda_1(x)|^2) + O(t^2\mathbb{E}|\Lambda_2(x)|^2) + O(|t|\mathbb{E}|\Lambda_3(x)|) + O(t^2\mathbb{E}|\Lambda_4(x)|^2) + O(t^2\mathbb{E}|\Lambda_5(x)|^2) \\
&\quad + O(t^2\mathbb{E}|\Lambda_1(x)\Lambda_2(x)|) + O(t^2\mathbb{E}|\Lambda_1(x)\Lambda_4(x)|) + O(t^2\mathbb{E}|\Lambda_1(x)\Lambda_5(x)|) \\
&\quad + O(t^2\mathbb{E}|\Lambda_2(x)\Lambda_4(x)|) + O(t^2\mathbb{E}|\Lambda_2(x)\Lambda_5(x)|) + O(t^2\mathbb{E}|\Lambda_4(x)\Lambda_5(x)|) \\
&= \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)} + \text{(V)} \\
&\quad + O(t^2 h_0^{2L+1}) + O(t^2 n^{-1}) + O(|t|\{n^{-1/2} h_0^{\frac{2L+1}{2}} + n^{-1}\}) + O(t^2 n^{-1} h_0) + O(t^2 n^{-1} h_0) \\
&\quad + O(t^2 n^{-1/2} h_0^{\frac{2L+1}{2}}) + O(t^2 n^{-1/2} h_0^{L+1}) + O(t^2 n^{-1/2} h_0^{L+1}) \\
&\quad + O(t^2 n^{-1} h_0^{1/2}) + O(t^2 n^{-1} h_0^{1/2}) + O(t^2 n^{-1} h_0) \\
&= \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)} + \text{(V)} + O(t^2 n^{-1}) + O(|t| n^{-1}),
\end{aligned} \tag{19}
$$

where the fourth equality follows from Lemmas 10, 14,15, 16, 17, and 18 and the final equality uses $h_0 = O(n^{-1/(2L+1)})$.

Define $\gamma(t) = \mathbb{E}\Big[e^{\frac{it}{\sqrt{nh}} S_i}\Big]$. We have

$$
\text{(I)} = \mathbb{E}\Big[e^{\frac{it}{\sqrt{nh}} \sum_{i=1}^n S_i}\Big] = \mathbb{E}\Big[e^{\frac{it}{\sqrt{nh}} S_1}\Big]^n = \gamma(t)^n, \tag{20}
$$

from, Lemma 5,

$$
\begin{aligned}
\text{(II)} &= \mathbb{E}\Big[e^{itS(x)}(it\Lambda_1(x))\Big] \\
&= \gamma(t)^{n-1} \frac{C_{PI} h_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}} n \mathbb{E}\Big[e^{\frac{it}{\sqrt{nh_0}} S_1} \mathcal{L}_1\Big] \left(\sum_{l=0}^{L-1} C_{\Gamma,l}(x) h_0^l\right)(it) \\
&= \gamma(t)^{n-1} \frac{C_{PI} n^{1/2} h_0^{\frac{2L+1}{2}}}{\mu_{20}^{1/2}} \mathbb{E}\Big[\Big\{1 + \frac{it}{(nh_0)^{1/2}} S_1 + \frac{(it)^2}{2nh_0} S_1^2\Big\}\mathcal{L}_1\Big] \left(\sum_{l=0}^{L-1} C_{\Gamma,l}(x) h_0^l\right)(it) + o\{(nh)^{-1}\} \\
&= \gamma(t)^{n-1} \frac{C_{PI} n^{1/2} h_0^{\frac{2L+1}{2}}}{\mu_{20}^{1/2}} \frac{1}{(nh_0)^{1/2}} \mathbb{E}[S_1 \mathcal{L}_1] \left(\sum_{l=0}^{L-1} C_{\Gamma,l}(x) h_0^l\right)(it)^2 \\
&\quad + O(n^{-1/2} h_0^{(2L+1)/2}) O(n) O(|t|^3 n^{-1} h_0^{-1}) O(h_0) \\
&= \gamma(t)^{n-1} \frac{C_{PI}}{\mu_{20}^{1/2}} \mathbb{E}[S_1 \mathcal{L}_1] \left(\sum_{l=0}^{L-1} C_{\Gamma,l}(x) h_0^{L+l}\right)(it)^2 + O(|t|^3 n^{-1/2} h_0^{\frac{(2L+1)}{2}}) \\
&= \gamma(t)^{n-1} \frac{C_{PI}}{\mu_{20}^{1/2}} \mathbb{E}[S_1 \mathcal{L}_1] \left(\sum_{l=0}^{L-1} C_{\Gamma,l}(x) h_0^{L+l}\right)(it)^2 + O(|t|^3 n^{-1}),
\end{aligned} \tag{21}
$$

from Lemma 5 and 6,

$$
\begin{aligned}
\text{(III)} &= \mathbb{E}\Big[e^{itS(x)}(it\Lambda_2(x))\Big] \\
&= \gamma(t)^{n-2} \frac{C_{PI} n(n-1)}{n^{3/2} h_0^{1/2} \mu_{20}^{1/2}} \mathbb{E}\Big[e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)} \Gamma_1 \mathcal{L}_2\Big](it)
\end{aligned}
$$

$$
\begin{aligned}
&= \gamma(t)^{n-2}\frac{C_{PI}n(n-1)}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}\mathbb{E}\Bigg[\Big\{1+\frac{it}{(nh_0)^{1/2}}(S_1+S_2) \\
&\quad + \frac{(it)^2}{2nh_0}(S_1+S_2)^2 + \frac{(it)^3}{6(nh_0)^{3/2}}(S_1+S_2)^3\Big\}\Gamma_1\mathcal{L}_2\Bigg](it) + o\{(nh)^{-1}\} \\
&= \gamma(t)^{n-2}\frac{C_{PI}n(n-1)}{n^{5/2}h_0^{3/2}\mu_{20}^{1/2}}\mathbb{E}[S_1\Gamma_1]\mathbb{E}[S_2\mathcal{L}_2](it)^3 + O(n^2)O(t^4n^{-3}h_0^{-2})O(h_0)O(h_0) \\
&= \gamma(t)^{n-2}\frac{C_{PI}n(n-1)}{n^{5/2}h_0^{3/2}\mu_{20}^{1/2}}\mathbb{E}[S_1\Gamma_1]\mathbb{E}[S_2\mathcal{L}_2](it)^3 + O(t^4n^{-1}) \\
&= \gamma(t)^{n-2}\frac{C_{PI}}{n^{1/2}h_0^{3/2}\mu_{20}^{1/2}}\mathbb{E}[S_1\Gamma_1]\mathbb{E}[S_2\mathcal{L}_2](it)^3 + O(t^4n^{-1}),
\end{aligned} \tag{22}
$$

from Lemma 5,

$$
\begin{aligned}
\text{(IV)} &= \mathbb{E}\left[e^{itS(x)}(it\Lambda_4(x))\right] \\
&= \frac{-C_{PI}n(n-1)}{2n^{3/2}h_0^{1/2}}\gamma(t)^{n-2}\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)}S_1\mathcal{L}_2\right](it) \\
&= \frac{-C_{PI}n^{1/2}}{2h_0^{1/2}}\gamma(t)^{n-2}\mathbb{E}\Bigg[\Bigg\{1+\frac{it}{(nh_0)^{1/2}}(S_1+S_2) \\
&\quad + \frac{(it)^2}{2nh_0}(S_1+S_2)^2 + \frac{(it)^3}{6(nh_0)^{3/2}}(S_1+S_2)^3\Bigg\}S_1\mathcal{L}_2\Bigg](it) + o\{(nh)^{-1}\} \\
&= \frac{-C_{PI}}{2n^{1/2}h_0^{3/2}}\gamma(t)^{n-2}\mathbb{E}[S_1^2]\mathbb{E}[S_2\mathcal{L}_2](it)^3 + O(n^{1/2}h_0^{-1/2})O(t^4n^{-3/2}h_0^{-3/2})O(h_0)O(h_0) \\
&= \frac{-C_{PI}}{2n^{1/2}h_0^{1/2}}\gamma(t)^{n-2}\mathbb{E}[S_2\mathcal{L}_2](it)^3 + O(t^4n^{-1})
\end{aligned} \tag{23}
$$

