NONPARAMETRIC ESTIMATION OF ADDITIVE MODELS WITH ERRORS-IN-VARIABLES

HAO DONG, TAISUKE OTSU, AND LUKE TAYLOR

Abstract. In the estimation of nonparametric additive models, conventional methods, such as backfitting and series approximation, cannot be applied when measurement error is present in a covariate. This paper proposes an estimator for such models by extending Horowitz and Mammen (2004)'s two-stage estimator to the errors-in-variables case. In the first stage, to adapt to the additive structure, we use a series approximation together with a ridge approach to deal with the ill-posedness brought by mismeasurement. We derive the uniform convergence rate of this first-stage estimator and characterize how the measurement error slows down the convergence rate for ordinary/super smooth cases. To establish the limiting distribution, we construct a second-stage estimator via one-step backfitting with a deconvolution kernel using the first-stage estimator. The asymptotic normality of the second-stage estimator is established for ordinary/super smooth measurement error cases. Finally, a Monte Carlo study and an empirical application highlight the applicability of the estimator.

1. Introduction

Since their inception, nonparametric additive regression models have received much attention in the econometrics and statistics literature (see, e.g., Horowitz, 2014, for a review). Their popularity is primarily driven by their ability to overcome the curse of dimensionality through imposing additive separability of covariates. Furthermore, this separability ensures easy interpretation and is a natural and realistic assumption in many economic models. For example, constant elasticity of substitution production functions take this form (Leontief, 1947), as do many models of consumer behavior (Deaton and Muellbauer, 1980).

In some situations, the curse of dimensionality can be particularly severe; measurement error in covariates is one such situation. In that case, the degree to which the convergence rate of nonparametric estimators deteriorates as the dimension of the covariates increases is less favorable. Unfortunately, economic data is often subject to contamination. Indeed, the frequent use of imprecise measurements of complex variables such as GDP and inflation, the reliance on survey data, and the inability to accurately measure intangible variables such as cognitive ability all lead to measurement error (see, e.g., Bound, Brown and Mathiowetz, 2001, Hu, 2017, and Schennach, 2020, for surveys in econometrics). Thus, nonparametric additive models can be particularly useful when dealing with contaminated data. Empirical examples of additive models which could have benefited from acknowledging measurement error include Xu and Lin (2015, 2016), who examine the impact of industrialization on carbon dioxide emissions, where industrialization is measured as industry value added as a proportion of GDP; and Dominici *et*

The authors would like to thank anonymous referees for helpful comments.

al. (2002), who study the health effects of air pollution, where air pollution is widely held to be imprecisely measured.

In answer to these issues, this paper studies estimation of the nonparametric additive regression model with a mismeasured covariate:

$$Y = \mu + g(X^*) + m_1(Z_1) + \cdots + m_D(Z_D) + U, \tag{1.1}$$

where Y is a response variable, μ is an unknown intercept, X^* is an error-free but unobservable covariate, $Z = (Z_1, \ldots, Z_D)$ are observable covariates, U is an error term, and $(g, m_1, \ldots, m_D)$ are unknown functions to be estimated. If X^* is observable, it is a standard nonparametric additive model with the identity link function, which has been well studied in the literature; see, e.g., Stone (1985, 1986), Buja, Hastie and Tibshirani (1989), Linton and Nielsen (1995), Linton and Härdle (1996), Opsomer and Ruppert (1997), Mammen, Linton and Nielsen (1999), Opsomer (2000), Horowitz and Mammen (2004), and Ozabaci, Henderson and Su (2014). However, when X^* is mismeasured, these conventional methods to estimate the unknown functions are generally inconsistent.

We consider estimation of the nonparametric additive regression model in (1.1) when the measurement X on X^* involves a classical measurement error. More precisely, throughout the paper, we assume that the measurement X is generated by

$$X = X^* + \epsilon, \tag{1.2}$$

where X^* is scalar, and ϵ is a measurement error independent of X^*. In Remarks 1 and 3 at the end of Section 2, we discuss generalizations to relax these assumptions.

We develop an estimator for the unknown functions $g, m_1, \ldots, m_D$ and intercept μ using the observables (Y, X, Z) generated by (1.1) and (1.2) and study its asymptotic properties. In particular, we extend the two-stage approach of Horowitz and Mammen (2004) to deal with the measurement error using deconvolution techniques. In the first stage, Horowitz and Mammen (2004) estimated the unknown functions by a series approximation method. In the presence of measurement error, the coefficients in the series approximation are estimated by the ridge-based regularized estimator as in Hall and Meister (2007). In the second stage, Horowitz and Mammen (2004) implemented one-step backfitting based on local linear regression to achieve asymptotic normality of the estimator. In our case, this step is carried out using a nonparametric deconvolution kernel regression.

There is an extensive literature on nonparametric additive models when all covariates are accurately measured; see the papers cited above. A rare exception to assuming the accurate measurement of covariates is a recent paper by Han and Park (2018). In particular, they also focus on classical measurement error, and develop a new estimator for additive models by extending the smoothed backfitting approach of Mammen, Linton and Nielsen (1999). However, there are two major differences between our work and theirs. First, our second-stage estimator achieves asymptotic normality, which is useful for statistical inference, while they only derive the convergence rate of their estimator. Moreover, our first-stage estimator converges at a faster rate than theirs. Second, our two-stage estimator can handle both ordinary smooth and supersmooth errors, while their method cannot be easily adapted to the case of supersmooth measurement error, which is a particularly important class of error distribution as it includes the normal distribution. As such, this paper contributes to the literature on nonparametric additive models by developing the first estimator that achieves asymptotic normality in the face of measurement error in a covariate, and can handle supersmooth errors.

We also contribute to the literature on nonparametric deconvolution methods for measurement error models. In particular, we employ the ridge-based regularization method by Hall and Meister (2007) to estimate moments involving error-free unobservable covariates. Also, for backfitting in the second stage, we apply a nonparametric deconvolution kernel regression; see, e.g., Stefanski and Carroll (1990), Carroll and Hall (1988), Fan (1991a, 1991b), Fan and Masry (1992), Fan and Truong (1993), Delaigle, Hall and Meister (2008), and Hall and Lahiri (2008).

The rest of this paper is organized as follows. Section 2 introduces the basic setup and develops our two-stage estimator. Section 3 presents our main results: we derive the convergence rate of the first-stage estimator in Section 3.1, and establish the limiting distribution of the second-stage estimator in Section 3.2. We focus on the case when the measurement error distribution is known to researchers in Sections 2 and 3, and relax the known error distribution assumption using auxiliary data in Section 4. Sections 5 and 6 present a simulation study and an empirical application, respectively. Finally, Section 7 concludes. All proofs are contained in the Appendix.

Notation. Throughout the paper, let $\|f\|_2 = \left(\int |f(w)|^2 dw\right)^{1/2}$ be the L_2-norm of a function $f : \mathbb{R} \to \mathbb{C}$, $L_2(\mathbb{R}) = \{f : \|f\|_2 < \infty\}$ be the L_2-space, and $\langle f_1, f_2 \rangle = \int f_1(w)\overline{f_2(w)}dw$ be the inner product in $L_2(\mathbb{R})$, where $\bar{c}$ denotes the complex conjugate of $c \in \mathbb{C}$, $f^{\mathrm{ft}}(t) = \int f(x)e^{\mathrm{i}tx}dx$ be the Fourier transform of f with $\mathrm{i} = \sqrt{-1}$. Also, let $\|A\| = [\mathrm{tr}(A^\dagger A)]^{1/2}$ be the Frobenius norm of a complex matrix A, where $A^\dagger$ denotes A's conjugate transpose, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ be the largest and smallest eigenvalues of a Hermite matrix A, respectively, and $\delta_{k,k'}$ be the Kronecker delta, which is equal to 0 if $k \neq k'$, and equal to 1 if $k = k'$.

2. Setup and estimator

Before presenting our estimator, we first show that the functions $g, m_1, \ldots, m_D$ and intercept μ in the model (1.1) can be identified from the distribution of observables (Y, X, Z). To this end, we impose the following assumptions.

Assumption 1. **(1):** *ϵ is independent of (Y, X^*, Z), the density f_ϵ of ϵ is known, and $f_\epsilon^{\mathrm{ft}}(t) \neq 0$ for all $t \in \mathbb{R}$.*

(2): *The density $f_{X^*,Z}$ of (X^*, Z) is bounded away from zero on $\mathcal{I} \times [-1,1]^D$, where $\mathcal{I}$ is a known compact subset of the support of X^*, and $[-1,1]$ is the support of Z_d for $d = 1, \ldots, D$.*

(3): *$E[U|X^*, Z] = 0$ and $g, m_1, \ldots, m_D$ are normalized as*

$$\int_{\mathcal{I}} g(w)dw = \int_{-1}^{1} m_1(w)dw = \cdots = \int_{-1}^{1} m_D(w)dw = 0. \tag{2.1}$$

Assumption 1 (1) claims the measurement error is classical in nature and f_ϵ^{ft} is non-vanishing everywhere, which is commonly used in the literature on nonparametric estimation with measurement error (see, Meister, 2009, for a review). The assumption f_ϵ being known is restrictive in economic applications. In Section 4, we discuss how to relax this by using repeated measurements on X^*. Assumption 1 (3) contains normalizations required for identification.

Assumption 1 (2) requires all covariates to be continuously distributed. In Remark 2 below, we discuss how to include correctly measured discrete covariates. As in Horowitz and Mammen (2004), we assume that the observable covariates Z are supported on $[-1,1]^D$. This is an innocuous assumption as an invertible transformation of Z that satisfies it can be used in place of Z. However, this argument fails for the unobservable covariate X^*. Indeed, such a transformation does not preserve the additive structure in (1.2) except when the transformation is linear. Thus, even though the distribution of ϵ is known, it is difficult to recover the distribution of X^* from the transformation of X through deconvolution. Also, since the support of ϵ is typically unknown, so too is the support of X^*. With these considerations, we do not impose any condition on the support of X, X^* or ϵ, but focus on estimation of the function g over some known compact set $\mathcal{I}$ of interest. $f_{X^*,Z}$ is assumed to be bounded away from zero on $\mathcal{I}\times[-1,1]^D$ so that the conditional expectations (on the event $(X^*,Z)\in\mathcal{I}\times[-1,1]^D$) are well defined.

Under Assumption 1, all unknown objects in model (1.1) are identified. This result is summarized in Theorem 1 as follows.

Theorem 1. *Under Assumption 1, the functions $g, m_1,\ldots,m_D$ and intercept μ are identified.*

This theorem follows from an application of the marginal integration argument for the nonparametric additive model combined with the identification of the density $f_{Y,X^*,Z}$ of (Y,X^*,Z) based on deconvolution methods. The proof is given in Appendix A.

We now introduce our estimation strategy. For expository purposes, we tentatively assume that the error-free covariate X^* is observed. To estimate μ, m_d over $[-1,1]$, and g over the subset $\mathcal{I}$ under the normalization in (2.1), the first-stage estimation of Horowitz and Mammen (2004) is used by minimizing

$$\sum_{j=1}^n \mathbb{I}\{X_j^*\in\mathcal{I}\}\left[Y_j-\mu-\sum_{k=1}^{\kappa}p_k(X_j^*)\theta_k^0-\sum_{d=1}^{D}\sum_{k=1}^{\kappa}q_k(Z_{d,j})\theta_k^d\right]^2, \tag{2.2}$$

with respect to $\theta=(\mu,\theta_1^0,\ldots,\theta_\kappa^0,\theta_1^1,\ldots,\theta_\kappa^1,\ldots,\theta_1^D,\ldots,\theta_\kappa^D)'$, where $\mathbb{I}\{\cdot\}$ is the indicator function, $\{p_k\}_{k=1}^\infty$ and $\{q_k\}_{k=1}^\infty$ are basis functions supported on $\mathcal{I}$ and $[-1,1]$, respectively, and κ is a tuning parameter characterizing the accuracy of the series approximation. The trimming term $\mathbb{I}\{X_j^*\in\mathcal{I}\}$ appears because we are only interested in estimating g over $\mathcal{I}$.

If X^* is mismeasured, the unobservability of X^* renders this method infeasible. Also, replacing X_j^* with the observable X_j and applying least squares estimation for the above criterion would yield inconsistent estimates in general. To estimate θ in (2.2), we consider the population counterpart of (2.2), that is

$$E[\mathbb{I}\{X^*\in\mathcal{I}\}Y^2]+\theta' E[P_\kappa P_\kappa']\theta-2E[YP_\kappa']\theta, \tag{2.3}$$

where $P_\kappa = (p_0(X^*), p_1(X^*), \ldots, p_\kappa(X^*), q_{01}(Z_1), \ldots, q_{0\kappa}(Z_1), \ldots, q_{01}(Z_D), \ldots, q_{0\kappa}(Z_D))'$ with $p_0(X^*) = \mathbb{I}\{X^* \in \mathcal{I}\}$ and $q_{0k}(Z_d) = p_0(X^*)q_k(Z_d)$ for $k = 1, \ldots, \kappa$ and $d = 1, \ldots, D$. Thus, once we have estimators for $E[P_\kappa P_\kappa']$ and $E[YP_\kappa']$, denoted $\hat{E}[P_\kappa P_\kappa']$ and $\hat{E}[YP_\kappa']$, respectively, θ can be estimated by

$$\hat{\theta} = (\Re \hat{E}[P_\kappa P_\kappa'])^{-1} \Re \hat{E}[YP_\kappa'], \tag{2.4}$$

where $\Re\{\cdot\}$ denotes the real part of a complex-valued matrix or vector, and the inverse here may be the Moore-Penrose inverse. Based on this, the first-stage estimators of g and m_d for $d = 1, \ldots, D$ are given by

$$\hat{g}(x^*) = \sum_{k=1}^{\kappa} p_k(x^*)\hat{\theta}_k^0, \qquad \hat{m}_d(z_d) = \sum_{k=1}^{\kappa} q_k(z_d)\hat{\theta}_k^d. \tag{2.5}$$

To implement the estimator in (2.5) based on (2.4), we must estimate elements of $E[P_\kappa P_\kappa']$ and $E[YP_\kappa']$.

We first consider estimation of $E[Yp_k(X^*)]$, which appears in $E[YP_\kappa']$. By Plancherel's isometry (see Lemma 1 (1) in Appendix D), $E[Yp_k(X^*)]$ can be expressed as

$$\begin{aligned} E[Yp_k(X^*)] &= \langle mf_{X^*}, p_k \rangle = \frac{1}{2\pi} \left\langle [mf_{X^*}]^{\mathrm{ft}}, p_k^{\mathrm{ft}} \right\rangle \\ &= \frac{1}{2\pi} \int E[Ye^{\mathrm{i}tX}] \frac{p_k^{\mathrm{ft}}(-t)}{f_\epsilon^{\mathrm{ft}}(t)} dt, \end{aligned}$$

where $m(\cdot) = E[Y|X^* = \cdot]$, and the last equality follows by the law of iterated expectations and Assumption 1 (1). A naive estimator of this moment could be obtained by replacing $E[Ye^{\mathrm{i}tX}]$ by its sample analog $n^{-1}\sum_{j=1}^n Y_j e^{\mathrm{i}tX_j}$. However, it is well known that $n^{-1}\sum_{j=1}^n Y_j e^{\mathrm{i}tX_j}/f_\epsilon^{\mathrm{ft}}(t)$ is a poor estimator for $f_{X^*}^{\mathrm{ft}}(t)$ in the tails. Intuitively, the estimation error of the sample analog can be severely amplified in the tails, such that the above integral may not be well-defined. Some form of regularization is commonly introduced in such a situation. Here we employ the ridge approach of Hall and Meister (2007) and suggest estimating $E[Yp_k(X^*)]$ by

$$\hat{E}[Yp_k(X^*)] = \frac{1}{2\pi} \int \left(\frac{1}{n} \sum_{j=1}^n Y_j e^{\mathrm{i}tX_j} \right) \frac{p_k^{\mathrm{ft}}(-t) f_\epsilon^{\mathrm{ft}}(-t) |f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} dt, \tag{2.6}$$

where $r \geq 0$ is a tuning parameter to control the smoothness of the integrand and $n^{-\zeta}$ with $\zeta > 0$ is a ridge term to keep the denominator away from zero.

Also note that the moments $E[Yp_k(X^*)]$ and $E[Yq_{0k}(Z_d)]$, which appear in $E[YP_\kappa']$, and $E[p_k(X^*)]$, $E[q_{0k}(Z_d)]$, $E[p_k(X^*)p_l(X^*)]$, $E[p_k(X^*)q_{0l}(Z_d)]$ and $E[q_{0k}(Z_d)q_{0l}(Z_{d'})]$, which appear in $E[P_\kappa P_\kappa']$, can all be expressed in a general form $E[p(X^*)Q]$ by different choices of a known function p and a random variable Q, where Q is a known function of Y and Z. For example, $E[p(X^*)Q]$ equals $E[p_k(X^*)q_{0l}(Z_d)]$ when $p(X^*) = p_k(X^*)$ and $Q = q_l(Z_d)$. By a similar argument, the estimator of $E[p(X^*)Q]$ is constructed as

$$\hat{E}[p(X^*)Q] = \frac{1}{2\pi} \int \left(\frac{1}{n} \sum_{j=1}^n Q_j e^{\mathrm{i}tX_j} \right) \frac{p^{\mathrm{ft}}(-t) f_\epsilon^{\mathrm{ft}}(-t) |f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} dt.$$

Then, by applying these estimators to elements in (2.4), we obtain the first-stage estimator (2.5).

In the literature on nonparametric deconvolution methods, the kernel approach is more frequently used than the ridge. However, the kernel-based method is not adaptive. This is because, to obtain the optimal convergence rate, the smoothness of the target function must be known so that the kernel function can be chosen to adapt to it. Indeed, this disadvantage of the kernel approach becomes more severe when there are multiple targets to be estimated simultaneously. In such situations, even if the smoothness of all targets are known, choosing a kernel function to adapt for each component is a nontrivial task. It would be even more challenging when the number of targets grows with the sample size, as is the case considered in this paper. Compared to the kernel-based method, the ridge approach can adapt remarkably well to targets with different smoothness via cross-validation, as shown in Hall and Meister (2007). On the other hand, the kernel approach requires fewer tuning parameters. In particular, the ridge approach involves two tuning parameters, r and ζ, while the kernel approach uses only one: the bandwidth. However, we claim that while the choice of the ridging parameter ζ is important to the performance of our estimator, the choice of r is far less so. A detailed discussion on the choice of r and ζ is left to Section 3.1.

To conduct statistical inference, we construct a second-stage estimator for which we can establish its asymptotic distribution. If X^* is observable, we can implement one-step backfitting as in Horowitz and Mammen (2004), where the second-stage estimator of g is given by a nonparametric regression of the residuals of the first stage, $Y_j - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})$, on the covariate X_j^*. When X^* is mismeasured and unobservable, we modify this second-stage by applying a deconvolution kernel regression. In particular, let

$$\mathbb{K}_h(w) = \frac{1}{2\pi} \int e^{-\mathrm{i}tw} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt,$$

be the deconvolution kernel, where K is a kernel function and h is a bandwidth. The second-stage estimator of g is defined as

$$\tilde{g}(x^*) = \frac{\sum_{j=1}^{n} \mathbb{K}_h(x^* - X_j)\left[Y_j - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})\right]}{\sum_{j=1}^{n} \mathbb{K}_h(x^* - X_j)}. \tag{2.7}$$

The second-stage estimator of m_d, however, cannot use a direct application of the deconvolution kernel regression because the unobservable X^* is now present in the dependent variable $Y_j - \hat{\mu} - \hat{g}(X_j^*) - \sum_{d' \neq d}^{D} \hat{m}_{d'}(Z_{d',j})$. Moreover, it enters in a nonlinear fashion, instead of simply acting as a covariate. One initial thought would be to first estimate $g(x^*) + m_d(z_d)$ by the deconvolution kernel regression of $Y_j - \hat{\mu} - \sum_{d' \neq d}^{D} \hat{m}_{d'}(Z_{d',j})$ on $(X_j^*, Z_{d,j})$, then deduct $\hat{g}(x^*)$. This, however, would make the estimator of m_d dependent on the choice of x^*, which would not be welcome in practice. Alternatively, we consider the standard kernel regression of $Y_j - \hat{\mu} - \sum_{d' \neq d}^{D} \hat{m}_{d'}(Z_{d',j})$ on $Z_{d,j}$, and then deduct an estimator of $E[g(X^*)|Z_d]$ to estimate m_d. The conditional moment $E[g(X^*)|Z_d]$ can be estimated based on estimates of g and the joint density of X^* and Z_d. For the joint density of X^* and Z_d, we use the deconvolution density estimator. For the unknown function g, it is natural to employ its first-stage estimator $\hat{g}$. However, since $\hat{g}(x^*)$ is a valid estimator of $g(x^*)$ only when $x^* \in \mathcal{I}$, the second-stage estimation of m_d

should be conditional on $X^* \in \mathcal{I}$. In particular, we consider

$$
\begin{aligned}
m_d(z_d) &= E\Big[Y - \mu - g(X^*) - \sum_{d' \neq d} m_{d'}(Z_{d'}) | Z_d = z_d, X^* \in \mathcal{I}\Big] \\
&= \frac{\int_{\mathcal{I}} E\big[Y - \mu - g(X^*) - \sum_{d' \neq d} m_{d'}(Z_{d'}) | Z^d = z_d, X^* = x^*\big] f_{Z_d, X^*}(z_d, x^*) dx^*}{\int_{\mathcal{I}} f_{Z_d, X^*}(z_d, x^*) dx^*},
\end{aligned}
$$

which suggests the following second-stage estimator of m_d:

$$
\tilde{m}_d(z_d) = \frac{\sum_{j=1}^n \int_{\mathcal{I}} \mathbb{K}_h(x^* - X_j)\big[Y_j - \hat{\mu} - \hat{g}(x^*) - \sum_{d' \neq d} \hat{m}_{d'}(Z_{d',j})\big] dx^* K_h(z_d - Z_{d,j})}{\sum_{j=1}^n \int_{\mathcal{I}} \mathbb{K}_h(x^* - X_j) dx^* K_h(z_d - Z_{d,j})}, \quad (2.8)
$$

with $K_h(w) = K(w/h)$ for a (conventional) kernel function K.

Remark 1. [Case of vector X] We note that the proposed method can be generalized to the case of vector X, i.e.,

$$
Y = \mu + g_1(X_1^*) + \cdots + g_L(X_L^*) + m_1(Z_1) + \cdots + m_D(Z_D) + U,
$$

where $X_1^*, \ldots, X_L^*$ are unobservable, and instead we observe noisy measurements $X_1, \ldots, X_L$. Suppose the measurement errors $\epsilon_1, \ldots, \epsilon_L$ are classical and mutually independent. In this case, the first-stage estimator can be constructed similarly. The second-stage estimator can then be obtained as

$$
\tilde{g}_l(x_l^*) = \frac{\sum_{j=1}^n \int_{\mathcal{I}_{l-}} \prod_{l'=1}^L \mathbb{K}_h(x_{l'}^* - X_{l',j}) \Big[Y_j - \hat{\mu} - \sum_{l' \neq l} \hat{g}_{l'}(x_{l'}^*) - \sum_{d=1}^D \hat{m}_d(Z_{d,j})\Big] dx_{l-}^*}{\sum_{j=1}^n \int_{\mathcal{I}_{l-}} \prod_{l'=1}^L \mathbb{K}_h(x_{l'}^* - X_{l',j}) dx_{l-}^*},
$$

$$
\tilde{m}_d(z_d) = \frac{\sum_{j=1}^n \int_{\mathcal{I}} \prod_{l'=1}^L \mathbb{K}_h(x_{l'}^* - X_{l',j})\big[Y_j - \hat{\mu} - \sum_{l=1}^L \tilde{g}_l(x_l^*) - \sum_{d' \neq d} \hat{m}_{d'}(Z_{d',j})\big] dx^* K_h(z_d - Z_{d,j})}{\sum_{j=1}^n \int_{\mathcal{I}} \prod_{l'=1}^L \mathbb{K}_h(x_{l'}^* - X_{l',j}) dx^* K_h(z_d - Z_{d,j})},
$$

for $l = 1, \ldots, L$ and $d = 1, \ldots, D$, where $\mathcal{I}_{l-} = \mathcal{I}_1 \times \cdots \times \mathcal{I}_{l-1} \times \mathcal{I}_{l+1} \times \cdots \times \mathcal{I}_L$, $\mathcal{I} = \mathcal{I}_1 \times \cdots \times \mathcal{I}_L$, $dx_{l-}^* = dx_1^* \ldots dx_{l-1}^* dx_{l+1}^* \ldots dx_L^*$, and $dx^* = dx_1^* \ldots dx_L^*$. We expect that analogous results to those in the next section can be established for this estimator as well.

Remark 2. [Case of correctly measured discrete covariates] Even though Assumption 1 (4) requires all covariates to be continuous, the proposed method can be generalized to the case where some correctly measured covariates are discrete (with finite support), i.e.,

$$
Y = \mu + W'\beta + g(X^*) + m_1(Z_1) + \cdots + m_D(Z_D) + U, \quad (2.9)
$$

where $W = (W_1, \ldots, W_S)'$ is a vector of indicators for specific values of correctly measured discrete covariates and $\beta = (\beta_1, \ldots, \beta_S)'$ are the corresponding slope parameters. In this case, the first-stage estimator can be constructed using $\tilde{P}_\kappa = (p_0(X^*)W_1, \ldots, p_0(X^*)W_S, P_\kappa')'$ instead of P_κ as in (2.4) and (2.5), and the second-stage estimator follows by using $\tilde{Y}_j = Y_j - \tilde{\mu} - W_j'\tilde{\beta}$ in place of Y_j as in (2.7) and (2.8), where $\tilde{\mu}$ and $\tilde{\beta}$ are the first-stage estimators of μ and β based on $\tilde{P}_\kappa$.

Remark 3. [Non-classical measurement error] The classical measurement error assumption is restrictive in many cases and has often come under criticism; see, e.g., Bound, Brown and Mathiowetz (2001) and Hyslop and Imbens (2001). Consequently, many papers have begun to move beyond this assumption to provide results for the non-classical error case; see, e.g., Hu and Schennach (2008), Gottschalk and Huynh (2010), Bonhomme and Robin (2010), Hu and Sasaki (2015), and An, Wang and Xiao (2020). See also Schennach (2020, Section 3.6) for a survey on the developments to deal with non-classical measurement errors. Furthermore, Schennach (2019) showed that nonparametric deconvolution methods can be applied under a weaker assumption than independence, known as subindependence. However, in each of these cases, this research built on earlier work on the same models with classical measurement error. For the additive regression model, there is still much work to be completed in the classical error case; thus, we see our work as a stepping stone to the more widely applicable case of non-classical error.

Although the estimator is constructed based on the classical measurement error assumption, there are cases with non-classical error that our method can be applied to. As an example, consider the case where W is a noisy measurement of W^* which entails a measurement error $W^*(\nu - 1)$, where ν is a random error that is independent of W^*. In this case, our estimator can be applied after implementing a log-transformation and by treating $\log(W)$ as X, $\log(W^*)$ as X^*, and $\log(\nu)$ as ϵ. As another example, consider the case where X is a noisy measurement of X^* which entails a measurement error $(\phi - 1)X^* + \nu$, for constant $\phi \neq 1$, where ν is a random error that is independent of X^*. In this case, if $\phi \neq 0$, we can treat ϕX^* as the true underlying covariate instead of X^*,[1] and our estimator can be directly applied to estimate $g, m_1, \ldots, m_D$.

3. Asymptotic properties

3.1. **First-stage estimator.** We now study the asymptotic properties of the first-stage estimator in (2.5). Let $\mathcal{F}_{\alpha,c} = \{f \in L_2(\mathbb{R}) : \int |f^{\text{ft}}(t)|^2(1+|t|^2)^\alpha dt \leq c\}$ denote the Sobolev class of order $\alpha > 0$ and $c > 0$.[2] We impose the following assumptions.

Assumption 2. **(1):** $\{Y_j, X_j, Z_j\}_{j=1}^n$ *is an i.i.d. sample of* (Y, X, Z) *satisfying (1.1), (1.2),* $E[Y^2|X^*, Z] < \infty$*, and* f_Z *is bounded.*

(2): f_{X^*}, $f_{X^*|Z_d=z_d}$, $f_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}$, $E[Y|X^*]f_{X^*}$, *and* $E[Y|X^* = \cdot, Z_d = z_d]f_{X^*|Z_d=z_d}$ *belong to* $\mathcal{F}_{\alpha,c_{\text{sob}}}$ *for all* $d, d' = 1, \ldots, D$ *and* $z_d, z_{d'} \in [-1, 1]$.

(3): $\{p_k\}_{k=1}^\infty$ *and* $\{q_k\}_{k=1}^\infty$ *are basis functions supported on* $\mathcal{I}$ *and* $[-1, 1]$*, respectively, and satisfy* $\int p_k(w)dw = \int q_k(w)dw = 0$ *and* $\langle p_k, p_{k'}\rangle = \langle q_k, q_{k'}\rangle = \delta_{k,k'}$ *for all* k, k'.