where the final equality uses $\mathbb{E}[S_1^2] = h_0$, and from Lemma 5,

$$
\begin{aligned}
\text{(V)} &= \mathbb{E}\left[e^{itS(x)}(it\Lambda_5(x))\right] \\
&= \frac{-C_{PI}n}{2n^{3/2}h_0^{1/2}}\gamma(t)^{n-1}\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}S_1}S_1\mathcal{L}_1\right](it) \\
&= \frac{-C_{PI}}{2n^{1/2}h_0^{1/2}}\gamma(t)^{n-1}\mathbb{E}\left[\left\{1+\frac{it}{\sqrt{nh_0}}S_1\right\}S_1\mathcal{L}_1\right](it) + o\{(nh)^{-1}\} \\
&= \frac{-C_{PI}}{2n^{1/2}h_0^{1/2}}\gamma(t)^{n-1}\mathbb{E}[S_1\mathcal{L}_1](it) + O(n^{-1/2}h_0^{-1/2})O(t^2n^{-1/2}h_0^{-1/2})O(h_0) \\
&= \frac{-C_{PI}}{2n^{1/2}h_0^{1/2}}\gamma(t)^{n-1}\mathbb{E}[S_1\mathcal{L}_1](it) + O(t^2n^{-1})
\end{aligned} \tag{24}
$$

then

$$
\begin{aligned}
\chi_{PI}(t) &= \text{(I)} + \text{(II)} + \text{(III)} + \text{(IV)} + \text{(V)} + O(t^2n^{-1}) + O(|t|n^{-1}) \\
&= \gamma(t)^n + \gamma(t)^{n-1}\frac{C_{PI}}{\mu_{20}^{1/2}}\mathbb{E}[S_1\mathcal{L}_1]\left(\sum_{l=0}^{L-1}C_{\Gamma,l}h_0^{L+l}\right)(it)^2 \\
&\quad + \gamma(t)^{n-2}\frac{C_{PI}}{\mu_{20}^{1/2}}\frac{1}{n^{1/2}h_0^{3/2}}\mathbb{E}[S_1\Gamma_1]\mathbb{E}[S_2\mathcal{L}_2](it)^3 \\
&\quad - \left\{\gamma(t)^{n-2}(it)^3 + \gamma(t)^{n-1}(it)\right\}\frac{C_{PI}}{2n^{1/2}h_0^{1/2}}\mathbb{E}[S_1\mathcal{L}_1]
\end{aligned}
$$

$$+O\Big((|t|+t^2+|t|^3+t^4)n^{-1}\Big).$$

For $m=0,1,2$, by Feller (1971, p535-536),

$$\gamma(t)^{n-m}=\exp\left(\frac{-t^2}{2}\right)\left\{1+\frac{\mu_{30}\mu_{20}^{-3/2}}{6n^{1/2}h_0^{1/2}}(it)^3+\frac{\mu_{40}\mu_{20}^{-2}}{24nh_0}(it)^4+\frac{\mu_{30}^2\mu_{20}^{-3}}{72nh_0}(it)^6\right\}$$
$$+o\Big((nh_0)^{-1}(t^4+|t|^9)e^{-t^2/4}\Big).$$

By (20), (21), (22), (23) and (24), noting $\mathbb{E}[S_1\mathcal{L}_1]=h_0\mu_{20}^{-1/2}\rho_{11}$, $\mathbb{E}[S_1\Gamma_1]=h_0\mu_{20}^{-1/2}\xi_{11}$, and $\mathbb{E}[S_2\mathcal{L}_2]=h_0\mu_{20}^{-1/2}\rho_{11}$,

$$\begin{aligned}\chi_{PI}(t)&=\exp\left(\frac{-t^2}{2}\right)\Bigg[\left\{1+\frac{\mu_{30}\mu_{20}^{-3/2}}{6n^{1/2}h_0^{1/2}}(it)^3+\frac{\mu_{40}\mu_{20}^{-2}}{24nh_0}(it)^4+\frac{\mu_{30}^2\mu_{20}^{-3}}{72nh_0}(it)^6\right\}\\&\quad+C_{PI}\rho_{11}\mu_{20}^{-1}\left(\sum_{l=0}^{L-1}C_{\Gamma,l}(x)h_0^{L+1+l}\right)(it)^2+C_{PI}\frac{\rho_{11}\xi_{11}\mu_{20}^{-3/2}h_0^{1/2}}{n^{1/2}}(it)^3\\&\quad-C_{PI}\frac{\mu_{20}^{-1/2}\rho_{11}h_0^{1/2}}{2n^{1/2}}\{(it)^3+(it)\}\Bigg]\\&\quad+O\Big((|t|+t^2+|t|^3+t^4)n^{-1}\Big)+o\Big((nh_0)^{-1}(t^4+|t|^9)e^{-t^2/4}\Big)\\&=\tilde{\chi}_{PI}(t)+O\Big((|t|+t^2+|t|^3+t^4)n^{-1}\Big)+o\Big((nh_0)^{-1}(t^4+|t|^9)e^{-t^2/4}\Big).\end{aligned}$$

This implies

$$(A)=\int_{-p}^{p}\left|\frac{\chi_{PI}(t)-\tilde{\chi}_{PI}(t)}{t}\right|dt=o\{(nh_0)^{-1}\}$$

Next, we confirm $(B)=o\{(nh_0)^{-1}\}$, for $p\leqq|t|\leqq n^{\frac{2L}{2L+1}}\log n$. Define

$$\begin{aligned}S(x;m)&\equiv\frac{1}{n^{1/2}h_0^{1/2}}\sum_{i=1}^{m}S_i,\\\Lambda_1(x;m)&\equiv\frac{C_{PI}h_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}}\sum_{i=1}^{m}\mathcal{L}_i\left(\sum_{l=0}^{L-1}C_{\Gamma,l}(x)h_0^{l}\right),\\\Lambda_2(x;m)&\equiv\frac{C_{PI}}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}\sum_{i=1}^{m}\sum_{j\neq i}^{m}\Gamma_i\mathcal{L}_j,\\\Lambda_4(x;m)&\equiv-\frac{C_{PI}}{2n^{3/2}h_0^{1/2}}\sum_{i=1}^{m}\sum_{j\neq i}^{m}S_i\mathcal{L}_j\\\Lambda_5(x;m)&\equiv-\frac{C_{PI}}{2n^{3/2}h_0^{1/2}}\sum_{i=1}^{m}S_i\mathcal{L}_i\end{aligned}$$

then

$$\begin{aligned}|\chi_{PI}(t)|&=|\mathbb{E}e^{it(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_3(x)+\Lambda_4(x)+\Lambda_5(x))}|\\&\leqq|\mathbb{E}e^{it(S(x)+\Lambda_1(x)+\Lambda_2(x)+\Lambda_4(x)+\Lambda_5(x)}|+O(|t||\mathbb{E}\Lambda_3(x)|)\\&\leqq\Big|\mathbb{E}e^{it(S(x)+(\Lambda_1(x)-\Lambda_1(x;m))+(\Lambda_2(x)-\Lambda_2(x;m))+(\Lambda_4(x)-\Lambda_4(x;m))+(\Lambda_5(x)-\Lambda_5(x;m)))}\\&\qquad\times\Big\{1+it\Lambda_1(x;m)\Big\}\Big\{1+it\Lambda_2(x;m)\Big\}\Big\{1+it\Lambda_4(x;m)\Big\}\Big\{1+it\Lambda_5(x;m)\Big\}\Big|\end{aligned}$$

$$
\begin{aligned}
&+O(t^2\{\mathbb{E}\Lambda_1(x;m)^2+\mathbb{E}\Lambda_2(x;m)^2+\mathbb{E}\Lambda_4(x;m)^2+\mathbb{E}\Lambda_5(x;m)^2\})+O(|t||\mathbb{E}\Lambda_3(x)|)\\
\leqq&\left|\mathbb{E}e^{it(S(x)+(\Lambda_1(x)-\Lambda_1(x;m))+(\Lambda_2(x)-\Lambda_2(x;m))+(\Lambda_4(x)-\Lambda_4(x;m))+(\Lambda_5(x)-\Lambda_5(x;m)))}\right|\\
&+|t|\Big|\mathbb{E}e^{it(S(x)+(\Lambda_1(x)-\Lambda_1(x;m))+(\Lambda_2(x)-\Lambda_2(x;m))+(\Lambda_4(x)-\Lambda_4(x;m))+(\Lambda_5(x)-\Lambda_5(x;m)))}\\
&\qquad\times\{\Lambda_1(x;m)+\Lambda_2(x;m)+\Lambda_4(x;m)+\Lambda_5(x;m)\}\Big|\\
&+O(t^2\{\mathbb{E}\Lambda_1(x;m)^2+\mathbb{E}\Lambda_2(x;m)^2+\mathbb{E}\Lambda_4(x;m)^2+\mathbb{E}\Lambda_5(x;m)^2\\
&\qquad+\mathbb{E}|\Lambda_1(x;m)\Lambda_2(x;m)|+\mathbb{E}|\Lambda_1(x;m)\Lambda_4(x;m)|+\mathbb{E}|\Lambda_1(x;m)\Lambda_5(x;m)|\\
&\qquad+\mathbb{E}|\Lambda_2(x;m)\Lambda_4(x;m)|+\mathbb{E}|\Lambda_2(x;m)\Lambda_5(x;m)|+\mathbb{E}|\Lambda_4(x;m)\Lambda_5(x;m)|\})\\
&+O(|t||\mathbb{E}\Lambda_3(x)|).
\end{aligned}
\tag{25}
$$