(4): $\lambda_{\min}(E[P_\kappa P_\kappa']) \geq \underline{\lambda} > 0$ *for all* κ, $\sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D} \|P_\kappa(x^*, z)\| = O(\kappa^{1/2})$ *as* $\kappa \to \infty$*, and there exists* $\theta_0 = (\mu_0, \theta_0^0, \theta_0^1, \ldots, \theta_0^D)$ *such that*

$$\sup_{x^*\in\mathcal{I}} |g(x^*) - P_{\kappa,0}'(x^*)\theta_0^0| = O(\kappa^{-2}), \qquad \sup_{z_d\in[-1,1]} |m_d(z_d) - P_{\kappa,d}'(z_d)\theta_0^d| = O(\kappa^{-2}),$$

[1]If ϕX^* is treated as the true underlying covariate and $\phi \neq 0$, we have $Y - \mu - m_1(Z_1) - \cdots - m_D(Z_D) - U = \tilde{g}(\phi X^*)$, with $\tilde{g}(\cdot) = g(\cdot/\phi)$. Then the normalization required in Assumption 1 (6) becomes $\int_{\tilde{\mathcal{I}}} \tilde{g}(w)dw = 0$, with $\tilde{\mathcal{I}} = \{\tilde{w} = \phi w : w \in \mathcal{I}\}$. Since $\mathcal{I}$ is the range of X^* chosen by researchers, instead of $\tilde{\mathcal{I}}$ (which is unknown without a specific value of ϕ), we may directly decide the range of interest of ϕX^* based on limited knowledge of ϕ (for example, a set of possible values of ϕ).

[2]Even though it seems somewhat different, the Sobolev condition imposed here is essentially equivalent to the one used in Meister (2009, eq. (2.30)), which imposes $\int |f^{\text{ft}}(t)|^2|t|^{2\alpha}dt < c$. First, it is easy to see that $\int |f^{\text{ft}}(t)|^2(1+|t|^2)^\alpha dt < c$ implies $\int |f^{\text{ft}}(t)|^2|t|^{2\alpha}dt < c$. For the other direction, we have $\int |f^{\text{ft}}(t)|^2(1+|t|^2)^\alpha dt \leq 2^\alpha \int_{|t|\leq 1} |f^{\text{ft}}(t)|^2 dt + 2^\alpha \int |f^{\text{ft}}(t)|^2|t|^{2\alpha}dt < c'$, where the first inequality follows by $2^\alpha|t|^{2\alpha} \geq (1+|t|^2)^\alpha \Leftrightarrow |t| \geq 1$, and the second inequality follows by $f \in L_2(\mathbb{R})$ and Meister (2009, eq. (2.30)).

where $P_{\kappa,0}(x^*) = (p_1(x^*), \ldots, p_\kappa(x^*))$ *and* $P_{\kappa,d}(z_d) = (q_1(z_d), \ldots, q_\kappa(z_d))$ *for* $d = 1, \ldots, D$.

(5): $r \geq 0$, $\zeta > 0$, *and* $\kappa \to \infty$ *as* $n \to \infty$.

Assumption 2 (1) is standard for cross-section data. Extensions to more general data environments are beyond the scope of this paper. Assumption 2 (2) lists the Sobolev conditions for several densities and regression functions, these restrict the smoothness of the underlying objects to control orders of the bias terms from estimation. Assumption 2 (3) contains conditions on the basis functions $\{p_k\}_{k=1}^\infty$ and $\{q_k\}_{k=1}^\infty$. Similar conditions are used in Horowitz and Mammen (2004) for the first-stage estimator without measurement error. Assumption 2 (4) is commonly used for series-based estimation; see, e.g., Newey (1997, Assumptions 2 and 3). Assumption 2 (5) contains mild requirements on the tuning constants, r and ζ, for the ridge regularization, and κ for the series approximation. See the remark at the end of this subsection for further discussion.

Although Assumptions 2 (2) and (4) both contain smoothness conditions, they focus on different objects for different purposes: Assumption 2 (2) imposes smoothness of densities and products of densities and regression functions to control the estimation bias, while Assumption 2 (4) is imposed on each of the nonparametric components of the regression function to control the series approximation error. This approximation bias term contributes to the last terms of the convergence rates in Theorem 2 below and indicates that the series length κ needs to diverge fast enough to achieve the desired convergence rate. As explained in Newey (1997, pp.150), for spline basis functions, since our nonparametric components are all univariate, Assumption 2 (4) essentially requires that g and m_d are twice continuously differentiable, which is consistent with Assumption 2 (2) when $\alpha = 2$.

It is known in the literature that the convergence rate of a deconvolution-based estimator depends on the smoothness of the measurement error density f_ϵ. Intuitively, the deconvolution-based estimators typically involve the characteristic function of ϵ in the denominator. The smoother f_ϵ is, the faster its characteristic function decays to zero in the tails, slowing down the convergence of the resulting estimator. Therefore, for the density of the measurement error f_ϵ, we consider the following two categories that are commonly employed in the deconvolution literature.

f_ϵ is said to be *ordinary smooth* of order β, if there exist constants $c_{\text{os},1} > c_{\text{os},0} > 0$ and $\beta > 0$ such that

$$c_{\text{os},0}(1 + |t|)^{-\beta} \leq |f_\epsilon^{\text{ft}}(t)| \leq c_{\text{os},1}(1 + |t|)^{-\beta} \quad \text{for all } t \in \mathbb{R}.$$

f_ϵ is said to be *supersmooth* of order γ, if there exist constants $c_{\text{ss},1} > c_{\text{ss},0} > 0$, $\mu > 0$, and $\gamma > 0$ such that

$$c_{\text{ss},0} \exp(-\mu|t|^\gamma) \leq |f_\epsilon^{\text{ft}}(t)| \leq c_{\text{ss},1} \exp(-\mu|t|^\gamma) \quad \text{for all } t \in \mathbb{R}.$$

In particular, the characteristic function of an ordinary smooth error distribution decays at a polynomial rate, while the characteristic function of a supersmooth error distribution decays at an exponential rate. Typical examples of ordinary smooth densities are the Laplace and gamma densities, and examples of supersmooth densities include the normal and Cauchy densities. To facilitate the discussion of the convergence rate of the first-stage estimator, we impose the following assumptions to specify the smoothness of the error distribution.

Assumption 3. *f_ϵ is ordinary smooth of order $\beta > 1/2$ and* $\max\{\kappa^2 n^{2\zeta+\frac{\zeta}{\beta}-1}, \kappa n^{-\frac{2\alpha\zeta}{\beta}}\} \to 0$ *as* $n \to \infty$.

Assumption 4. *f_ϵ is supersmooth of order $\gamma > 0$, $r = 0$, $0 < \zeta < \frac{1}{4}$, and* $\kappa(\log n)^{-\frac{2\alpha}{\gamma}} \to 0$ *as* $n \to \infty$.

Besides the supersmooth condition on the measurement error distribution, Assumption 4 contains further requirements on smoothing parameters r and ζ. Given $r = 0$, Assumption 4 guarantees that the variance of the first-stage estimation error converges to zero at a polynomial rate and is dominated by the bias of the first-stage estimation error, which converges to zero at a logarithmic rate in the supersmooth case. Under these assumptions, the convergence rate of the first-stage estimator in (2.5) is obtained as follows.

Theorem 2. *Suppose that Assumptions 1 and 2 hold true.*

(1): *Under Assumption 3, it holds*

$$\|\hat{\theta} - \theta_0\| = O_p\left(\kappa n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa^{\frac{1}{2}} n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-2}\right),$$

$$\sup_{x^* \in \mathcal{I}} |\hat{g}(x^*) - g(x^*)| = O_p\left(\kappa^{\frac{3}{2}} n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right),$$

$$\sup_{z_d \in [-1,1]} |\hat{m}_d(z_d) - m_d(z_d)| = O_p\left(\kappa^{\frac{3}{2}} n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right),$$

for $d = 1, \ldots, D$.

(2): *Under Assumption 4, it holds*

$$\|\hat{\theta} - \theta_0\| = O_p\left(\kappa^{\frac{1}{2}} (\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-2}\right),$$

$$\sup_{x^* \in \mathcal{I}} |\hat{g}(x^*) - g(x^*)| = O_p\left(\kappa (\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-\frac{3}{2}}\right),$$

$$\sup_{z_d \in [-1,1]} |\hat{m}_d(z_d) - m_d(z_d)| = O_p\left(\kappa (\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-\frac{3}{2}}\right),$$

for $d = 1, \ldots, D$.

It is worth noting that the number of regressors D does not appear in the convergence rate obtained in Theorem 2; this is due to the additive structure of the regression function combined with the series approximation. This immunity of the additive model to the curse of dimensionality is well-documented for the error-free case; we contribute to the literature by showing it continues to hold in the face of measurement error.

The first two terms in the convergence rates of Theorem 2 (1) and the first terms in Theorem 2 (2) are due to estimation variances, which indicate that to achieve the desired convergence rate, the series length κ cannot diverge too quickly.

The last terms in the convergence rates above characterize the magnitudes of the series approximation errors, which are identical to those of the error-free case; see Horowitz and Mammen

(2004, Theorem 1). For $\hat{g}$ and $\hat{m}_d$ in the ordinary smooth case, the first two terms $\kappa^{\frac{3}{2}}n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}}$ and $\kappa n^{-\frac{\alpha\zeta}{\beta}}$ in the convergence rates characterize the magnitudes of the estimation bias and variance, respectively. For the supersmooth case, the term $\kappa(\log n)^{-\frac{\alpha}{\gamma}}$ characterizes the magnitude of the estimation bias, while the variance of the estimation error is dominated under Assumption 4. If the smoothness parameters α, β, and γ are known, we can choose κ and ζ to achieve the optimal convergence rates. In particular, when f_ϵ is ordinary smooth, by setting $\kappa = n^{\frac{2\alpha}{7\alpha+10\beta+5}}$ and $\zeta = \frac{5\beta}{7\alpha+10\beta+5}$, the optimal convergence rate of $\hat{g}$ and $\hat{m}_d$ is obtained as $n^{-\frac{3\alpha}{7\alpha+10\beta+5}}$. When f_ϵ is supersmooth, by setting $\kappa = (\log n)^{\frac{2\alpha}{5\gamma}}$, the optimal convergence rate of $\hat{g}$ and $\hat{m}_d$ is obtained as $(\log n)^{-\frac{3\alpha}{5\gamma}}$.

Remark 4. [Comparison with Han and Park (2018)] In the ordinary smooth case, we can compare our rate results to those obtained in Han and Park (2018) for their smoothed backfitting estimator. In particular, when $\alpha = 2$ and $\beta > 1/2$,[3] Han and Park (2018) showed that their backfitting estimator of g achieves the uniform convergence rate $n^{-\frac{1}{4+4\beta}}$, which is slower than the convergence rate $n^{-\frac{6}{19+10\beta}}$ of our first-stage estimator $\hat{g}$.[4] The difference is due to backfitting. In particular, the estimation variance of the backfitting estimator of Han and Park (2018) is dominated by two components. One comes solely from the measurement error, which would still exist if $D = 1$. The other arises from backfitting in conjunction with measurement errors, which dominates if $\beta > 1/2$. As we consider the estimation of all nonparametric components simultaneously in the first stage, our first-stage estimator $\hat{g}$ converges faster than the backfitting estimator of Han and Park (2018). However, Han and Park (2018) can handle the case when $\beta \leq 1/2$, which cannot be covered in this paper since $\beta > 1/2$ is required by the ridge-based regularization as in Lemma 3 (1).[5]

Remark 5. [Comparison with the error-free case] In the error-free case, by Horowitz and Mammen (2004, Theorem 1), the optimal convergence rate to estimate g and m_d is $n^{-\frac{3}{10}}$, which is obtained by setting $\kappa = n^{\frac{1}{5}}$. When f_ϵ is ordinary smooth of order $\beta > 1/2$ and $\alpha = 2$,[6] the optimal convergence rate of our $\hat{g}$ and $\hat{m}_d$ is slower than $n^{-\frac{1}{4}}$; thus, it is slower than $n^{-\frac{3}{10}}$. In the case of supersmooth f_ϵ, $\hat{g}$ and $\hat{m}_d$ converge at a logarithmic rate, which is certainly slower than the polynomial rate obtained in Horowitz and Mammen (2004). However, these slower convergence rates are quite reasonable given the contaminated nature of the sample.

Remark 6. [Choice of tuning parameters] To implement the first-stage estimator, we need to choose three tuning parameters, κ, r, and ζ. For the series length κ, to the best of our knowledge,

[3]We set $\alpha = 2$ because Han and Park (2018) assumed that f_j is twice continuously differentiable in their Assumption K4. See Meister (2009, Section A.2) for the relationship between order of differentiability and choice of α.

[4]Even though Han and Park (2018) considered a different setup where all covariates are mismeasured, the convergence rate of their smoothed backfitting method would remain the same when only one covariate is mismeasured, as it is independent of their number of covariates d. This is a natural result when the regression function has an additive structure. Therefore, the uniform convergence rate presented in Han and Park (2018, Corollary 3.5) can be directly compared to Theorem 2 when the smoothing parameters α and β are the same.

[5]We note that if $\beta > 1/2$, then f_ϵ is bounded and continuous.

[6]Similarly to the previous case, here we set $\alpha = 2$ because Horowitz and Mammen (2004) assume that m_j is twice continuously differentiable in their Assumption A2.

there is no theoretical study on the optimal choice even for the error-free additive model. As suggested in Horowitz and Lee (2005), one practical way is to construct a BIC-type criterion function for κ, and choose κ to minimize it. In our setup, the BIC-type criterion is obtained from the sample counterpart of the least squares objective function (2.3) with a penalty term for κ. For the tuning parameters r and ζ in the ridge-type regularization, we can follow the suggestions in Hall and Meister (2007). The choice of r, which controls the shape of the smoothing regime, is less important. For example, Hall and Meister (2007) set $r = 2$ for the ordinary smooth case and $r = 0$ for the supersmooth case in their numerical study. On the other hand, ζ plays the role of the ridge smoothing parameter, and its choice is crucial. For example, the moment estimator in (2.6) is interpreted as the one for $E[Yp_k(X^*)] = \int m(x)f_{X^*}(x)p_k(x)dx$. Thus, we can adapt the cross-validation method in Hall and Meister (2007, pp. 1539-40), which minimizes an estimate of $\int |\widehat{m(x)f_{X^*}}(x) - m(x)f_{X^*}(x)|^2 dx$ with respect to ζ, to the criterion weighted by $p_k(x)^2$.

3.2. **Second-stage estimator.** In this subsection, we derive the asymptotic distribution of the second-stage estimators $\tilde{g}$ and $\tilde{m}_d$. To this end, we impose the following additional assumptions.