The first term of (25) is bounded as below.

$$
\begin{aligned}
&\left|\mathbb{E}e^{itS(x;m)}\mathbb{E}e^{it((S(x)-S(x;m))+(\Lambda_1(x)-\Lambda_1(x;m))+(\Lambda_2(x)-\Lambda_2(x;m)+(\Lambda_4(x)-\Lambda_4(x;m)+(\Lambda_5(x)-\Lambda_5(x;m)))}\right|\\
&=\left|\mathbb{E}e^{itS(x;m)}\right|\left|\mathbb{E}e^{it((S(x)-S(x;m))+(\Lambda_1(x)-\Lambda_1(x;m))+(\Lambda_2(x)-\Lambda_2(x;m)+(\Lambda_4(x)-\Lambda_4(x;m))+(\Lambda_5(x)-\Lambda_5(x;m)))}\right|\\
&\leqq\left|\mathbb{E}e^{itS(x;m)}\right|=|\gamma(t)|^m.
\end{aligned}
\tag{26}
$$

Similarly, the second term of (25) devided by $|t|$ is bounded by

$$
|\mathbb{E}\{e^{itS(x;m)}\Lambda_1(x;m)\}|+|\mathbb{E}\{e^{itS(x;m)}\Lambda_2(x;m)\}|+|\mathbb{E}\{e^{itS(x;m)}\Lambda_4(x;m)\}|+|\mathbb{E}\{e^{itS(x;m)}\Lambda_5(x;m)\}|,
$$

where each term is bounded as follows. Let $C(x)$ be some positive and bounded generic function.

$$
\begin{aligned}
|\mathbb{E}\{e^{itS(x;m)}\Lambda_1(x;m)\}|&=\left|\gamma(t)^{m-1}\frac{C_{PI}C_{\Gamma,0}(x)h_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}}m\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}S_1(x)}\mathcal{L}_1\right]\right|+s.o.\\
&\leqq|\gamma(t)|^{m-1}\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}}|C_{PI}C_{\Gamma,0}(x)|\left|\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}S_1(x)}\mathcal{L}_1\right]\right|+s.o.\\
&\leqq|\gamma(t)|^{m-1}\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}}|C_{PI}C_{\Gamma,0}(x)|\mathbb{E}\left|e^{\frac{it}{\sqrt{nh_0}}S_1(x)}\mathcal{L}_1\right|+s.o.\\
&\leqq|\gamma(t)|^{m-1}\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}}|C_{PI}C_{\Gamma,0}(x)|\mathbb{E}\left|e^{\frac{it}{\sqrt{nh_0}}S_1(x)}\right|E|\mathcal{L}_1|+s.o.\\
&\leqq|\gamma(t)|^{m-1}\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}\mu_{20}^{1/2}}|C_{PI}C_{\Gamma,0}(x)|\mathbb{E}|\mathcal{L}_1|+s.o.\\
&\leqq C(x)|\gamma(t)|^{m-1}\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}},
\end{aligned}
\tag{27}
$$

where the final inequality uses Lemma 7.

$$
\begin{aligned}
|\mathbb{E}\{e^{itS(x;m)}\Lambda_2(x;m)\}|&=\left|\gamma(t)^{m-2}\frac{C_{PI}}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}m(m-1)\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)}\Gamma_1\mathcal{L}_2\right]\right|\\
&\leqq|\gamma(t)|^{m-2}\frac{m(m-1)}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}|C_{PI}|\left|\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)}\Gamma_1\mathcal{L}_2\right]\right|\\
&\leqq|\gamma(t)|^{m-2}\frac{m(m-1)}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}|C_{PI}|\mathbb{E}\left|e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)}\right|\mathbb{E}|\Gamma_1\mathcal{L}_2|
\end{aligned}
$$

$$
\begin{aligned}
&\leqq |\gamma(t)|^{m-2}\frac{m(m-1)}{n^{3/2}h_0^{1/2}\mu_{20}^{1/2}}|C_{PI}|\mathbb{E}|\Gamma_1\mathcal{L}_2| \\
&\leqq C(x)|\gamma(t)|^{m-2}\frac{m(m-1)h_0^{1/2}}{n^{3/2}}, \qquad (28)
\end{aligned}
$$

where the final inequality uses Lemma 7.

$$
\begin{aligned}
|\mathbb{E}\{e^{itS(x;m)}\Lambda_4(x;m)\}| &= \left|\frac{C_{PI}m(m-1)}{2n^{3/2}h_0^{1/2}}\gamma(t)^{m-2}\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)}S_1\mathcal{L}_2\right]\right| \\
&\leqq |\gamma(t)|^{m-2}\frac{m(m-1)}{2n^{3/2}h_0^{1/2}}|C_{PI}|\left|\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)}S_1\mathcal{L}_2\right]\right| \\
&\leqq |\gamma(t)|^{m-2}\frac{m(m-1)}{2n^{3/2}h_0^{1/2}}|C_{PI}|\mathbb{E}\left|e^{\frac{it}{\sqrt{nh_0}}(S_1+S_2)}\right|\mathbb{E}|S_1\mathcal{L}_2| \\
&\leqq |\gamma(t)|^{m-2}\frac{m(m-1)}{2n^{3/2}h_0^{1/2}}|C_{PI}|\mathbb{E}|S_1\mathcal{L}_2| \\
&\leqq C(x)|\gamma(t)|^{m-2}\frac{m(m-1)h_0^{1/2}}{2n^{3/2}}, \qquad (29)
\end{aligned}
$$

where the final inequality uses Lemma 5 and 7.

$$
\begin{aligned}
|\mathbb{E}\{e^{itS(x;m)}\Lambda_5(x;m)\}| &= \left|\frac{C_{PI}m}{2n^{3/2}h_0^{1/2}}\gamma(t)^{m-1}\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}S_1}S_1\mathcal{L}_1\right]\right| \\
&\leqq |\gamma(t)|^{m-1}\frac{C_{PI}m}{2n^{3/2}h_0^{1/2}}|C_{PI}|\left|\mathbb{E}\left[e^{\frac{it}{\sqrt{nh_0}}S_1}S_1\mathcal{L}_1\right]\right| \\
&\leqq |\gamma(t)|^{m-1}\frac{C_{PI}m}{2n^{3/2}h_0^{1/2}}|C_{PI}|\mathbb{E}\left|e^{\frac{it}{\sqrt{nh_0}}S_1}\right|\mathbb{E}|S_1\mathcal{L}_1| \\
&\leqq |\gamma(t)|^{m-1}\frac{C_{PI}m}{2n^{3/2}h_0^{1/2}}|C_{PI}|\mathbb{E}|S_1\mathcal{L}_1| \\
&\leqq C(x)|\gamma(t)|^{m-1}\frac{mh_0^{1/2}}{2n^{3/2}} \qquad (30)
\end{aligned}
$$

where the final inequality uses Lemma 5. (27), (28), (29) and (30) imply

$$
\begin{aligned}
&|t|\Big|\mathbb{E}e^{it(S(x)+(\Lambda_1(x)-\Lambda_1(x;m))+(\Lambda_2(x)-\Lambda_2(x;m))+(\Lambda_4(x)-\Lambda_4(x;m))+(\Lambda_5(x)-\Lambda_5(x;m))} \\
&\qquad \times\{\Lambda_1(x;m)+\Lambda_2(x;m)+\Lambda_4(x;m)+\Lambda_5(x;m)\}\Big| \\
&\leqq C(x)\left\{|\gamma(t)|^{m-1}\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}}+|\gamma(t)|^{m-2}\frac{m^2h_0^{1/2}}{n^{3/2}}+|\gamma(t)|^{m-1}\frac{mh_0^{1/2}}{n^{3/2}}\right\}|t|. \qquad (31)
\end{aligned}
$$

Then, (25), (26), and (31) yield

$$
\begin{aligned}
|\chi_{PI}(t)| \leqq |\gamma(t)|^m &+ C(x)\left\{|\gamma(t)|^{m-1}\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}}+|\gamma(t)|^{m-2}\frac{m^2h_0^{1/2}}{n^{3/2}}+|\gamma(t)|^{m-1}\frac{mh_0^{1/2}}{n^{3/2}}\right\}|t| \\
&+O(t^2\{\mathbb{E}\Lambda_1(x;m)^2+\mathbb{E}\Lambda_2(x;m)^2+\mathbb{E}\Lambda_4(x;m)^2+\mathbb{E}\Lambda_5(x;m)^2 \\
&\qquad +\mathbb{E}|\Lambda_1(x;m)\Lambda_2(x;m)|+\mathbb{E}|\Lambda_1(x;m)\Lambda_4(x;m)|+\mathbb{E}|\Lambda_1(x;m)\Lambda_5(x;m)| \\
&\qquad +\mathbb{E}|\Lambda_2(x;m)\Lambda_4(x;m)|+\mathbb{E}|\Lambda_2(x;m)\Lambda_5(x;m)|+\mathbb{E}|\Lambda_4(x;m)\Lambda_5(x;m)|\}) \\
&+O(|t||\mathbb{E}\Lambda_3(x)|)
\end{aligned}
$$

$$\leqq C(x)|\gamma(t)|^{m-2}\left[1+\left\{\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}}+\frac{m^2h_0^{1/2}}{n^{3/2}}+\frac{mh_0^{1/2}}{n^{3/2}}\right\}|t|\right]$$
$$+O\left(t^2\left\{\frac{mh_0^{2L+1}}{n}+\frac{m^2}{n^3}+\frac{m^{3/2}h_0^{(2L+1)/2}}{n^2}\right\}\right)+O\left(|t|\left\{n^{-1/2}h_0^{(2L+1)/2}+n^{-1}\right\}\right)$$

where second inequality uses $|\gamma(t)|\leqq 1$ and Lemma 11,19,20,21,22 and 23.

We evaluate (B), partioning its range of integration into two parts, $p\leqq|t|\leqq\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}$ and $\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}\leqq|t|\leqq n^{\frac{2L}{2L+1}}\log n$.

(i) For $p\leqq|t|\leqq\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}$

Applying Taylor expansion to $e^{\frac{it}{\sqrt{nh_0}}S_1(x)}$ with respect to t, we have

$$|\gamma(t)-1-\frac{t^2}{2n}|\leqq\frac{|t|^3\mu_{20}^{-3/2}\mu_{30}}{6n^{3/2}h_0^{1/2}},$$

then for $|t|\leqq\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}$,

$$|\gamma(t)|\leqq 1-\frac{t^2}{2n}+\frac{|t|^3\mu_{20}^{-3/2}\mu_{30}}{6n^{3/2}h_0^{1/2}}\leqq 1-\frac{t^2}{2n}+\frac{t^2}{6n}=1-\frac{t^2}{3n}\leqq\exp\left(-\frac{t^2}{3n}\right),$$

then

$$|\chi_{PI}(t)|\leqq C(x)|\gamma(t)|^{m-2}\left[1+\left\{\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}}+\frac{m^2h_0^{1/2}}{n^{3/2}}+\frac{mh_0^{1/2}}{n^{3/2}}\right\}|t|\right]$$
$$+O\left(t^2\left\{\frac{mh_0^{2L+1}}{n}+\frac{m^2}{n^3}+\frac{m^{3/2}h_0^{(2L+1)/2}}{n^2}\right\}\right)+O\left(|t|\left\{n^{-1/2}h_0^{(2L+1)/2}+n^{-1}\right\}\right)$$
$$\leqq C(x)\exp\left(-\frac{(m-2)t^2}{3n}\right)\left[1+\left\{\frac{mh_0^{\frac{2L+1}{2}}}{n^{1/2}}+\frac{m^2h_0^{1/2}}{n^{3/2}}+\frac{mh_0^{1/2}}{n^{3/2}}\right\}|t|\right]$$
$$+O\left(t^2\left\{\frac{m}{n^2}+\frac{m^2}{n^3}+\frac{m^{3/2}}{n^{5/2}}\right\}\right)+O\left(|t|n^{-1}\}\right).$$

Using (A.21) in Nishiyama and Robinson (2000), we can take $m=[9n\log n/t^2]$ since $1\leqq m\leqq n-1$ holds for $p\leqq|t|\leqq\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}$ and sufficiently large n.

Because $m\geqq(9n\log n)/t^2-1$, for $|t|\leqq\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}$

$$\exp\left(-\frac{(m-2)t^2}{3n}\right)=\exp\left(-\frac{(m+1)t^2}{3n}\right)\exp\left(\frac{3t^2}{3n}\right)\leqq C\exp(-3\log n)\leqq\frac{C}{n^3},$$

and this implies, using $m\leqq(9n\log n)/t^2$,

$$|\chi_{PI}(t)|\leqq\frac{C(x)}{n^3}\left[1+n^{1/2}(\log n)h_0^{\frac{2L+1}{2}}\frac{1}{|t|}+n^{1/2}(\log n)^2h_0^{1/2}\frac{1}{|t|^3}+n^{-1/2}(\log n)h_0^{1/2}\frac{1}{|t|}\right]$$
$$+O\left(n^{-1}(\log n)+n^{-1}(\log n)^2\frac{1}{t^2}+n^{-1}(\log n)^{3/2}\frac{1}{|t|}\right)+O(|t|n^{-1})$$

Therefore, dropping the integral range $p\leqq|t|\leqq\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}$ on the right-hand side,

$$\int_{p\leqq|t|\leqq\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}}\left|\frac{\chi_{PI}(t)}{t}\right|dt$$

$$\leqq C(x)\left[\left\{n^{-3}+n^{-1}(\log n)\right\}\int\frac{dt}{|t|}+\left\{n^{-5/2}(\log n)h_0^{\frac{2L+1}{2}}+n^{-7/2}(\log n)h_0^{1/2}\right.\right.$$
$$\left.\left.+n^{-1}(\log n)^{3/2}\right\}\int\frac{dt}{t^2}+n^{-1}(\log n)^2\int\frac{dt}{|t|^3}+n^{-5/2}(\log n)^2h_0^{1/2}\int\frac{dt}{t^4}\right]+O(n^{-1})$$
$$=o\{(nh_0)^{-1}\}$$

(ii) For $\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}\leqq|t|\leqq n^{\frac{2L}{2L+1}}\log n$, there exist $\eta\in(0,1)$, such that $|\gamma(t)|\leqq 1-\eta$ from Assumption 9. We can take $m=[-3\log n/\log(1-\eta)]$ since $1\leqq m\leqq n-1$ for sufficiently large n. Then $\chi_{PI}(t)$ is bounded as follow.

$$|\chi_{PI}(t)|$$
$$\leqq C(1-\eta)^{-3\log n/\log(1-\eta)}$$
$$\times\left[1+\left\{\frac{h_0^{\frac{2L+1}{2}}}{n^{1/2}}+\frac{h_0^{1/2}}{n^{3/2}}\right\}|t|\left(\frac{-3\log n}{\log(1-\eta)}\right)+\frac{h_0^{1/2}}{n^{3/2}}|t|\left(\frac{-3\log n}{\log(1-\eta)}\right)^2\right]$$
$$+O\left(t^2\left\{n^{-2}\left(\frac{-3\log n}{\log(1-\eta)}\right)+n^{-3}\left(\frac{-3\log n}{\log(1-\eta)}\right)^2+n^{-5/2}\left(\frac{-3\log n}{\log(1-\eta)}\right)^{3/2}\right\}\right)$$

Noting that

$$(1-\eta)^{-3\log n/\log(1-\eta)}=(1-\eta)^{\log n^{-3}/\log(1-\eta)}=(1-\eta)^{\log_{(1-\eta)}n^{-3}}=n^{-3},$$

$$\int_{\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}}\leqq|t|\leqq n^{\frac{2L}{2L+1}}\log n}\left|\frac{\chi_{PI}(t)}{t}\right|dt$$
$$=O\left(\frac{\log(n^{\frac{2L}{2L+1}}\log n)}{n^3}+\frac{\log n}{n^3}\left\{\frac{h_0^{\frac{2L+1}{2}}}{n^{1/2}}+\frac{h_0^{1/2}}{n^{3/2}}\right\}(n^{\frac{2L}{2L+1}}\log n)+\frac{(\log n)^2}{n^3}\frac{h_0^{1/2}}{n^{3/2}}(n^{\frac{2L}{2L+1}}\log n)\right)$$
$$+O\left(n^{\frac{2L}{2L+1}}\log n\left\{n^{-2}(\log n)+n^{-2}(\log n)^2+n^{-3/2}(\log n)^{3/2}\right\}\right)$$
$$=o\{(nh_0)^{-1}\}$$