Assumption 5. **(1):** *f_{X^*} is continuously differentiable, $\|f_X\|_\infty < \infty$, and g is twice continuously differentiable.*

(2): $\sup_x E[|U|^{2+\eta}|X = x] < \infty$ *for some constant $\eta > 0$.*

(3): *$\int wK(w)dw = 0$, $\int w^2K(w)dw < \infty$, $\|K^{\mathrm{ft}}\|_\infty < \infty$, and $\|K^{\mathrm{ft}\prime}\|_\infty < \infty$.*

(4): *$h \to 0$ as $n \to \infty$.*

Assumption 5 (1) contains smoothness conditions on the density f_{X^*} and the regression function g, which are used to control the estimation bias. Assumption 5 (2) is used to apply Lyapunov's central limit theorem. Assumption 5 (3) concerns the kernel function K, which is commonly employed to control the bias from nonparametric estimation. Assumption 5 (4) is standard for kernel-based estimators (as used in the second-stage estimator).

For the ordinary smooth case, we impose the following assumptions.

Assumption 6. **(1):** *$\|f_\epsilon^{\mathrm{ft}\prime}\|_\infty < \infty$, $|s|^\beta |f_\epsilon^{\mathrm{ft}}(s)| \to c_\epsilon$, and $|s|^{\beta+1}|f_\epsilon^{\mathrm{ft}\prime}(s)| \to \beta c_\epsilon$ for some constant $c_\epsilon > 0$ as $|s| \to \infty$.*

(2): *$\int |s|^\beta \{|K^{\mathrm{ft}}(s)| + |K^{\mathrm{ft}\prime}(s)|\}ds < \infty$, $\int |s|^{2\beta}|K^{\mathrm{ft}}(s)|^2 ds < \infty$.*

(3): *$\kappa^3 n^{2\zeta + \frac{\zeta}{\beta} - 1} \to 0$ and $\kappa n^{-\frac{\alpha\zeta}{\beta}} \to 0$ as $n \to \infty$.*

Assumption 6 (1) is commonly used in deconvolution problems with an ordinary smooth error. It goes further than Assumption 3, as Assumption 6 (1) characterizes the exact limit, rather than the upper and lower bounds, of the error characteristic function and its derivative in the tails. Assumption 6 (2) requires smoothness of the kernel function K. Assumption 6 (3) is required to eliminate estimation error from the first stage. According to Theorem 2, it guarantees that the first-stage estimator is uniformly consistent when the measurement error is ordinary smooth of order β. To derive the asymptotic distribution of $\tilde{g}$, we add the following assumptions.

Assumption 7. **(1):** *For each $x^* \in \mathcal{I}$, $E[|g(X^*) + U - g(x^*)|^2|X = x]$ as a function of x is continuous for almost all x.*

(2): $nh^{2\beta+1} \to \infty$ *as* $n \to \infty$.

Assumption 7 (1) is a technical assumption. Given Assumption 5, it would be satisfied if all densities are continuous. Assumption 7 (2) imposes an upper bound on the speed at which the bandwidth h decays to zero; this controls the estimation variance brought by the measurement error, and thus is characterized by the smoothness order of the measurement error distribution.

For the supersmooth case, we impose the following assumptions.

Assumption 8. **(1):** K^{ft} *is supported on* $[-1, 1]$.
(2): $\kappa(\log n)^{-\frac{\alpha}{\gamma}} \to 0$ *as* $n \to \infty$.

Assumption 8 (1) assumes the kernel function K is of infinite-order, rather than adapting the smoothness of the kernel function to that of the measurement error density as in the ordinary smooth case. Assumption 8 (2), parallel to Assumption 6 (3), asymptotically eliminates estimation error from the first stage. According to Theorem 2, it guarantees that the first-stage estimator is uniformly consistent when the measurement error density is supersmooth. To derive the asymptotic distribution of $\tilde{g}$, we add the following assumptions.

Assumption 9. **(1):** $nhe^{-2\mu h^{-\gamma}} \to \infty$ *as* $n \to \infty$.
(2): $E|G_{1,n,1}|^2 he^{-2\mu h^{-\gamma}} \to \infty$ *as* $n \to \infty$*, where* $G_{1,n,1}$ *is defined in Appendix B.2.*

Assumption 9 (1) requires the bandwidth h to go to zero at a logarithmic rate at most, this is due to the error characteristic function in the denominator decaying at an exponential rate. Assumption 9 (2) is a technical assumption used to verify Lyapunov's condition in the proof of Theorem 3. Primitive conditions, as in Fan and Masry (1992, Condition 3.1), could be derived. To keep the exposition simple, following Delaigle, Fan and Carroll (2009), we stick to the current form of Lyapunov's condition.

Under these assumptions, the asymptotic distribution of the second-stage estimator $\tilde{g}$ is obtained as follows. Let $\text{Bias}\{\tilde{g}(x^*)\} = g(x^*) - E[\tilde{g}(x^*)]$ and $Var[\tilde{g}(x^*)]$ be the variance of $\tilde{g}(x^*)$.

Theorem 3. *Suppose Assumptions 1, 2, and 5 hold true.*

(1): *Under Assumptions 3, 6, and 7, it holds*

$$\frac{\tilde{g}(x^*) - g(x^*) - Bias\{\tilde{g}(x^*)\}}{\sqrt{Var[\tilde{g}(x^*)]}} \xrightarrow{d} N(0,1).$$

(2): *Under Assumptions 4, 8, and 9, it holds*

$$\frac{\tilde{g}(x^*) - g(x^*) - Bias\{\tilde{g}(x^*)\}}{\sqrt{Var[\tilde{g}(x^*)]}} \xrightarrow{d} N(0,1).$$

The asymptotic normality of $\tilde{g}$ is provided in a normalized form. It is interesting to note that the measurement error barely has any effect on the bias term $\text{Bias}\{\tilde{g}(x^*)\}$. Indeed, it can be shown that the dominant term of $\text{Bias}\{\tilde{g}(x^*)\}$ is the same as that of Horowitz and Mammen's (2004) second-stage estimator of g, which is of order h^2. On the other hand, the measurement error affects the manner of divergence of $Var[\tilde{g}(x^*)]$ to infinity. In particular, when f_ϵ is ordinary smooth, as shown in Appendix B.2, $Var[\tilde{g}(x^*)]$ explodes at the rate $h^{-(2\beta+1)}$. In the case of supersmooth f_ϵ, deriving the exact exploding rate of $Var[\tilde{g}(x^*)]$ is difficult in general. Thus, the lower bound on the exploding rate of $Var[\tilde{g}(x^*)]$ is obtained under Assumption 9 rather than the exact rate, as shown in Appendix B.2.

Similarly to the second-stage estimator of Horowitz and Mammen (2014) in the error-free case, our second-stage estimator is oracle in the sense that its asymptotic distribution is the same as if all other nonparametric components were known. In particular, as shown in Appendix C, the asymptotic distribution of $\tilde{g}$ is characterized by the deconvolution kernel regression estimator with dependent variable $Y - \mu - \sum_{d=1}^{D} m_d(Z_d)$ and bandwidth h, and the choice of the bandwidth h is independent of the choice of the first-stage tuning parameters κ and ζ. In fact, both Assumption 6 (3) and Assumption 8 (2), which are separately imposed to guarantee the asymptotic negligibility of the first-stage estimation error for the ordinary smooth case and the supersmooth case, respectively, only involve κ and ζ but not h.

Since X^* is not directly observable, it is difficult to adapt the penalized least squares method in Horowitz and Mammen (2004) to select the bandwidth parameter h in the second-stage estimator. Even for the conventional nonparametric deconvolution regression, it is not clear how to implement a standard data-driven selection of h, such as cross-validation (see, pp. 123-5 of Meister, 2009). One practical way to select h is to apply the SIMEX-based cross-validation method in Delaigle and Hall (2008) by setting the dependent variable as $Y_j - \hat{\mu} - \sum_{d=1}^{D} \hat{m}_d(Z_{d,j})$ in the second-stage estimation. However, the theoretical analysis of this is beyond the scope of this paper.

We now consider the asymptotic distribution of $\tilde{m}_d$. For the ordinary smooth case, we impose the following assumptions.

Assumption 10. **(1):** $\mathcal{I} = \operatorname{supp} g = [b_1, b_2]$.

(2): *f_ϵ is ordinary smooth of order $\beta \geq 2$.*

(3): *$E[|g(X^*) + m_d(Z_d) + U - m_d(z_d)|^2 | X = x, Z_d = z]$ is continuous for $d = 1, \ldots, D$ and almost all $(x, z) \in \mathcal{I} \times [-1, 1]$.*

(4): $\sup_s \left| g^{\text{ft}}\left(-\frac{s}{h}\right) \frac{s}{h^2} \right| \to 0$ *as $n \to \infty$.*

(5): *$nh^{2\beta} \to \infty$ as $n \to \infty$.*

Assumption 10 (1) assumes that $\mathcal{I}$ equals $\operatorname{supp} X^*$ and is a closed interval with known boundary points $b_1 < b_2$. It is stronger than Assumption 1 (4), where we assume that $\mathcal{I}$ is a compact subset of $\operatorname{supp} X^*$. However, this assumption is difficult to avoid in the current derivation of the asymptotic normality of $\tilde{m}_d$ because there is an additional integration of x^* over $\mathcal{I}$ in the definition of $\tilde{m}_d$, and we must be specific regarding the smoothness of this integration. In Assumption 10 (2), we require $\beta \geq 2$, which is a technical assumption to guarantee $\int |K^{\text{ft}}(s)||s|^{\beta-2} ds < \infty$. Assumption 10 (3) plays a similar role as Assumption 7 (1). Again, given Assumption 5, it would be satisfied if all densities are continuous. Assumption 10 (4) is an additional smoothness condition on g to ensure the estimation noise of $\hat{g}$ is negligible in the estimation of m_d. In particular, it requires that g^{ft} should decay to zero fast enough. Assumption 10 (5) imposes an upper bound on the decay rate of h to zero. This is different from Assumption 7 (2) due to the additional integration with respect to x^* in the definition of $\tilde{m}_d$.

To derive the asymptotic distribution of $\tilde{m}_d$ for the supersmooth case, we impose the following assumptions.

Assumption 11. **(1):** $\mathcal{I} = \operatorname{supp} g = [b_1, b_2]$.

(2): *$nh^3 e^{-2\mu h^{-\gamma}} \to \infty$ as $n \to \infty$.*

(3): *$E|G^d_{1,n,1}|^2 h^3 e^{-2\mu h^{-\gamma}} \to \infty$ as $n \to \infty$, where $G^d_{1,n,1}$ is defined in Appendix B.3 for $d = 1, \dots, D$.*

Assumption 11 (2) plays a similar role as Assumption 9 (1). This assumption requires the bandwidth h to decay at an even slower rate due to the extra integration in the definition of $\tilde{m}_d$. Assumption 11 (3) is a technical assumption used to verify Lyapunov's condition in the proof of Theorem 4, which is imposed to keep the presentation simple. Similar to Assumption 9 (2), primitive conditions, like Fan and Masry (1992, Condition 3.1), could be derived.

The asymptotic distribution of the second-stage estimator $\tilde{m}_d$ for m_d is obtained as follows. Let $\text{Bias}\{\tilde{m}_d(z_d)\} = m_d(z_d) - E[\tilde{m}_d(z_d)]$ and $Var[\tilde{m}_d(z_d)]$ be the variance of $\tilde{m}_d(z_d)$.

Theorem 4. *Suppose Assumptions 1, 2, and 5 hold true.*

(1): *Under Assumption 6 and 10, it holds*

$$\frac{\tilde{m}_d(z_d) - m_d(z_d) - Bias\{\tilde{m}_d(z_d)\}}{\sqrt{Var[\tilde{m}_d(z_d)]}} \xrightarrow{d} N(0,1).$$

(2): *Under Assumption 4, 8, and 11, it holds*

$$\frac{\tilde{m}_d(z_d) - m_d(z_d) - Bias\{\tilde{m}_d(z_d)\}}{\sqrt{Var[\tilde{m}_d(z_d)]}} \xrightarrow{d} N(0,1).$$

Similar to $\tilde{g}$, the asymptotic normality of $\tilde{m}_d$ is also provided in a normalized form. Again, it can be shown that the dominant term of $\text{Bias}\{\tilde{m}_d(z_d)\}$ is the same as that of the error-free second-stage estimator of m_d as in Horowitz and Mammen (2004), which has the order h^2, while the measurement error slows down the divergence rate of $Var[\tilde{m}_d(z_d)]$ to infinity. In particular, when f_ϵ is ordinary smooth, as shown in Appendix B.3, $Var[\tilde{m}_d(z_d)]$ diverges at the rate $h^{-2\beta}$, which is slower than that of $\tilde{g}$ due to the additional integration with respect to x^*. In the case of supersmooth f_ϵ, again, the lower bound on the divergence rate of $Var[\tilde{m}_d(z_d)]$ is obtained under Assumption 11 rather than the exact rate, as shown in Appendix B.3. By similar arguments for $\tilde{g}$, as for the second-stage estimator of Horowitz and Mammen (2004) in the error-free case, $\tilde{m}_d$ is oracle in the sense that its asymptotic distribution is the same as if all other nonparametric components were known.

4. Case of unknown measurement error distribution

4.1. **Setup and estimator.** A major limitation of our estimator is the assumption that the measurement error density f_ϵ is known, which is unrealistic in many settings. In this section, we relax this assumption by considering f_ϵ to be unknown but repeated measurements on X^* are available. Suppose we have two independent noisy measurements of the error-free covariate X^*, i.e.,

$$X_j = X_j^* + \epsilon_j, \quad \text{and} \quad X_j^r = X_j^* + \epsilon_j^r, \tag{4.1}$$

for $j = 1, \ldots, n$. The following assumption is imposed for identification of f_ϵ.

Assumption 12. *ϵ^r has the same distribution as ϵ, and is independent of (Y, X^*, ϵ). Furthermore, f_ϵ^{ft} is real-valued.*

These assumptions are commonly used in the literature when the measurement error distribution is symmetric around zero. In particular, Assumption 12 requires that ϵ^r is an independent copy of ϵ, and f_ϵ^{ft} is real-valued if f_ϵ is symmetric around zero. Under Assumption 12, the measurement error distribution can be identified by $f_\epsilon^{\mathrm{ft}}(t) = |E[\cos\{t(X - X^r)\}]|^{1/2}$. Given an i.i.d. sample $\{X_j, X_j^r\}_{j=1}^n$ of (X, X^r), following Delaigle, Hall and Meister (2008), f_ϵ^{ft} can be estimated by

$$\check{f}_\epsilon^{\mathrm{ft}}(t) = \left| \frac{1}{n} \sum_{j=1}^n \cos\{t(X_j - X_j^r)\} \right|^{1/2}.$$

Based on this estimator of f_ϵ^{ft}, we propose to estimate g and m_d by

$$\check{g}(x^*) = \sum_{k=1}^{\kappa} p_k(x^*)\check{\theta}_k^0, \qquad \check{m}_d(z_d) = \sum_{k=1}^{\kappa} q_k(z_d)\check{\theta}_k^d, \tag{4.2}$$

where $\check{\theta} = (\Re\check{E}[P_\kappa P_\kappa'])^{-1}\Re\check{E}[YP_\kappa']$, and elements of $\check{E}[YP_\kappa']$ and $\check{E}[P_\kappa P_\kappa']$ are constructed as

$$\check{E}[p(X^*)Q] = \frac{1}{2\pi} \int \left(\frac{1}{n} \sum_{j=1}^n Q_j e^{\mathrm{i}tX_j} \right) \frac{p^{\mathrm{ft}}(-t)\check{f}_\epsilon^{\mathrm{ft}}(t)}{\{\check{f}_\epsilon^{\mathrm{ft}}(t) \vee n^{-\zeta}\}^2} dt,$$

which is obtained by replacing f_ϵ^{ft} in $\hat{E}[p(X^*)Q]$ by the estimator $\check{f}_\epsilon^{\mathrm{ft}}$ and using the fact that $\check{f}_\epsilon^{\mathrm{ft}}(-t) = \check{f}_\epsilon^{\mathrm{ft}}(t)$ and $\check{f}_\epsilon^{\mathrm{ft}}(t) \geq 0$ for all $t \in \mathbb{R}$.

4.2. **Asymptotic properties.** To study the asymptotic properties of $\check{g}, \check{m}_1, \ldots, \check{m}_D$, we focus on a special case where $\mathcal{I}$ coincides with the support of X^*. Even though it is rare in practice, this simplifying assumption allows us to characterize the additional challenges that must be met in order to extend our methods to the case when f_ϵ is unknown. Alternatively, we could assume that the density of X^* is bounded away from zero over the range of integration; such an assumption is not uncommon in works of this type. The known support assumption, together with some additional assumptions used to derive the convergence rates of $\check{g}, \check{m}_1, \ldots, \check{m}_D$, are summarized as follows.

Assumption 13. **(1):** *$\mathcal{I}$ is the support of X^*.*
(2): *$E|\epsilon|^{2+\varsigma} < \infty$ for some $\varsigma > 0$.*
(3): *$\max_{d=1,\ldots,D} E|q_k(Z_d)|^2 < \infty$ and $\int |p_k^{\mathrm{ft}}(t)|dt < \infty$ for $k = 1, \ldots$.*

Assumption 13 (2) is a mild condition required by Lemma 11 in Appendix D, which is used to characterize the uniform convergence rate of the empirical characteristic function of $\epsilon - \epsilon^r$ over an expanding region. Assumption 13 (3) contains additional conditions on the basis functions than Assumption 2. In particular, the first part is imposed to guarantee that terms like $\frac{1}{n}\sum_{j=1}^n |q_k(Z_{d,j})|$ and $\frac{1}{n}\sum_{j=1}^n |q_k(Z_{d,j})q_l(Z_{d',j})|$ are stochastically bounded, and the second part is a regularity condition on p_k used to characterize the convergence rates of $\check{g}, \check{m}_1, \ldots, \check{m}_D$.

For the tuning parameter κ, when f_ϵ is unknown, in addition to Assumption 3 and 4, we make the following assumptions for the ordinary smooth case and supersmooth case, respectively.

Assumption 14. $\kappa^2 n^{3\zeta} \max \left\{ \begin{array}{c} n^{-1/2} \log \left(c_{\mathrm{os},1}^{1/\beta} n^{\zeta/\beta} - 1 \right), \\ \max_{1 \le k \le \kappa} \int_{|t| > c_{\mathrm{os},0}^{1/\beta} n^{\zeta/\beta} - 1} |p_k^{\mathrm{ft}}(t)| dt \end{array} \right\} \to 0$ *as* $n \to \infty$.

Assumption 15. $\kappa^2 n^{3\zeta} \max \left\{ \begin{array}{c} n^{-1/2} \log \left(\mu^{-1/\gamma} \log(c_{\mathrm{ss},1} n^{\zeta})^{1/\gamma} \right), \\ \max_{1 \le k \le \kappa} \int_{|t| > \mu^{-1/\gamma} \log(c_{\mathrm{ss},0} n^{\zeta})^{1/\gamma}} |p_k^{\mathrm{ft}}(t)| dt \end{array} \right\} \to 0$ *as* $n \to \infty$.

In the ordinary smooth case, Assumption 14 guarantees that if n is large enough, the moment matrix $\check{E}[P_\kappa P_\kappa']$ for the case when f_ϵ is estimated using repeated measurements on X^* is close to the moment matrix $\hat{E}[P_\kappa P_\kappa']$ for the case when f_ϵ is directly known, which is used in the proof of Theorem 5. Assumption 15 plays a similar role in the proof of Theorem 5 when f_ϵ is supersmooth.

Under these assumptions, the convergence rates of the estimators for the case of unknown f_ϵ are obtained as follows.

Theorem 5. *Suppose Assumption 1 holds except f_ϵ is unknown, and Assumptions 2, 12, and 13 hold true.*

(1): *Under Assumptions 3 and 14, it holds*

$$\|\check{\theta} - \theta_0\| = O_p \left(\kappa n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa^{\frac{1}{2}} n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-2} + \chi_{\mathrm{os},n,\zeta,\kappa}^{1/2} \right),$$

$$\sup_{x^* \in \mathcal{I}} |\check{g}(x^*) - g(x^*)| = O_p \left(\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}} + \kappa^{1/2} \chi_{\mathrm{os},n,\zeta,\kappa}^{1/2} \right),$$

$$\max_{1 \le d \le D} \sup_{z_d \in [-1,1]} |\check{m}_d(z_d) - m_d(z_d)| = O_p \left(\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}} + \kappa^{1/2} \chi_{\mathrm{os},n,\zeta,\kappa}^{1/2} \right),$$

where $\chi_{\mathrm{os},n,\zeta,\kappa} = \kappa^2 n^{-1/2 + 3\zeta} \log(c_{\mathrm{os},1}^{1/\beta} n^{\zeta/\beta} - 1) + \kappa^2 n^{3\zeta} \max_{1 \le k \le \kappa} \int_{|t| > c_{\mathrm{os},0}^{1/\beta} n^{\zeta/\beta} - 1} |p_k^{\mathrm{ft}}(t)| dt.$

(2): *Under Assumptions 4 and 15, it holds*

$$\|\check{\theta} - \theta_0\| = O_p \left(\kappa^{\frac{1}{2}} (\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-2} + \chi_{\mathrm{ss},n,\zeta,\kappa}^{1/2} \right),$$

$$\sup_{x^* \in \mathcal{I}} |\check{g}(x^*) - g(x^*)| = O_p \left(\kappa (\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-\frac{3}{2}} + \kappa^{1/2} \chi_{\mathrm{ss},n,\zeta,\kappa}^{1/2} \right),$$

$$\max_{1 \le d \le D} \sup_{z_d \in [-1,1]} |\check{m}_d(z_d) - m_d(z_d)| = O_p \left(\kappa (\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-\frac{3}{2}} + \kappa^{1/2} \chi_{\mathrm{ss},n,\zeta,\kappa}^{1/2} \right),$$

where $\chi_{\mathrm{ss},n,\zeta,\kappa} = \kappa^2 n^{3\zeta} \max_{1 \le k \le \kappa} \int_{|t| > \mu^{-1/\gamma} \log(c_{\mathrm{ss},0} n^{\zeta})^{1/\gamma}} |p_k^{\mathrm{ft}}(t)| dt.$

Theorem 5 (1) presents the convergence rates of $\check{g}, \check{m}_1, \ldots, \check{m}_D$ when f_ϵ is ordinary smooth. Compared to Theorem 2 (1) where f_ϵ is known, the additional components that contain $\chi_{\mathrm{os},n,\zeta,\kappa}$ characterize the estimation error of $\check{f}_\epsilon^{\mathrm{ft}}$. In particular, the first component of $\chi_{\mathrm{os},n,\zeta,\kappa}$ captures the estimation error of $\check{f}_\epsilon^{\mathrm{ft}}$ over an expanding region, and the second component is the upper bound of the estimation error of $\check{f}_\epsilon^{\mathrm{ft}}$ in the tails, obtained through the regularity of the basis functions p_k and the smoothness of f_ϵ. Theorem 5 (2) presents the convergence rates of $\check{g}, \check{m}_1, \ldots, \check{m}_D$ when f_ϵ is supersmooth. Similarly to the ordinary smooth case, the additional components that contain $\chi_{\mathrm{ss},n,\zeta,\kappa}$ characterize the estimation error of $\check{f}_\epsilon^{\mathrm{ft}}$. In contrast to the ordinary smooth case, since κ can only diverge at a logarithm rate, the additional component that captures the estimation error of $\check{f}_\epsilon^{\text{ft}}$ over an expanding region is dominated and does not appear in the final rate.

5. Finite sample properties

In this section, the finite sample properties of our estimator are investigated and compared to the estimators of Han and Park (2018) and Horowitz and Mammen (2004). Note that the estimator of Horowitz and Mammen (2004) is not designed to deal with measurement error, so this constitutes simply ignoring the measurement error issue.

The following data generating process is considered

$$Y = 1 + g(X^*) + m_1(Z_1) + m_2(Z_2) + U,$$

where $(X^* Z_1, Z_2)$ are each drawn from $N(0, 1/3)$ with correlation of 0.25 between each variable, and U is drawn from $N(0, 1/3)$ and is independent of $(X^* Z_1, Z_2)$. While X^* is assumed unobservable, we suppose $X = X^* + \epsilon$ is observed, where ϵ is mutually independent and independent of (X^*, Z_1, Z_2, U). We consider two cases for the density of ϵ. For the ordinary smooth case, ϵ has a zero mean Laplace distribution with variance of 1/9. For the supersmooth case, ϵ has a normal distribution with zero mean and variance of 1/9.

For the regression functions, we take $m_1(z) = z - z^2$, $m_2(z) = \sin(\pi z)$, and consider three specifications for $g(z)$:

$$\begin{aligned}
\text{DGP1 :} & g(z) = z - z^2, \\
\text{DGP2 :} & g(z) = \arctan(\pi z), \\
\text{DGP3 :} & g(z) = \cos(\pi z).
\end{aligned}$$

Note that each function is further standardized such that Assumption 1 (6) is satisfied, where we truncate the range of integration at the 5% and 95% quantiles of X^*.

Throughout the simulation study, we use the kernel proposed in Fan (1992) which has a Fourier transform given by

$$K^{\text{ft}} = \mathbb{I}\{|t| \leq 1\}(1 - t^2)^3.$$

This kernel satisfies all necessary assumptions given in Section 3. As basis functions for our approach, we use polynomials standardized to satisfy Assumptions 2 (4) and (5). For the choice of tuning parameters for the first-stage estimator, we follow the suggestions in Remark 6 of Section 3.1, while the bandwidth for the second-stage estimator is selected using the SIMEX approach of Delaigle and Hall (2008). The bandwidth for the estimator of Han and Park (2018) was also chosen using this SIMEX approach. For the method of Horowitz and Mammen (2004), we use cross-validation to choose the bandwidths and use the same polynomials as for our estimator.

Results for two sample sizes, 500 and 1000, are provided. The mean integrated squared error (multiplied by 10) for each estimator under each setting is given in Tables 1 - 3 and are based on

1000 Monte Carlo replications. 'DOT' refers to the estimator of this paper, while 'HP' and 'HM' refer to the methods of Han and Park (2018) and Horowitz and Mammen (2004), respectively.

Table 1: DGP 1

Estimator	g				m_1				m_2			
Error Type	**OS**		**SS**		**OS**		**SS**		**OS**		**SS**	
Sample Size	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**
DOT	0.30	0.26	0.42	0.38	0.26	0.23	0.24	0.21	0.18	0.11	0.16	0.10
HP	0.22	0.15	0.39	0.32	0.21	0.18	0.18	0.15	0.35	0.27	0.34	0.27
HM	0.51	0.46	0.58	0.55	0.24	0.21	0.24	0.21	0.16	0.10	0.16	0.10

Table 2: DGP 2

Estimator	g				m_1				m_2			
Error Type	**OS**		**SS**		**OS**		**SS**		**OS**		**SS**	
Sample Size	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**
DOT	0.21	0.16	0.49	0.40	0.26	0.23	0.25	0.22	0.19	0.12	0.17	0.11
HP	0.30	0.22	0.57	0.45	0.12	0.07	0.11	0.07	0.33	0.25	0.31	0.24
HM	0.73	0.64	0.90	0.81	0.24	0.22	0.24	0.22	0.16	0.10	0.16	0.10

Table 3: DGP 3

Estimator	g				m_1				m_2			
Error Type	**OS**		**SS**		**OS**		**SS**		**OS**		**SS**	
Sample Size	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**	**500**	**1000**
DOT	0.47	0.34	1.13	0.98	0.26	0.23	0.24	0.22	0.19	0.11	0.18	0.11
HP	0.53	0.35	1.49	1.30	0.21	0.16	0.19	0.16	0.35	0.27	0.34	0.27
HM	1.40	1.27	1.79	1.67	0.24	0.22	0.24	0.22	0.17	0.11	0.17	0.11

First, as would be expected, the MISE for each estimator falls as the sample size increases and when the function to be estimated is closer to linearity. Furthermore, as suggested by the theoretical results, the performance of the estimators is better in the case of ordinary smooth measurement error than with supersmooth error. It is also unsurprising to see that m_1 and m_2 are estimated with a lower MISE than g since the regressor associated with g is the only one to suffer from measurement error.

In all settings, the method of Horowitz and Mammen (2004) is clearly dominated by the other two methods when estimating g; this is to be expected since the approach of Horowitz and Mammen (2004) is designed to be used only with perfectly measured regressors. However, it is interesting to note that this estimator performs admirably when estimating m_1 and m_2, showing a marginal improvement over the method of this paper and generally giving lower MISE than the estimator of Han and Park (2018).

When comparing the estimator of this paper to that of Han and Park (2018), it appears that neither approach dominates the other. When the function to estimate is closer to linearity, i.e. g in DGP 1 and m_1 in all three DGPs, the estimator of Han and Park (2018) is preferable. However, when the function exhibits more nonlinearity, our estimator dominates. Interestingly, the difference between the estimators does not appear to depend on the smoothness of the measurement error density. This suggests that although Han and Park (2018) do not discuss the asymptotic properties of their estimator under supersmooth error, it is likely to remain consistent in this case.

6. Empirical application

In this section, we use our estimator to analyze the black-white wage gap. In particular, we aim to shed light on the differing wage-returns to cognitive ability, tenure, and education across race. This topic dates back to at least Blinder (1973) and has received much attention in the economics literature (see, for example, Card and Lemieux, 1996, Chay and Lee, 2000, and Lang and Manove, 2011). While many papers include cognitive ability as a control - or explicitly estimate the return to cognitive ability - few account for the inherent measurement error present in this variable. Schennach (2007) is one such exception. In that paper, she presents a nonlinear - but parametric - model of the black-white wage gap while taking seriously the issue of measurement error. See Lang and Lehman (2012) for a comprehensive review of this vast literature.