Finally, we evaluate (C). For some constant C,

$$(C)=\int_{p\leqq|t|}\frac{1}{|t|}e^{\frac{-t^2}{2}}\left|1+\frac{\mu_{30}\mu_{20}^{-3/2}}{6n^{1/2}h_0^{1/2}}(it)^3+\frac{\mu_{40}\mu_{20}^{-2}}{24nh_0}(it)^4+\frac{\mu_{30}^2\mu_{20}^{-3}}{72nh_0}(it)^6\right.$$
$$+C_{PI}\rho_{11}\mu_{20}^{-1}\left(\sum_{l=0}^{L-1}C_{\Gamma,l}(x)h_0^{L+l+1}\right)(it)^2+C_{PI}\frac{\rho_{11}\xi_{11}\mu_{20}^{-3/2}h_0^{1/2}}{n^{1/2}}(it)^3$$
$$\left.-C_{PI}\frac{\mu_{20}^{-1/2}\rho_{11}h_0^{1/2}}{2n^{1/2}}\{(it)^3+(it)\}\right|dt$$
$$\leqq C\left[\int_p^\infty\frac{1}{t}e^{\frac{-t^2}{2}}dt+\frac{h_0^{1/2}}{n^{1/2}}\int_p^\infty t^2e^{\frac{-t^2}{2}}dt+\frac{1}{nh_0}\int_p^\infty(t^3+t^5)e^{\frac{-t^2}{2}}dt\right.$$
$$\left.+h_0^{L+1}\int_p^\infty te^{\frac{-t^2}{2}}dt+\frac{h_0^{1/2}}{n^{1/2}}\int_p^\infty t^2e^{\frac{-t^2}{2}}dt+\frac{h_0^{1/2}}{n^{1/2}}\int_p^\infty(t^2+1)e^{\frac{-t^2}{2}}dt\right]$$

Since $p=\min\left\{\frac{n^{1/2}h_0^{1/2}}{\mu_{20}^{-3/2}\mu_{30}},\log n\right\}$,the first integral is smaller than $p^{-2}\int_p^\infty te^{-t^2/2}dt=p^{-2}e^{-p^2/2}=o(n^{-1})$, the second and fifth integrals are smaller than $p^{-1}\int_p^\infty t^3e^{-t^2/2}dt=p^{-1}e^{-p^2/2}(p^2+2)=o(n^{-1})$, the third integral is $\int_p^\infty(t^3+t^5)e^{-t^2/2}dt=$

$e^{-p^2/2}(p^4+5p^2+10) = o(n^{-1})$, the fourth integral is $\int_p^\infty te^{-t^2/2}dt = e^{-p^2/2} = o(n^{-1})$, and the final integral is $p^{-1}e^{-t^2/2}(p^2+3) = o(n^{-1})$. It follows that $(C) = o\{(nh_0)^{-1}\}$. Thus the expansion is valid. □

A.4 Proof of Theorem 3.2

Proof. Define ϵ and ϵ_{PI} as follows,

$$\epsilon \equiv \mathbb{P}(S(x) \leqq z) - \Phi(z) - \phi(z)\Big[(nh_0)^{-1/2}p_1(z) + (nh_0)^{-1}p_2(z)\Big],$$

$$\begin{aligned}\epsilon_{PI} &\equiv \mathbb{P}(S_{\mathrm{PI}}(x) \leqq z) - \Phi(z)\\ &\quad - \phi(z)\Big[(nh_0)^{-1/2}p_1(z) + h_0^{L+1}p_{3,0}(z) + n^{-1/2}h_0^{1/2}p_4(z) + (nh_0)^{-1}p_2(z)\Big].\end{aligned}$$

Then we have,

$$\begin{aligned}&\sup_{z\in\mathbb{R}}\Big|\mathbb{P}(S(x) \leqq z) - \mathbb{P}(S_{PI}(x) \leqq z) - \phi(z)\big[h_0^{L+1}p_{3,0}(z) + n^{-1/2}h_0^{1/2}p_4(z)\big]\Big|\\ &= \sup_{z\in\mathbb{R}}|\epsilon - \epsilon_{PI}| = o(h_0^{L+1} + n^{-1/2}h_0^{1/2}).\end{aligned}$$

□

A.5 About Remark 2.3

Recall we can expand $(\hat{h}-h_0)/h_0$ as follows,

$$\begin{aligned}\frac{\hat{h}-h_0}{h_0} &= \frac{-C_{\mathrm{PI}}}{n}\sum_{i=1}^n\Big\{f^{(2L)}(X_i) - \mathbb{E}f^{(2L)}(X_i)\Big\}\\ &\qquad + \frac{Cb^{L_p}}{n}\sum_{i=1}^n\Big\{f^{(2L+L_p)}(X_i) - \mathbb{E}f^{(2L+L_p)}(X_i)\Big\} + o_p(n^{-1/2}b^{L_p})\\ &= O_p(n^{-1/2}) + O_p(n^{-1/2}b^{L_p}) + o_p(n^{-1/2}b^{L_p})\end{aligned}$$

In Remark 2.3, we stated that we can make the second and subsequent terms as small as we like by letting kernel order L_p be high enough. In this section we provide an explanation for this statement.

Hall and Marron (1987) have shown MSE of $\hat{I}_L$ is given as follows,

$$MSE(\hat{I}_L) = O(n^{-2}b^{-8L-1} + n^{-1} + b^{2L_p}). \tag{32}$$

This implies the MSE optimal bandwidth sequence is $b = cn^{\frac{-2}{8L+2L_p+1}}$ for some positive constant c.

In order for the effect of the second term to be negligible, it is sufficient that the following conditions hold from Cauchy-Schwarz inequality (note that the Cauchy-Schwarz inequality should not provide sharp bounds, so the second term may have no effect under milder conditions). Suppose that $b = cn^q$ ($q < 0$) for some positive constant c,

$$\begin{aligned}&\sqrt{nh_0}\left\{\mathbb{E}\left(\frac{Cb^{L_p}}{n}\sum_{i=1}^n\Big\{f^{(2L+L_p)}(X_i) - \mathbb{E}f^{(2L+L_p)}(X_i)\Big\}\right)^2 \mathbb{E}\Gamma^2_{KDE_1}\right\}^{1/2} = o\{(nh_0)^{-1}\}\\ \Longleftrightarrow\ & O\{(nh_0)^{1/2}\}O(n^{-1/2}b^{L_p})O(h_0^L) = o\{(nh_0)^{-1}\}\\ \Longleftrightarrow\ & (nh_0)^{1/2}(n^{-1/2}b^{L_p})h_0^L \ll (nh_0)^{-1}\end{aligned}$$

$$
\begin{aligned}
&\iff b^{L_p} h_0^{L+\frac{1}{2}} \ll (nh_0)^{-1} \\
&\iff b^{L_p} \ll n^{-1/2} h_0^{-1} \\
&\iff n^{qL_p} \ll n^{\frac{-2L+1}{2(2L+1)}} \iff qL_p < \frac{-2L+1}{2(2L+1)} \iff L_p > \frac{-2L+1}{2(2L+1)q} \qquad (33)
\end{aligned}
$$

where $A_n \ll B_n$ means $A_n = o(B_n)$. This implies what we mentioned in Remark 2.3. When one take $q = -2/(8L+2L_p+1)$ (MSE optimal), it is possible to make the effect of pilot bandwidth on the asymptotic structure of KDE up to the order of $O\{(nh_0)^{-1}\}$ negligible by using the higher-order kernel H which satisfies $L_p > \frac{(2L-1)(8L+1)}{4L+6}$. For example, when $L = 2$, $L_p \geqq 4$ satisfies the condition, which is not unrealistic choice of kernel orders L and L_p.