For our study, we use data from the National Longitudinal Survey of Youth 1979 (NLSY79). The dataset contains a sample of Americans who were aged between 14-22 when first interviewed in 1979. As is typical in work using the NLSY79 dataset, we restrict the sample to males who work in the formal labour market (Neal, 2004). One part of this extensive survey is the Armed Services Vocational Aptitude Battery (ASVAB), a series of multiple-choice tests designed to measure cognitive ability. The individuals are then periodically interviewed, with the most recent interview conducted in 2016. Among other variables, information on education, tenure, and wages is provided in these surveys.

Our interest lies in estimating the effect of cognitive ability, tenure, and education on wages as measured in 2016, and how these effects differ across race. Thus, we estimate an additive regression model using the approach of this paper where the log of hourly wages is the dependent variable and cognitive ability measured by the ASVAB test score (averaged over the ten tests administered), years of tenure, and years of education act as the regressors. All variables are standardized to have unit variance. To allow for the effect of race to be unconstrained, we estimate the model separately for blacks and whites. The sample size for blacks is 600, and for whites 1232.

We consider the ASVAB score to be a noisy measure of cognitive ability, where the noise is unrelated to the true underlying ability (see Dong, Otsu and Taylor, 2020, for evidence to the suitability of the classical measurement error assumption in this context). In contrast to the Monte Carlo evidence in Section 5, here we assume the distribution of the measurement error to be known, thus, we must specify choices. We assume the error distribution is normal and give results for a standard deviation of 0.1 and 0.3, respectively; this allows an examination of the sensitivity of the results to this choice. The parameters for the estimator are chosen in the same manner as in Section 5. In Figures 6.1 - 6.3, we plot each of the additive functions between the 1% and 99% quantile of the respective regressor. The blue lines refer to results for black individuals and the red lines for white individuals. The solid lines use a standard deviation of 0.1 for the measurement error and the dashed lines use 0.3.

First, it is interesting to see that the estimates for whites are quite insensitive to the assumed standard deviation of the measurement error. However, for blacks, this is not the case for all estimated functions; this likely reflect the smaller sample size for black men. Unsurprisingly, there is a positive effect for each of cognitive ability, tenure, and education on wages, with cognitive ability showing the largest effect. It is also clear from these plots that the relationship between each regressor and the outcome is nonlinear. Thus, a nonparametric analysis seems necessary in this setting. Finally, in line with much of the previous literature, the estimated wage-returns for these different attributes are higher for blacks than whites (see, for example, Lang and Manove, 2011).

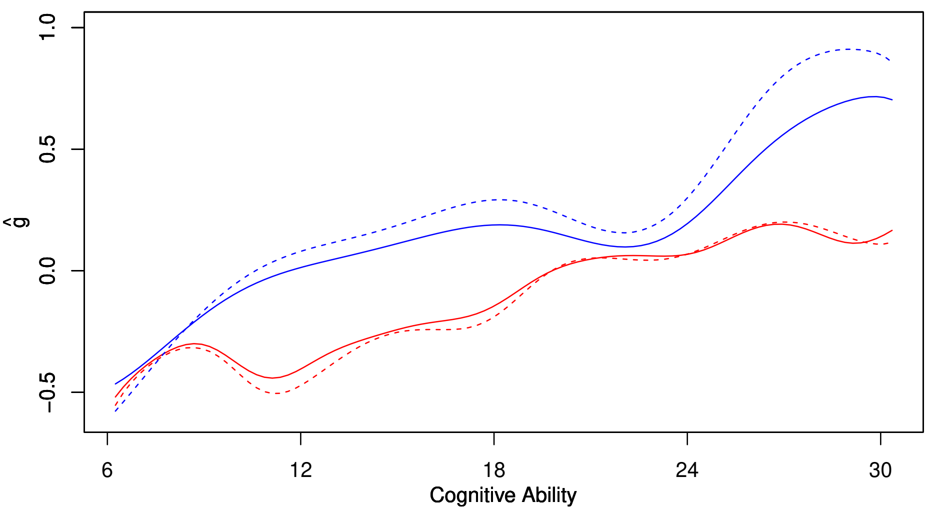

FIGURE 6.1. Plot of $\hat{g}$ between the 1% and 99% quantile of the ASVAB test score using the NLSY79 dataset. The blue lines give results for black individuals and the red lines for white individuals. The solid lines use a standard deviation of 0.1 for the measurement error and the dashed lines use 0.3.

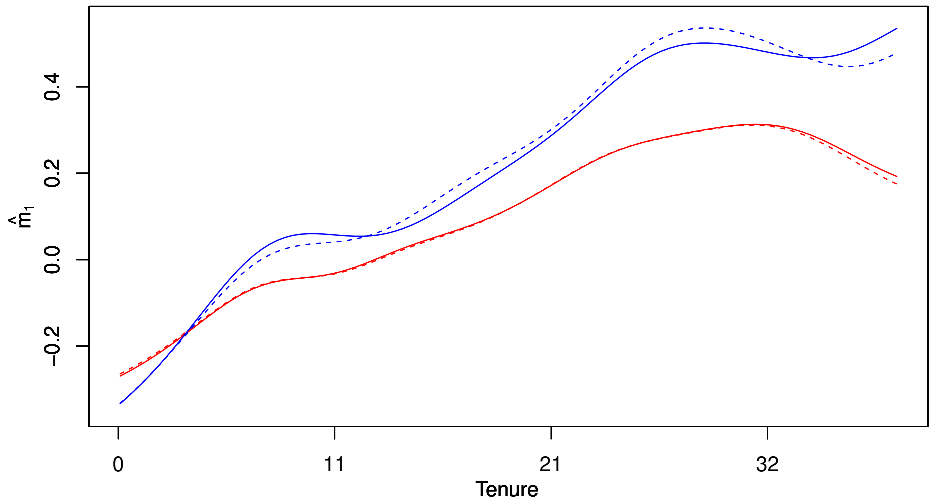

FIGURE 6.2. Plot of $\hat{m}_1$ between the 1% and 99% quantile of tenure (in years) using the NLSY79 dataset. The blue lines give results for black individuals and the red lines for white individuals. The solid lines use a standard deviation of 0.1 for the measurement error and the dashed lines use 0.3.

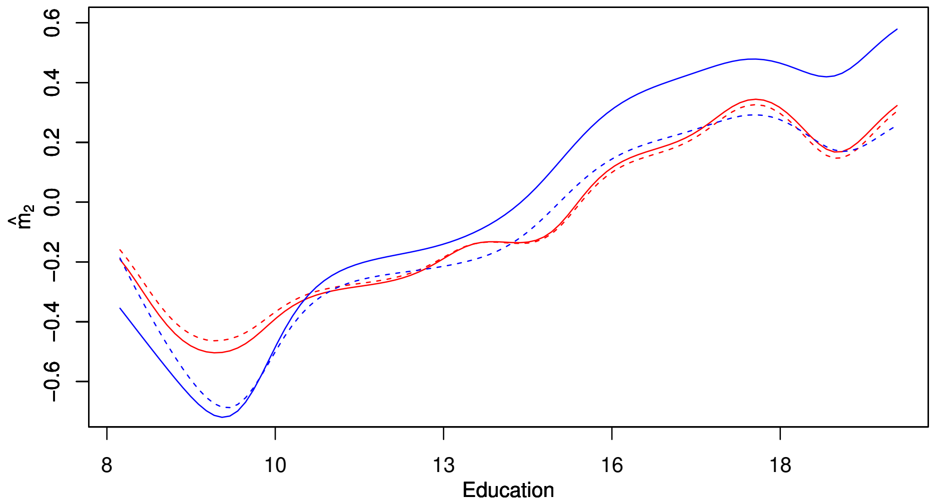

FIGURE 6.3. Plot of $\hat{m}_2$ between the 1% and 99% quantile of education (in years) using the NLSY79 dataset. The blue lines give results for black individuals and the red lines for white individuals. The solid lines use a standard deviation of 0.1 for the measurement error and the dashed lines use 0.3.

7. Conclusion

In this paper, we develop a novel nonparametric estimation strategy for additive models where one covariate is contaminated with classical measurement error. The estimation procedure is divided into two stages. In the first stage, to adapt to the additive structure of the regression function, we derive the first-stage estimator based on an orthonormal series approximation combined with a ridge parameter deconvolution; the ridge approach being used to deal with the ill-posedness brought by the measurement error.

The uniform convergence rate of our first-stage estimator is separately derived for cases of ordinary/super smooth measurement error. In particular, we find that the presence of measurement error slows down the convergence rate in general. In the case of ordinary smooth measurement error, our first-stage estimator can achieve a uniform convergence rate as fast as $n^{-\frac{6}{19+10\beta}}$, which is faster than $n^{-\frac{1}{4+4\beta}}$ as in Han and Park (2018) under the same smoothness condition when $\alpha = 2$ and $\beta > 1/2$. In the case of supersmooth measurement error which was not addressed by Han and Park (2018), our first-stage estimator can achieve a uniform convergence rate as fast as $(\log n)^{-\frac{3\alpha}{5\gamma}}$. These rate results, however, are slower than the error-free rate $n^{-\frac{3}{10}}$ as in Horowitz and Mammen (2004), which is expected given the contaminated nature of the sample. To establish the limiting distribution - which is important for statistical inference - we consider the second-stage estimator obtained by one-step backfitting using a deconvolution kernel based on our first-stage estimator. The method of constructing our second-stage estimator, however, depends on whether the nonparametric component to be estimated is associated with a mismeasured covariate or a correctly measured one. The asymptotic normality is established for both types of second-stage estimator, and for cases of ordinary/super smooth measurement error. Finally, a Monte Carlo study and an empirical application highlight the applicability of our estimator.

Further research is needed to explore optimal convergence rates, adaptive estimation, and extensions to models with non-identity link functions.

Appendix A. Proofs for Section2

A.1. **Proof of Theorem 1.** Let $z = (z_1, \ldots, z_D)$, $z_{-d} = (z_1, \ldots, z_{d-1}, z_{d+1}, \ldots, z_D)$, $A(\mathcal{I})$ be the length of the set $\mathcal{I}$, and $f^{\text{ft}}_{Y,X,Z}(y, \cdot, z)(t) = \int f_{Y,X,Z}(y, x, z)e^{\text{i}tx}dx$, where $f_{Y,X,Z}$ is the density of (Y, X, Z). By Assumption 1 (1) and Lemma 1 (2), $f_{Y,X^*,Z}$ is identified as

$$f_{Y,X^*,Z}(y, x^*, z) = \frac{1}{2\pi}\int e^{-\text{i}tx^*}\frac{f^{\text{ft}}_{Y,X,Z}(y, \cdot, z)(t)}{f^{\text{ft}}_{\epsilon}(t)}dt,$$

and the conditional mean $E[Y|X^*, Z]$ is also identified under Assumption 1 (2). Thus, using Assumption 1 (3), the intercept μ and the functions g and m_d for $d = 1, \ldots, D$ are identified as

$$\begin{aligned}
\mu &= 2^{-D}A(\mathcal{I})^{-1}\iint_{\mathcal{I}\times[-1,1]^D} E[Y|X^* = x^*, Z = z]dx^*dz,\\
g(x^*) &= 2^{-D}\int_{[-1,1]^D} E[Y|X^* = x^*, Z = z]dz - \mu,\\
m_d(z_d) &= 2^{-(D-1)}\int_{[-1,1]^{D-1}} E[Y|X^* = x^*, Z = z]dz_{-d} - \mu - g(x^*).
\end{aligned}$$

Appendix B. Proofs for Section 3

B.1. **Proof of Theorem 2 .** First, we show the convergence rate of $\|\hat{\theta} - \theta^*\|^2$. Let $\hat{M}_\kappa = \Re\hat{E}[P_\kappa P_\kappa']$, $\hat{C}_\kappa = \Re\hat{E}[YP_\kappa']$, $M_\kappa = E[P_\kappa P_\kappa']$, $C_\kappa = E[P_\kappa Y]$, $\theta^* = M_\kappa^{-1}C_\kappa$, and $r_\kappa = E[Y|X^*, Z] - P_\kappa'\theta_0$. Observe that

$$\begin{aligned}
\|\hat{\theta} - \theta^*\|^2 &= \|\hat{M}_\kappa^{-1}\hat{C}_\kappa - M_\kappa^{-1}C_\kappa\|^2 = \|\hat{M}_\kappa^{-1}(\hat{C}_\kappa - C_\kappa) + \hat{M}_\kappa^{-1}(M_\kappa - \hat{M}_\kappa)\theta^*\|^2\\
&\leq 2\|\hat{M}_\kappa^{-1}(\hat{C}_\kappa - C_\kappa)\|^2 + 2\|\hat{M}_\kappa^{-1}(M_\kappa - \hat{M}_\kappa)\theta^*\|^2\\
&\leq 2\lambda_{\max}(\hat{M}_\kappa^{-2})\{\|\hat{C}_\kappa - C_\kappa\|^2 + \|\hat{M}_\kappa - M_\kappa\|^2\|\theta^*\|^2\},
\end{aligned}$$

where the penultimate step follows from Jensen's inequality, and the last step follows from $\lambda_{\max}(A) = \sup_{\|\delta\|=1}\delta'A\delta$ and $\lambda_{\max}(A'A) \leq \|A\|^2$.

Since $\|\hat{M}_\kappa - M_\kappa\|^2 \leq \|\hat{E}[P_\kappa P_\kappa'] - M_\kappa\|^2$ and $\|\hat{C}_\kappa - C_\kappa\|^2 \leq \|\hat{E}[P_\kappa Y] - C_\kappa\|^2$, the orders of $\|\hat{M}_\kappa - M_\kappa\|^2$ and $\|\hat{C}_\kappa - C_\kappa\|^2$ follow from Lemma 4. For $\lambda_{\max}(\hat{M}_\kappa^{-2})$, we have

$$\lambda_{\max}(\hat{M}_\kappa^{-2}) = \lambda_{\min}^{-2}(\hat{M}_\kappa) \geq \{\inf_{\|\delta\|=1}\delta'(\hat{M}_\kappa - M_\kappa)\delta + \lambda_{\min}(M_\kappa)\}^2,$$

where $\left(\inf_{\|\delta\|=1}\delta'(\hat{M}_\kappa - M_\kappa)\delta\right)^2 \leq \|\hat{M}_\kappa - M_\kappa\|^2 \xrightarrow{p} 0$ under Assumption 3 or 4, and $\lambda_{\min}(M_\kappa) \geq \underline{\lambda} > 0$ as in Assumption 2 (4). For $\|\theta^*\|^2$, we have

$$\|\theta^*\|^2 = C_\kappa' M_\kappa^{-2}C_\kappa \leq \lambda_{\max}(M_\kappa^{-1})C_\kappa M_\kappa^{-1}C_\kappa \leq \underline{\lambda}^{-1}E[E[Y|X^*, Z]^2] < \infty,$$

where the third step follows from $C_\kappa = E[P_\kappa E[Y|X^*, Z]]$ and Theorem 1 of Tripathi (1999), and the last step follows from the boundedness of $g, m_1, \cdots, m_D$. Combining these results, we obtain

$$\|\hat{\theta} - \theta^*\|^2 = \begin{cases} O_p\left(\kappa^2 n^{2\zeta+\frac{\zeta}{\beta}-1} + \kappa n^{-\frac{2\alpha\zeta}{\beta}}\right) & \text{under Assumption 3,}\\ O_p\left(\kappa(\log n)^{-\frac{2\alpha}{\gamma}}\right) & \text{under Assumption 4.}\end{cases}$$

Therefore, the convergence rate of $\|\hat{\theta}-\theta_0\|$ follows from the triangle inequality and

$$\|\theta^*-\theta_0\|^2 = E[P_\kappa' r_k]M_\kappa^{-2}E[P_\kappa r_k] \le \lambda_{\max}(M_\kappa^{-1})E[P_\kappa' r_k]M_\kappa^{-1}E[P_\kappa r_k] \le \underline{\lambda}^{-1}E[r_k^2] = O(\kappa^{-4}),$$

where the first step follows from $\theta^* = \theta_0 + M_\kappa^{-1}E[P_\kappa r_k]$, the third step follows from Theorem 1 of Tripathi (1999), and the last step follows from Assumption 2 (4).

Next, we prove the convergence rates of $\hat{g}$ and $\hat{m}_d$. Since the proof is similar, we focus on $\hat{g}$. Let $\hat{\theta}^0$ be the vector of estimated coefficients in $\hat{\theta}$ corresponding to $P_{\kappa,0}$. Then, we have

$$\begin{aligned}
\sup_{x^*\in\mathcal{I}} |\hat{g}(x^*)-g(x^*)| &\le \sup_{x^*\in\mathcal{I}} |P_{\kappa,0}(x^*)'(\hat{\theta}^0-\theta_0^0)| + \sup_{x^*\in\mathcal{I}} |g(x^*)-P_{\kappa,0}'(X^*)\theta_0^0| \\
&\le \sup_{x^*\in\mathcal{I}} \|P_{\kappa,0}(x^*)\| \cdot \|\hat{\theta}^0-\theta_0^0\| + O(\kappa^{-2}) \\
&= \begin{cases} O_p\left(\kappa^{\frac{3}{2}}n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right) & \text{under Assumption 3,} \\ O_p\left(\kappa(\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-\frac{3}{2}}\right) & \text{under Assumption 4,} \end{cases}
\end{aligned}$$

where the second step uses the Cauchy-Schwartz inequality and Assumption 2 (4), and the last step follows from $\sup_{x^*\in\mathcal{I}}\|P_{\kappa,0}(x^*)\| \le \sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D}\|P_\kappa(x^*,z)\|$, Assumption 2 (4), $\|\hat{\theta}^0-\theta_0^0\| \le \|\hat{\theta}-\theta_0\|$, and the convergence rate of $\|\hat{\theta}-\theta_0\|$.

B.2. **Proof of Theorem 3 .** To simplify the presentation, in the following discussion we suppress dependence on x^*, the point at which g is evaluated. Let $\mathbb{A}_n = \frac{1}{n}\sum_{j=1}^n \mathbb{K}_h(x^*-X_j)$ and $a = f_{X^*}(x^*)\int K(w)dw$. Decompose $\tilde{g}-g = \frac{1}{n}\sum_{j=1}^n G_{n,j}$, where $G_{n,j} = G_{1,n,j}+G_{2,n,j}+G_{3,n,j}+G_{4,n,j}$ and

$$\begin{aligned}
G_{1,n,j} &= \frac{1}{2\pi a}\int e^{-\mathrm{i}t(x^*-X_j)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}\Big[Y_j-\mu-\sum_{d=1}^D m_d(Z_{d,j})-g(x^*)\Big]dt, \\
G_{2,n,j} &= \frac{1}{2\pi a}\int e^{-\mathrm{i}t(x^*-X_j)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}\Big[\mu+\sum_{d=1}^D m_d(Z_{d,j})-\hat{\mu}-\sum_{d=1}^D \hat{m}_d(Z_{d,j})\Big]dt, \\
G_{3,n,j} &= \frac{a-\mathbb{A}_n}{\mathbb{A}_n}G_{1,n,j}, \qquad G_{4,n,j} = \frac{a-\mathbb{A}_n}{\mathbb{A}_n}G_{2,n,j}.
\end{aligned}$$

The proof is divided into three steps. First, we consider the case where f_ϵ is ordinary smooth.

Step 1: Show

$$\frac{\sum_{j=1}^n G_{1,n,j} - nE[G_{1,n,1}]}{\sqrt{nVar[G_{1,n,1}]}} \xrightarrow{d} N(0,1). \tag{B.1}$$

By Lyapunov's central limit theorem, it is sufficient for (B.1) to show

$$\lim_{n\to\infty}\frac{E|G_{1,n,1}|^{2+\eta}}{n^{\eta/2}\Big[E|G_{1,n,1}|^2\Big]^{(2+\eta)/2}} = 0, \tag{B.2}$$

for some constant $\eta > 0$. Let $\mu_{g,2+\eta}(x) = E[|g(X^*)+U-g(x^*)|^{2+\eta}|X=x]f_X(x)$. By the law of iterated expectations, we can write $E|G_{1,n,1}|^{2+\eta}$ as

$$E|G_{1,n,1}|^{2+\eta} = \int_x\left|\frac{1}{2\pi a}\int_t e^{-\mathrm{i}t(x^*-x)}\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\right|^{2+\eta}\mu_{g,2+\eta}(x)dx. \tag{B.3}$$

If $\eta > 0$, we have

$$\begin{aligned} & E|G_{1,n,1}|^{2+\eta} \\ \leq\ & \frac{h^{-(\beta+1)\eta}}{(2\pi)^\eta a^{(2+\eta)}} \left(h^{\beta+1} \int \frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|} dt \right)^\eta \times \frac{h^{2\beta+1}}{4\pi^2} \int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 \mu_{g,2+\eta}(x) dx \\ =\ & O(h^{-(\beta+1)(\eta+2)+1}), \end{aligned} \tag{B.4}$$

where the equality follows by Lemmas 5 and 7. On the other hand, if $\eta = 0$, we have

$$\begin{aligned} E|G_{1,n,1}|^2 &= \frac{h^{-(2\beta+1)}}{a^2} \left(\frac{h^{2\beta+1}}{4\pi^2} \int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 \mu_{g,2+\eta}(x) dx \right) \\ &= \frac{h^{-(2\beta+1)} \mu_{g,2}(x^*)}{2\pi a^2 c_\epsilon^2} \int |s|^{2\beta} |K^{\mathrm{ft}}(s)|^2 ds \{1 + o_p(1)\}, \end{aligned} \tag{B.5}$$

where the second equality follows by Lemma 7. Thus, (B.4) and (B.5) together imply that (B.1) holds true if $nh \to \infty$ as $n \to \infty$.

Step 2: Show

$$\frac{\sum_{j=1}^n G_{2,n,j} - nE[G_{2,n,1}]}{\sqrt{nVar[G_{1,n,1}]}} \xrightarrow{p} 0. \tag{B.6}$$

For the numerator, we note

$$\sum_{j=1}^n G_{2,n,j} - nE[G_{2,n,1}] = O_p \left(\sqrt{nE|G_{2,n,1}|^2} \right), \tag{B.7}$$

and

$$\begin{aligned} E|G_{2,n,1}|^2 &= \int_x E \left[\left| \mu + \sum_{d=1}^D m_d(Z_{d,1}) - \hat{\mu} - \sum_{d=1}^D \hat{m}_d(Z_{d,1}) \right|^2 \middle| X = x \right] \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 f_X(x) dx \\ &\leq \left(|\hat{\mu} - \mu| + \sum_{d=1}^D \sup_{z_d \in [-1,1]} |\hat{m}_d(z_d) - m_d(z_d)| \right)^2 \\ &\quad \times 4\pi^2 h^{-(2\beta+1)} \left\{ \frac{h^{2\beta+1}}{4\pi^2} \int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 f_X(x) dx \right\} \\ &= O_p \left(\kappa^3 n^{2\zeta + \frac{\zeta}{\beta} - 1} h^{-(2\beta+1)} + \kappa^2 n^{-\frac{2\alpha\zeta}{\beta}} h^{-(2\beta+1)} + \kappa^{-3} h^{-(2\beta+1)} \right), \end{aligned} \tag{B.8}$$

where the last equality follows by Theorem 2 and Lemma 7. For the denominator,

$$\begin{aligned} aE[G_{1,n,1}] &= \frac{1}{2\pi} \int e^{-\mathrm{i}tx^*} K^{\mathrm{ft}}(th) \{E[e^{\mathrm{i}tX^*} g(X^*)] - E[e^{\mathrm{i}tX^*}] g(x^*)\} dt \\ &= E[K_h(x^* - X^*) g(X^*)] - E[K_h(x^* - X^*)] g(x^*) \\ &= \int K_h(x^* - w) g(w) f_{X^*}(w) dw - g(x^*) \int K_h(x^* - w) f_{X^*}(w) dw, \\ &= O(h^2), \end{aligned} \tag{B.9}$$

where the last equality follows by the second-order differentiability of f_{X^*}, the third-order differentiability of g, the symmetry of K, $\int K(w)w^2dw < \infty$, and the fact that

$$\begin{aligned}&\int K_h(x^*-w)g(w)f_{X^*}(w)dw - g(x^*)\int K_h(x^*-w)f_{X^*}(w)dw \\ =\ & f_{X^*}(x^*)g''(x^*)\int K(w)w^2dwh^2 + o(h^2).\end{aligned}$$

Then (B.9) and (B.5) imply that $Var[G_{1,n,1}]$ is strictly dominated by $E|G_{1,n,1}|^2$ for large n. Now by (B.5), we have

$$\frac{1}{Var[G_{1,n,1}]} = O(h^{(2\beta+1)}). \tag{B.10}$$

Thus, (B.6) holds true if $\kappa^{\frac{3}{2}}n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}} \to 0$ as $n \to \infty$.

Step 3: Show

$$\frac{\sum_{j=1}^n G_{k,n,j} - nE[G_{k,n,1}]}{\sqrt{nVar[G_{1,n,1}]}} \xrightarrow{p} 0, \tag{B.11}$$

for $k = 3, 4$. For this, it is sufficient to show $\mathbb{A}_n - a = o_p(1)$. To see this, note

$$\mathbb{A}_n = E[\mathbb{A}_n] + O_p\left(n^{-1/2}\left[E|\mathbb{K}_h(x^*-X)|^2\right]^{1/2}\right). \tag{B.12}$$

For the first term in (B.12), we have

$$\begin{aligned}E[\mathbb{A}_n] &= E\left[\frac{1}{2\pi}\int\frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)}e^{-\text{i}t(x^*-X)}dt\right] = \frac{1}{2\pi}\int e^{-\text{i}tx^*}K^{\text{ft}}(th)f_{X^*}^{\text{ft}}(t)dt \\ &= E[K_h(x^*-X^*)] = \int K(u)f_{X^*}(x^*-uh)du = a + O(h), \end{aligned} \tag{B.13}$$

where the second equality follows by Assumption 1 (1), the third equality follows by Plancherel's isometry (Lemma 1 (1)), the fourth equality follows by a change of variables, and the last equality follows by the differentiability of f_{X^*}. For the second term in (B.12), by Lemma 7, we have $E|\mathbb{K}_h(x^*-X)|^2 = O(h^{-(2\beta+1)})$ and thus

$$\mathbb{A}_n - a = O(h) + O_p(n^{-1/2}h^{-(\beta+1/2)}), \tag{B.14}$$

which implies that (B.11) follows by (B.1) and (B.6) if $h \to 0$ and $nh^{2\beta+1} \to \infty$.

Combining (B.1), (B.6), and (B.11), we have

$$\frac{\tilde{g}(x^*) - g(x^*) - \text{Bias}\{\tilde{g}(x^*)\}}{\sqrt{Var[G_{1,n,1}]}} \xrightarrow{d} N(0,1),$$

where $\text{Bias}\{\tilde{g}(x^*)\} = E[G_{n,1}]$. To conclude for the ordinary smooth case, note $Var[\tilde{g}(x^*)] = \frac{1}{n}Var\left[\sum_{k=1}^4 G_{k,n,1}\right]$. By the Cauchy-Schwartz inequality, the covariance terms are dominated by the variance terms, then for $Var[\tilde{g}(x^*)]/Var[G_{1,n,1}] \xrightarrow{p} 1$, it is sufficient to show $Var[G_{k,n,1}]/Var[G_{1,n,1}] \xrightarrow{p} 0$ for $k = 2, 3, 4$, which immediately follows by (B.8), (B.10), and (B.12).

The proof for the supersmooth case is similar to that of the ordinary smooth case, so we only state the differences here. First, we update the upper bound results. In Step 1 of the ordinary smooth case, to verify the Lyapunov condition (B.2), by (B.3), parallel to (B.4), for $\eta > 0$, we have

$$
\begin{aligned}
E|G_{1,n,1}|^{2+\eta} &\leq \frac{\sup_x \mu_{g,2+\eta}(x)}{(2\pi a)^{2+\eta}} \left(\int \frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|} dt \right)^\eta \int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 dx \\
&= O\left(h^{-(1+\eta)} e^{\mu(2+\eta)h^{-\gamma}} \right), \qquad \text{(B.15)}
\end{aligned}
$$

where the last equality follows by Lemma 8 and $\sup_x \mu_{g,2+\eta}(x) < \infty$. For the latter, we note $\|g\|_\infty < c_g$ for some $c_g > 0$ and

$$
|g(X^*) + U - g(x^*)|^{2+\eta} \leq \{|g(X^*)| + |U| + |g(x^*)|\}^{2+\eta} \leq \{2c_g + |U|\}^{2+\eta} \leq c_1 + c_2|U|^{2+\eta},
$$

for constants $c_1 = 2^{1+\eta}(2c_g)^{2+\eta}$ and $c_2 = 2^{1+\eta}$. Hence, $\sup_x \mu_{g,2+\eta}(x) < \infty$ follows by $\|f_X\|_\infty < \infty$ and $\sup_x E[|U|^{2+\eta}|X = x] < \infty$. By a similar argument as in (B.15), we have

$$
\begin{aligned}
\int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 f_X(x) dx &\leq \|f_X\|_\infty \int_x \left| \int_t e^{-\mathrm{i}t(x^*-x)} \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt \right|^2 dx \\
&= O\left(h^{-1} e^{2\mu h^{-\gamma}} \right), \qquad \text{(B.16)}
\end{aligned}
$$

where the equality follows by $\|f_X\|_\infty < \infty$ and Lemma 8. Therefore, for the parallel result to (B.8), by Theorem 2 and (B.16),

$$
E|G_{2,n,1}|^2 = O_p\left(\kappa(\log n)^{-\frac{\alpha}{\beta}} h^{-1} e^{2\mu h^{-\gamma}} + \kappa^{-\frac{3}{2}} h^{-1} e^{2\mu h^{-\gamma}} \right). \qquad \text{(B.17)}
$$

For the parallel result to (B.12), using (B.16), we have

$$
\mathbb{A}_n - a = O(h) + O_p\left(n^{-1/2} h^{-1/2} e^{\mu h^{-\gamma}} \right), \qquad \text{(B.18)}
$$

which implies that (B.11) still holds if $h \to 0$ and $nhe^{-2\mu h^{-\gamma}} \to \infty$.