B Lemmas

Lemma 1. *Under Assumptions 1, 4, 5, 6, 10, and 11,*

$$
\mathbb{E}\Gamma_{KDE_1} = \sum_{l=0}^{L-1} C_{\Gamma,l}(x) h_0^{L+l} + o(h_0^{2L-1}), \quad \textit{where} \quad C_{\Gamma,l}(x) \equiv -\left(\int u^{L+l} K(u)du\right) \frac{f^{(L+l)}(x)}{(L+l-1)!}
$$

Proof.

$$
\begin{aligned}
\mathbb{E}\Gamma_{KDE_1} &= \mathbb{E}\left[\frac{1}{nh_0}\sum_{i=1}^{n} K'\left(\frac{X_i - x}{h_0}\right)\left(\frac{X_i - x}{h_0}\right)\right] + \mathbb{E}\left[\frac{1}{nh_0}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_0}\right)\right] \\
&= \frac{1}{h_0}\int K'\left(\frac{z-x}{h_0}\right)\left(\frac{z-x}{h_0}\right) f(z)dz + \frac{1}{h_0}\int K\left(\frac{z-x}{h_0}\right) f(z)dz \\
&= \int K'(u) u f(x+uh_0)du + \int K(u) f(x+uh_0)du \\
&= -\int K(u) f(x+uh_0)du - \int K(u) u f'(x+uh_0) h_0 du + \int K(u) f(x+uh_0)du \\
&= -\int K(u) u f'(x+uh_0) h_0 du \\
&= -\int K(u) u\left\{f^{(1)}(x) + \cdots + \frac{f^{(L)}(x)}{(L-1)!}(uh_0)^{L-1} + \cdots + \frac{f^{(2L)}}{(2L-1)!}(uh_0)^{2L-1}\right\} h_0 du + o(h_0^{2L-1}) \\
&= -\sum_{l=0}^{L}\left(\int u^{L+l} K(u)du\right)\frac{f^{(L+l)}(x)}{(L+l-1)!} h_0^{L+l} + o(h_0^{2L-1}) \\
&\equiv \sum_{l=0}^{L-1} C_{\Gamma,l}(x) h_0^{L+l} + o(h_0^{2L-1})
\end{aligned}
$$

The fourth equality follows from integration by part of the first term and Assumption 2,11, the seventh equality follows from the expansion of $f'(x+uh_0)$ around $h_0 = 0$ and Assumption 2 and the eighth equality follows from Assumption 10. ◻

Lemma 2. *Under Assumptions 1, 2, 4,5, 10, and 12,*

$$
\mathbb{E}\Gamma_{KDE_2} = O(h_0^L)
$$

Proof.

$$
\mathbb{E}\Gamma_{KDE_2} = \mathbb{E}\left[\frac{1}{nh_0}\sum_{i=1}^{n} K''\left(\frac{X_i - x}{h_0}\right)\left(\frac{X_i - x}{h_0}\right)^2\right] + \mathbb{E}\left[\frac{4}{nh_0}\sum_{i=1}^{n} K'\left(\frac{X_i - x}{h_0}\right)\left(\frac{X_i - x}{h_0}\right)\right] + \mathbb{E}\left[\frac{2}{nh_0}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_0}\right)\right]
$$

$$
\begin{aligned}
&= \frac{1}{h_0}\int K''\left(\frac{z-x}{h_0}\right)\left(\frac{z-x}{h_0}\right)^2 f(z)dz + \frac{4}{h_0}\int K'\left(\frac{z-x}{h_0}\right)\left(\frac{z-x}{h_0}\right) f(z)dz + \frac{2}{h_0}\int K\left(\frac{z-x}{h_0}\right) f(z)dz \\
&= 2\int K(u)f(x+uh_0)du \\
&\quad + 4\left\{-\int K(u)f(x+uh_0)du - \int K(u)uf'(x+uh_0)h_0du\right\} \\
&\quad + \left\{-2\int K'(u)uf(x+uh_0) - \int K'(u)u^2 f'(x+uh_0)h_0du\right\} \\
&= 2\int K(u)f(x+uh_0)du \\
&\quad + 4\left\{-\int K(u)f(x+uh_0)du - \int K(u)uf'(x+uh_0)h_0du\right\} \\
&\quad + \left\{-2\left[-\int K(u)f(x+uh_0)du - \int K(u)uf'(x+uh_0)h_0du\right]\right\} \\
&\quad + \left\{-\left[-2\int K(u)uf'(x+uh_0)h_0du - \int K(u)u^2 f''(x+uh_0)h_0^2du\right]\right\} \\
&= \int K(u)u^2 f''(x+uh_0)h_0^2du = O(h_0^L)
\end{aligned}
$$

The third equality follows from integration by parts of the first and second terms and Assumption 12, the fourth equality follows from integration by parts of the second term and Assumption 12 and the final equality follows from the expansion of $f''(x+uh_0)$ around $h_0 = 0$ and Assumptions 2, 10. □

Lemma 3. *Under Assumptions 1, 2, 4,5, 10, and 12,*

$$
\mathbb{E}\left|\sqrt{nh_0}\left(\frac{\hat{h}-h_0}{h_0}\right)^2 \Gamma_{KDE_2}\right| = o\{(nh_0)^{-1}\}
$$

Proof. Similar to the proof of Theorem 1. □

Lemma 4. *Under Assumptions 1, 2, 4, 6,10, and 12,*

$$
\mathbb{E}[\Gamma_1\mathcal{L}_1] = O(h_0^{L+1})
$$

Proof. Letting $g(x) = f^{(L)}(x)f(x)$

$$
\begin{aligned}
\mathbb{E}[\Gamma_1\mathcal{L}_1] &= \mathbb{E}\left[\left\{K'\left(\frac{X_1-x}{h_0}\right)\left(\frac{X_1-x}{h_0}\right) + K\left(\frac{X_1-x}{h_0}\right) - \mathbb{E}\left[K'\left(\frac{X_1-x}{h_0}\right)\left(\frac{X_1-x}{h_0}\right) + K\left(\frac{X_1-x}{h_0}\right)\right]\right\}\mathcal{L}_1\right] \\
&= \mathbb{E}\left[\left\{K'\left(\frac{X_1-x}{h_0}\right)\left(\frac{X_1-x}{h_0}\right) + K\left(\frac{X_1-x}{h_0}\right)\right\} f^{(L)}(X_1)\right] - \mathbb{E}\left[\left\{K'\left(\frac{X_1-x}{h_0}\right)\left(\frac{X_1-x}{h_0}\right) + K\left(\frac{X_1-x}{h_0}\right)\right\}\right]\mathbb{E}[f^{(L)}(X_1)]
\end{aligned}
$$

We can compute the first term as follow.

$$
\begin{aligned}
&\mathbb{E}\left[\left\{K'\left(\frac{X_1-x}{h_0}\right)\left(\frac{X_1-x}{h_0}\right) + K\left(\frac{X_1-x}{h_0}\right)\right\} f^{(L)}(X_1)\right] \\
&= \int \left\{K'\left(\frac{z-x}{h_0}\right)\left(\frac{z-x}{h_0}\right) + K\left(\frac{z-x}{h_0}\right)\right\} f^{(L)}(z)f(z)dz \\
&= h_0\int \left\{K'(u)u + K(u)\right\} g(x+uh_0)du \\
&= h_0\int \left\{K'(u)u + K(u)\right\}\left\{g(x) + \cdots + \frac{g^{(L)}(x)}{L!}(uh_0)^L + o(h_0^L)\right\} du
\end{aligned}
$$

$$
\begin{aligned}
&= -h_0 \int K(u)l(x)du + h_0 \int K(u)l(x)du + h_0 \int K'(u)u \frac{g^{(L)}(x)}{L!}(uh_0)^L du + h_0 \int K(u) \frac{g^{(L)}(x)}{L!}(uh_0)^l du + o(h_0^{L+1}) \\
&= \frac{g^{(L)}(x)}{L!} h_0^{L+1} \int K'(u)u^{L+1} du + \frac{g^{(L)}(x)}{L!} h_0^{L+1} \int K(u)u^L du + o(h_0^{L+1}) \\
&= -(L+1) \frac{g^{(L)}(x)}{L!} h_0^{L+1} \int K(u)u^L du + \frac{g^{(L)}(x)}{L!} h_0^{L+1} \int K(u)u^L du + o(h_0^{L+1}) \\
&= O(h_0^{L+1})
\end{aligned}
$$

The fourth equality follows from the expansion of $l(x+uh_0)$ around $h_0 = 0$ and Assumption 6, and the fifth equality follows from integration by parts of the products of $K'(u)u$ and $l^{(k)}(x)u^k, (0 \leqq k \leqq L-1)$ and Assumption 6 and 12. Next, we can compute the second term similarly to the first term.

$$
\begin{aligned}
&\mathbb{E}[\left\{K'\left(\frac{X_1-x}{h_0}\right)\left(\frac{X_1-x}{h_0}\right) + K\left(\frac{X_1-x}{h_0}\right)\right\}]\mathbb{E}[f^{(L)}(X_1)] \\
&= \left(-(L+1)\frac{f^{(L)}(x)}{L!} h_0^{L+1} \int K(u)u^L du + \frac{f^{(L)}(x)}{L!} h_0^{L+1} \int K(u)u^L du\right)\mathbb{E}[f^{(L)}(X_1)] \\
&= O(h_0^{L+1})
\end{aligned}
$$

These imply the lemma holds. □

Lemma 5. *For any positive integer k and any non-negative integer l,*

$$\mathbb{E}|S_1^k \mathcal{L}_1^l| = O(h_0)$$

Proof. Straightforward. □

Lemma 6. *For any positive integer k, l,*

$$\mathbb{E}|S_1^k \Gamma_1^l| = O(h_0)$$

Proof. Straightforward. □

Lemma 7. *For any positive integer $k, l \geqq 2$,*

$$\mathbb{E}|\Gamma_1^l|^k = O(h_0^k), \quad \mathbb{E}|\mathcal{L}_1^l|^k = O(1)$$

Proof. Straightforward. □

Lemma 8. *For any positive integer r,*

$$\mathbb{E}|\Lambda_1(x)|^r = O(h_0^{\frac{r(2L+1)}{2}})$$

Proof. From Lemma 7, for any positive integer k, and some positive bounded function $C(x)$,

$$
\begin{aligned}
\mathbb{E}|\Lambda_1(x)|^{2k} = \mathbb{E}\Lambda_1(x)^{2k} &\lesssim \frac{h_0^{k(2L+1)}}{n^k}\mathbb{E}\left[\left(\sum_{i=1}^n \mathcal{L}_i\right)^{2k}\right] + s.o. \\
&= \frac{h_0^{k(2L+1)}}{n^k} n^k \mathbb{E}\left[\mathcal{L}_i^2\right]^k + s.o. = O(h_0^{k(2L+1)})
\end{aligned}
$$

From Holder's inequality, for $0 < r < s$, $\mathbb{E}|X|^r \leqq \{\mathbb{E}|X|^s\}^{r/s}$, thus for any positive integer k,

$$\mathbb{E}|\Lambda_1(x)|^{2k-1} \leqq \{\mathbb{E}|\Lambda_1(x)|^{2k}\}^{\frac{2k-1}{2k}} = O(h_0^{\frac{(2k-1)(2L+1)}{2}})$$

This implies the lemma holds. □

Lemma 9. *For any positive integer r,*

$$\mathbb{E}|\Lambda_2(x)|^r = O(n^{-r/2})$$

Proof. From Lemma 7, for any positive integer k,

$$\begin{aligned}\mathbb{E}|\Lambda_2(x)|^{2k} &\lesssim \frac{1}{n^{3k}h_0^k\mu_{20}^k}\mathbb{E}\left[\left(\sum_{i=1}^n\sum_{j\neq i}^n \Gamma_i\mathcal{L}_j\right)^{2k}\right] \\ &= \frac{n^k(n-1)^k}{n^{3k}h_0^k\mu_{20}^k}\mathbb{E}\left[\Gamma_1^2\right]^k\mathbb{E}\left[\mathcal{L}_2^2\right]^k + s.o. = O(n^{-k})\end{aligned}$$

Then, similarly to the evaluation of $\mathbb{E}|\Lambda_1(x)|^r$, the lemma holds. □

Lemma 10.

$$\mathbb{E}|\Lambda_1(x)\Lambda_2(x)| = O(n^{-1/2}h_0^{\frac{2L+1}{2}})$$

Proof. Lemma 8, 9 and Holder inequality implies

$$\mathbb{E}|\Lambda_1(x)\Lambda_2(x)| \leqq \mathbb{E}|\Lambda_1(x)|\mathbb{E}|\Lambda_2(x)| = O(h_0^{\frac{2L+1}{2}})O(n^{-1/2})$$

□

Lemma 11.

$$\mathbb{E}|\Lambda_3(x)| = O(n^{-1/2}h_0^{\frac{2L+1}{2}} + n^{-1})$$

Proof. From Lemma 4,

$$\begin{aligned}\mathbb{E}\Lambda_3(x)^2 &\lesssim \frac{1}{n^3h_0\mu_{20}}\mathbb{E}\left[\sum_{i=1}^n\sum_{j=1}^n \Gamma_i\mathcal{L}_i\Gamma_j\mathcal{L}_j\right] \\ &= \frac{1}{n^3h_0\mu_{20}}\mathbb{E}\left[\sum_{i=1}^n\sum_{j\neq i}^n \Gamma_i\mathcal{L}_i\Gamma_j\mathcal{L}_j + \sum_{i=1}^n\Gamma_i^2\mathcal{L}_i^2\right] \\ &= \frac{n(n-1)}{n^3h_0\mu_{20}}\mathbb{E}[\Gamma_1\mathcal{L}_1]^2 + \frac{1}{n^2h_0\mu_{20}}\mathbb{E}[\Gamma_1^2\mathcal{L}_1^2] \\ &= O(n^{-1}h_0^{-1})O(h_0^{2(L+1)}) + O(n^{-2}h_0^{-1})O(h_0)\end{aligned}$$

Similarly to the evaluation of $\mathbb{E}|\Lambda_1(x)|^r$, the lemma holds. □

Lemma 12.

$$\mathbb{E}|\Lambda_4(x)|^2 = O(n^{-1}h_0)$$

Proof. The proof is similar to Lemma 9. □

Lemma 13.

$$\mathbb{E}|\Lambda_5(x)|^2 = O(n^{-1}h_0)$$

Proof. The proof is similar to Lemma 11. □

Lemma 14.

$$\mathbb{E}|\Lambda_1(x)\Lambda_4(x)| = O(n^{-1/2}h_0^{L+1})$$

Proof. From Cauchy-Schwarz inequality and Lemma 8 and 12, this lemma holds. □

Lemma 15.

$$\mathbb{E}|\Lambda_1(x)\Lambda_5(x)| = O(n^{-1/2}h_0^{L+1})$$

Proof. From Cauchy-Schwarz inequality and Lemma 8 and 13, this lemma holds. □

Lemma 16.

$$\mathbb{E}|\Lambda_2(x)\Lambda_4(x)| = O(n^{-1}h_0^{1/2})$$

Proof. From Cauchy-Schwarz inequality and Lemma 9 and 12, this lemma holds. □

Lemma 17.

$$\mathbb{E}|\Lambda_2(x)\Lambda_5(x)| = O(n^{-1}h_0^{1/2})$$

Proof. From Cauchy-Schwarz inequality and Lemma 9 and 13, this lemma holds. □

Lemma 18.

$$\mathbb{E}|\Lambda_4(x)\Lambda_5(x)| = O(n^{-1}h_0)$$

Proof. From Cauchy-Schwarz inequality and Lemma 12 and 13, this lemma holds. □

Lemma 19.

$$\mathbb{E}|\Lambda_1(x;m)|^2 = O\left(\frac{mh_0^{2L+1}}{n}\right)$$

Proof. From Lemma 7,

$$\begin{aligned}
&\mathbb{E}\,|\Lambda_1(x;m)|^2 \\
&\quad \lesssim \frac{h_0^{2L+1}}{n}\mathbb{E}\left[\sum_{i=1}^{m}\sum_{j=1}^{m}\mathcal{L}_i\mathcal{L}_j\right] = \frac{h_0^{2L+1}}{n}\mathbb{E}\left[\sum_{i=1}^{m}\sum_{j\neq i}^{m}\mathcal{L}_i\mathcal{L}_j + \sum_{i=1}^{m}\mathcal{L}_i^2\right] = O\left(\frac{mh_0^{2L+1}}{n}\right)
\end{aligned}$$

□

Lemma 20.

$$\mathbb{E}|\Lambda_2(x;m)|^2 = O\left(\frac{m^2}{n^3}\right)$$

Proof.

$$\begin{aligned}
&\mathbb{E}|\Lambda_2(x;m)|^2 \\
&\quad \lesssim \frac{1}{n^3h_0}\mathbb{E}\left[\sum_{i=1}^{m}\sum_{j\neq i}^{m}\sum_{k=1}^{m}\sum_{l\neq k}^{m}\Gamma_i\mathcal{L}_j\Gamma_k\mathcal{L}_l\right] = \frac{m(m-1)}{n^3h_0}\mathbb{E}[\Gamma_1^2]\mathbb{E}[\mathcal{L}_2^2] = O\left(\frac{m^2}{n^3}\right)
\end{aligned}$$

□

Lemma 21.

$$\mathbb{E}|\Lambda_4(x;m)|^2 = O\left(\frac{m^2}{n^3}\right)$$

Proof. Proof is similar to Lemma 20 □

Lemma 22.

$$\mathbb{E}|\Lambda_5(x;m)|^2 = O\left(\frac{m}{n^3}\right)$$

Proof.

$$\begin{aligned}
\mathbb{E}|\Lambda_5(x;m)|^2 &\lesssim \frac{1}{n^3 h_0}\mathbb{E}\left[\sum_{i=1}^{m}\sum_{j=1}^{m} S_i\mathcal{L}_i S_j\mathcal{L}_j\right] \\
&= \frac{1}{n^3 h_0}\mathbb{E}\left[\sum_{i=1}^{m}\sum_{j\neq i}^{m} S_i\mathcal{L}_i S_j\mathcal{L}_j + \sum_{i=1}^{m} S_i^2\mathcal{L}_i^2\right] \\
&= \frac{m}{n^3 h_0}\mathbb{E}[S_1^2\mathcal{L}_1^2] = O\left(\frac{m}{n^3}\right)
\end{aligned}$$

□

Lemma 23.

$$\begin{aligned}
\mathbb{E}|\Lambda_1(x;m)\Lambda_2(x;m)| &= O\left(\frac{m^3 h_0^{2L+1}}{n^4}\right)^{1/2} = O\left(\frac{m^{3/2} h_0^{(2L+1)/2}}{n^2}\right) \\
\mathbb{E}|\Lambda_1(x;m)\Lambda_4(x;m)| &= O\left(\frac{m^3 h_0^{2L+1}}{n^4}\right)^{1/2} = O\left(\frac{m^{3/2} h_0^{(2L+1)/2}}{n^2}\right) \\
\mathbb{E}|\Lambda_1(x;m)\Lambda_5(x;m)| &= O\left(\frac{m^2 h_0^{2L+1}}{n^4}\right)^{1/2} = O\left(\frac{m h_0^{(2L+1)/2}}{n^2}\right) \\
\mathbb{E}|\Lambda_2(x;m)\Lambda_4(x;m)| &= O\left(\frac{m^4}{n^6}\right)^{1/2} = O\left(\frac{m^2}{n^3}\right) \\
\mathbb{E}|\Lambda_2(x;m)\Lambda_5(x;m)| &= O\left(\frac{m^3}{n^6}\right)^{1/2} = O\left(\frac{m^{3/2}}{n^3}\right) \\
\mathbb{E}|\Lambda_4(x;m)\Lambda_5(x;m)| &= O\left(\frac{m^3}{n^6}\right)^{1/2} = O\left(\frac{m^{3/2}}{n^3}\right)
\end{aligned}$$

Proof. From Cauchy-Schwarz inequality and Lemma 19, 20, 21 and 22, this lemma holds. □

C Derivation of Expression for $p_1(z), p_3(z)$ and $p_4(z)$

For $p_1(z)$, we have,

$$\begin{aligned}
p_1(z) &= \frac{-1}{6}\mu_{30}\mu_{20}^{-3/2}(z^2-1) \\
&= \frac{-1}{6}\frac{\kappa_{03}f(x) - 3\kappa_{02}f(x)^2 h_0 + \kappa_{22}f^{(2)}(x)h_0^2/2 + o(h_0^2)}{[\kappa_{02}f(x) - f(x)^2 h_0 + \{\kappa_{23}f^{(2)}(x)/2 + 2f(x)^3\}h^2 + o(h_0^2)]^{3/2}}(z^2-1) \\
&= \frac{-1}{6}[\kappa_{03}f(x) - 3\kappa_{02}f(x)^2 h_0 + \kappa_{22}f^{(2)}(x)h_0^2/2 + o(h_0^2)]
\end{aligned}$$

$$
\begin{aligned}
&\times [\{\kappa_{02} f(x)\}^{-3/2} \\
&\qquad - \frac{3}{2}\{\kappa_{02} f(x)\}^{-5/2}(f(x)^2 h_0 - \{\frac{\kappa_{23} f^{(2)}(x)}{2} + 2f(x)^3\}h_0^2) \\
&\qquad + \frac{15}{8}\{\kappa_{02} f(x)\}^{-7/2} f(x)^4 h_0^2](z^2-1) + o(h_0^2) \\
= &\frac{-1}{6}\Bigg[\kappa_{02}^{-3/2}\kappa_{03} f(x) - 3\left\{\frac{f(x)^{1/2}}{\kappa_{02}^{1/2}} - \frac{\kappa_{03} f(x)^{1/2}}{2\kappa_{02}^{5/2}}\right\}h_0 \\
&\quad + \left\{\frac{-3}{4}\{\kappa_{02} f(x)\}^{-5/2}\kappa_{03}\kappa_{23} f^{(2)}(x) f(x) - 3\{\kappa_{02} f(x)\}^{-5/2}\kappa_{03} f(x)^4\right. \\
&\qquad \left. + \frac{15}{8}\{\kappa_{02} f(x)\}^{-7/2}\kappa_{03} f(x)^5 + \frac{9}{2}\kappa_{02}^{-3/2} f(x)^{3/2}\right\}h_0^2\Bigg](z^2-1) + o(h_0^2) \\
\equiv &\gamma_{1,0}(x)(z^2-1) + \gamma_{1,1}(x)(z^2-1)h_0 + \gamma_{1,2}(x)(z^2-1)h_0^2 + o(h_0^2).
\end{aligned}
$$

For $p_{3,0}(z)$, we have

$$
\begin{aligned}
p_{3,0}(z) &= -C_{PI} C_{\Gamma,0}(x)\rho_{11}\mu_{20}^{-1} z \\
&= -C_{PI} C_{\Gamma,0}(x)\frac{\mathcal{L}(x) f(x) + O(h_0^L)}{[\kappa_{02} f(x) - f(x)^2 h_0 + o(h_0)]^{-1}} z \\
&= -C_{PI} C_{\Gamma,0}(x)[\mathcal{L}(x) f(x) + O(h_0^L)] \\
&\qquad \times [\{\kappa_{02} f(x)\}^{-1} - \{\kappa_{02} f(x)\}^{-2}\left(f(x)^2 h_0\right)]z + o(h_0) \\
&= -C_{PI} C_{\Gamma,0}(x)\kappa_{02}^{-1}\mathcal{L}(x) z + C_{PI} C_{\Gamma,0}(x)\kappa_{02}^{-2}\mathcal{L}(x) f(x) z h_0 + o(h_0) \\
&\equiv \gamma_{3,1,0}(x) z + \gamma_{3,1,1}(x) z h_0 + o(h_0),
\end{aligned}
$$

while for $p_4(z)$,

$$
\begin{aligned}
p_4(z) &= -C_{PI}\rho_{11}\xi_{11}\mu_{20}^{-3/2}(z^2-1) + \frac{1}{2}C_{PI}\rho_{11}\mu_{20}^{-1/2} z^2 \\
&= -C_{PI}\frac{\mathcal{L}(x) f(x)\{\tau_0 f(x) + o(h_0)\}}{[\kappa_{02} f(x) - f(x)^2 h_0 + o(h_0)]^{3/2}}(z^2-1) + \frac{1}{2}C_{PI}\frac{\mathcal{L}(x) f(x) + O(h_0^L)}{[\kappa_{02} f(x) - f(x)^2 h_0 + o(h_0)]^{1/2}} z^2 \\
&= -C_{PI}\mathcal{L}(x) f(x)\{\tau_0 f(x) + o(h_0)\} \\
&\qquad \times [\{\kappa_{02} f(x)\}^{-3/2} - \frac{3}{2}\{\kappa_{02} f(x)\}^{-5/2} f(x)^2 h_0 + o(h_0)\}](z^2-1) \\
&+ \frac{1}{2}C_{PI}\mathcal{L}(x) f(x) \\
&\qquad \times [\{\kappa_{02} f(x)\}^{-1/2} - \frac{1}{2}\{\kappa_{02} f(x)\}^{-3/2} f(x)^2 h_0 + o(h_0)\}] z^2 \\
&= -C_{PI}\kappa_{02}^{-3/2}\tau_0\mathcal{L}(x) f(x)^{1/2}(z^2-1) + \frac{3}{2}C_{PI}\kappa_{02}^{-5/2}\tau_0\mathcal{L}(x) f(x)^{3/2}(z^2-1)h_0 \\
&\qquad + \frac{1}{2}C_{PI}\kappa_{02}^{-1/2}\mathcal{L}(x) f(x)^{1/2} z^2 - \frac{1}{4}C_{PI}\kappa_{02}^{-3/2}\mathcal{L}(x) f(x)^{3/2} z^2 h_0 + o(h_0) \\
&\equiv \{\gamma_{4,1,0}(x)(z^2-1) + \gamma_{4,2,0}(x) z^2\} + \{\gamma_{4,1,1}(x)(z^2-1) + \gamma_{4,2,1}(x) z^2\}h_0 + o(h_0).
\end{aligned}
$$

□

References

[1] Esseen, C. G. (1945). Fourier Analysis of Distribution Functions: A Mathematical Study of the Laplace-Gaussian Law. *Acta Mathematica, 77*, 1-125.

[2] Feller, W. (1971). *An Introduction to the Theory of Probability and Its Applications*, vol. II. New York: Wiley.

[3] Hall, P., and Marron, J. S. (1987). Estimation of Integrated Squared Density Derivatives. *Statistics and Probability Letters, 6*, 109–15.

[4] Nishiyama, Y., and Robinson, P. M. (2000). Edgeworth Expansions for Semiparametric Averaged Derivatives. *Econometrica, 68*, 931–79.