To verify Lyapunov's condition (B.2) and to check that the first-stage estimation error is negligible as in (B.6), besides (B.15), we also need the parallel result to (B.5). However, it is difficult to derive the parallel result to Lemma 7 in general for the case of supersmooth f_ϵ. In the deconvolution literature, the lower bound of $E|G_{1,n,1}|^2$ is commonly used to verify (B.2) in the case of supersmooth f_ϵ. Primitive conditions, like Fan and Masry (1992, Condition 3.1), can be imposed to this end. In this paper, to avoid unnecessary complication, we directly assume the lower bound of $E|G_{1,n,1}|^2$ in Assumption 8 (3). Hence, under Assumption 8 (3), both (B.2) and (B.6) hold true, and the conclusion follows.

B.3. **Proof of Theorem 4 .** Similar to the proof of Theorem 3, in the following discussion we suppress dependence on z_d, the point at which m_d is evaluated. Let $\mathbb{A}_n^d = \frac{1}{n}\sum_{j=1}^n \int_{x^* \in \mathcal{I}} \mathbb{K}_h(x^* - X_j) dx^* K_h(z_d - Z_{d,j})$ and $a^d = \int_{x^* \in \mathcal{I}} f_{X^*, Z_d}(x^*, z_d) dx^* \left(\int K(w) dw \right)^2$. First, similar to the proof of Theorem 3, we have $\tilde{m}_d(z_d) - m_d(z_d) = \frac{1}{n}\sum_{j=1}^n G_{n,j}^d$, where $G_{n,j}^d = G_{1,n,j}^d + G_{2,n,j}^d + G_{3,n,j}^d + G_{4,n,j}^d$

and

$$
\begin{aligned}
G^d_{1,n,j} &= \frac{1}{a^d}\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X_j)dx^* K_h(z_d-Z_{d,j})\Big[Y_j-\mu-\sum_{d'\neq d} m_{d'}(Z_{d',j})-m_d(z_d)\Big] \\
&\quad -\frac{1}{a^d}\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X_j)g(x^*)dx^* K_h(z_d-Z_{d,j}), \\
G^d_{2,n,j} &= \frac{1}{a^d}\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X_j)dx^* K_h(z_d-Z_{d,j})\Big[\mu+\sum_{d'\neq d} m_{d'}(Z_{d',j})-\hat{\mu}-\sum_{d'\neq d}\hat{m}_{d'}(Z_{d',j})\Big] \\
&\quad -\frac{1}{a^d}\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X_j)\{\hat{g}(x^*)-g(x^*)\}dx^* K_h(z_d-Z_{d,j}), \\
G^d_{3,n,j} &= \frac{a^d-\mathbb{A}^d_n}{\mathbb{A}^d_n}G^d_{1,n,j}, \qquad G^d_{4,n,j} = \frac{a^d-\mathbb{A}^d_n}{\mathbb{A}^d_n}G^d_{2,n,j},
\end{aligned}
$$

and the rest of the proof follows in three steps. First, we consider the ordinary smooth case.

Step 1: Show

$$
\lim_{n\to\infty}\frac{E|G^d_{1,n,1}|^{2+\eta}}{n^{\eta/2}\Big[E|G^d_{1,n,1}|^2\Big]^{(2+\eta)/2}}=0, \tag{B.19}
$$

for some constant $\eta>0$. For the numerator, by Jensen's inequality,

$$
\begin{aligned}
E|G^d_{1,n,1}|^{2+\eta} &\leq \frac{2^{(1+\eta)}}{(a^d)^{2+\eta}}E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* K_h(z_d-Z_d)\{m_d(Z_d)+g(X^*)+U-m_d(z_d)\}\right|^{2+\eta} \\
&\quad +\frac{2^{(1+\eta)}}{(a^d)^{2+\eta}}E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)g(x^*)dx^* K_h(z_d-Z_d)\right|^{2+\eta}.
\end{aligned}
$$

For the first term, we have

$$
\begin{aligned}
& E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* K_h(z_d-Z_d)\{m_d(Z_d)+g(X^*)+U-m_d(z_d)\}\right|^{2+\eta} \\
&= O\left(h^{-\eta}\left(\int\frac{|K^{\mathrm{ft}}(th)|}{|f^{\mathrm{ft}}_\epsilon(t)|}dt\right)^\eta E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* K_h(z_d-Z_d)\right|^2\right) = O(h^{-(\eta+2)\beta-2\eta}),
\end{aligned}
$$

where the first equality follows by the law of iterated expectations, $\|m_d\|_\infty<\infty$, $\|g\|_\infty<\infty$, $\sup_{u,v}E[|U|^{2+\eta}|X=u,Z_d=v]<\infty$, and

$$
\begin{aligned}
& E\left[|m_d(Z_d)+g(X^*)+U-m_d(z_d)|^{2+\eta}\,|X=u,Z_d=v\right] \\
\leq\; & 4^{1+\eta}\left(2\|m_d\|_\infty^{2+\eta}+\|g\|_\infty^{2+\eta}+E[|U|^{2+\eta}|X=u,Z_d=v]\right),
\end{aligned}
$$

and the second equality follows by Lemmas 5 and 9.

By a very similar argument, we have

$$
E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)g(x^*)dx^* K_h(z_d-Z_d)\right|^{2+\eta} = O(h^{-(\eta+2)\beta-2\eta}),
$$

which implies $E|G^d_{1,n,1}|^{2+\eta}=O(h^{-(\eta+2)\beta-2\eta})$. Also, by Lemma 9, there exists a constant $c>0$ such that $E|G^d_{1,n,1}|^2\geq ch^{-2\beta}$ for all n large enough. Thus, (B.19) holds true if $nh^4\to\infty$ as $n\to\infty$.

Step 2: Show

$$\frac{E|G^d_{2,n,1}|^2}{Var(G^2_{1,n,1})} \to 0. \tag{B.20}$$

For the numerator, we have

$$\begin{aligned} E|G^d_{2,n,1}|^2 &\leq \frac{2}{a^{2d}} E\left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* K_h(z_d - Z_d)\big[\mu + \sum_{d'\neq d} m_{d'}(Z_{d'}) - \hat{\mu} - \sum_{d'\neq d} \hat{m}_{d'}(Z_{d'})\big] \right|^2 \\ &\quad + \frac{2}{a^{2d}} E\left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)\{\hat{g}(x^*) - g(x^*)\}dx^* K_h(z_d - Z_d) \right|^2. \end{aligned} \tag{B.21}$$

For the first term, we have

$$\begin{aligned} & E\left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* K_h(z_d - Z_d)\big[\mu + \sum_{d'\neq d} m_{d'}(Z_{d'}) - \hat{\mu} - \sum_{d'\neq d} \hat{m}_{d'}(Z_{d'})\big] \right|^2 \\ = & \int_{u,v} E\left[\left| \mu + \sum_{d'\neq d} m_{d'}(Z_{d'}) - \hat{\mu} - \sum_{d'\neq d} \hat{m}_{d'}(Z_{d'}) \right|^2 \middle| X = u, Z_d = v \right] \\ & \times \left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-u)dx^* K_h(z_d - v) \right|^2 f_{X,Z_d}(u,v)dudv \\ \leq & \left(|\hat{\mu} - \mu| + \sum_{d'\neq d} \sup_{z_{d'}\in[-1,1]} |\hat{m}_{d'}(z_{d'}) - m_{d'}(z_{d'})| \right)^2 h^{-2\beta} \left\{ h^{2\beta} E\left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* K_h(z_d - Z_d) \right|^2 \right\} \\ = & O_p\left(h^{-2\beta}\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + h^{-2\beta}\kappa n^{-\frac{\alpha\zeta}{\beta}} + h^{-2\beta}\kappa^{-\frac{3}{2}} \right), \end{aligned}$$

where the first equality follows by the law of iterated expectations and the last equality follows by Theorem 2 and Lemma 9. By a similar argument, we have

$$\begin{aligned} & E\left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)\{\hat{g}(x^*) - g(x^*)\}dx^* K_h(z_d - Z_d) \right|^2 \\ = & O_p\left(h^{-2\beta}\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + h^{-2\beta}\kappa n^{-\frac{\alpha\zeta}{\beta}} + h^{-2\beta}\kappa^{-\frac{3}{2}} \right), \end{aligned}$$

which implies $E|G^d_{2,n,1}|^2 = O_p\left(h^{-2\beta}\kappa^{\frac{3}{2}} n^{\zeta + \frac{\zeta}{2\beta} - \frac{1}{2}} + h^{-2\beta}\kappa n^{-\frac{\alpha\zeta}{\beta}} + h^{-2\beta}\kappa^{-\frac{3}{2}} \right)$. For the denominator, by Lemma 9, we have $E|G^d_{1,n,1}|^2 \geq ch^{-2\beta}$. Also, we note

$$\begin{aligned} a^d E[G^d_{1,n,1}] &= \int_{x^*\in\mathcal{I}} \frac{1}{2\pi} \int e^{-\mathrm{i}tx^*} K^{\mathrm{ft}}(th) \left\{ E[\{m_d(Z_d) + g(X^*)\}K_h(z_d - Z_d)|X^*]f_{X^*} \right\}^{\mathrm{ft}}(t)dtdx^* \\ &\quad - \int_{x^*\in\mathcal{I}} \frac{m_d(z_d) + g(x^*)}{2\pi} \int e^{-\mathrm{i}tx^*} K^{\mathrm{ft}}(th) \left\{ E[K_h(z_d - Z_d)|X^*]f_{X^*} \right\}^{\mathrm{ft}}(t)dtdx^* \\ &= \int_{x^*\in\mathcal{I}} E[\{m_d(Z_d) + g(X^*)\}K_h(x^* - X^*)K_h(z_d - Z_d)]dx^* \\ &\quad - \int_{x^*\in\mathcal{I}} \{m_d(z_d) + g(x^*)\}E[K_h(x^* - X^*)K_h(z_d - Z_d)]dx^* = O(h^2), \end{aligned} \tag{B.22}$$

where the first equality follows by Assumption 1 (1), the second equality follows by the convolution theorem (Lemma 1 (2)), and the last equality follows by the twice continuous differentiability of g, m_d, and f_{X^*,Z_d}, the symmetry of K, $\int K(w)w^2dw < \infty$, and the following fact

$$\begin{aligned}
& \int K_h(x^* - w_1)K_h(z_d - w_2)\{g(w_1) + m_d(w_2)\}f_{X^*,Z_d}(w_1, w_2)dw \\
& -\{g(x^*) + m_d(z_d)\}\int K_h(x^* - w_1)K_h(z_d - w_2)f_{X^*,Z_d}(w_1, w_2)dw \\
= & \int K(w_1)K(w_2)\big[g(x^* - w_1h) + m_d(z_d - w_2h)\big]f_{X^*,Z_d}(x^* - w_1h, z_d - w_2h)dw \\
& -\{g(x^*) + m_d(z_d)\}\int K(w_1)K(w_2)f_{X^*,Z_d}(x^* - w_1h, z_d - w_2h)dw \\
= & f_{X^*,Z_d}(x^*, z_d)\{g''(x^*) + m_d''(z_d)\}\int K(w)w^2dw\int K(w)dwh^2 + o(h^2).
\end{aligned}$$

Since $Var[G_{1,n,1}]$ is dominated by $E|G_{1,n,1}|^2$, we obtain

$$\frac{1}{Var[G^d_{1,n,1}]} = O(h^{2\beta}). \tag{B.23}$$

Thus, (B.20) holds true if $\kappa^{\frac{3}{2}}n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}} \to 0$ as $n \to \infty$.

Step 3: Show

$$\mathbb{A}^d_n - a^d = o_p(1). \tag{B.24}$$

To see this, we note

$$\mathbb{A}^d_n = E[\mathbb{A}^d_n] + O_p\left(n^{-1/2}\left[E\left|\int_{x^*\in\mathcal{I}}\mathbb{K}_h(x^* - X)dx^*K_h(z_d - Z_d)\right|^2\right]^{1/2}\right).$$

For the first term $E[\mathbb{A}^d_n]$, we have

$$\begin{aligned}
E[\mathbb{A}^d_n] &= \int_{x^*\in\mathcal{I}}\frac{1}{2\pi}\int_t e^{-\mathrm{i}tx^*}K^{\mathrm{ft}}(th)\big\{E[K_h(z_d - Z_d)|X^*]f_{X^*}\big\}^{\mathrm{ft}}(t)dtdx^* \\
&= \int_{x^*\in\mathcal{I}}E[K_h(x^* - X^*)K_h(z_d - Z_d)]dx^* \\
&= \int_{x^*\in\mathcal{I}}\int_{u,v}K(u)K(v)f_{X^*,Z_d}(x^* - uh, z_d - vh)dudvdx^* = a^d + O(h^2),
\end{aligned}$$

where the first equality follows by the law of iterated expectations, the second equality follows by Plancherel's isometry (Lemma 1 (1)), the third equality follows by a change of variables, and the last equality follows by the standard bias reduction argument using the twice continuous differentiability of f_{X^*,Z_d}, the symmetry of K, $\int K(w)w^2dw < \infty$, and the compactness of $\mathcal{I}$. For the second-order term, by Lemma 9, we have $E\left|\int_{x^*\in\mathcal{I}}\mathbb{K}_h(x^* - X)dx^*K_h(z_d - Z_d)\right|^2 = O(h^{-2\beta})$, and it follows

$$\mathbb{A}^d_n - a^d = O(h) + O_p(n^{-1/2}h^{-\beta}),$$

which implies (B.24) holds true if $h \to 0$ and $nh^{2\beta} \to \infty$ as $n \to \infty$.

Combining (B.19), (B.20), and (B.24), by a similar argument as in the proof of Theorem 3, we have

$$\frac{\tilde{m}_d(z_d) - m_d(z_d) - \mathrm{Bias}\{\tilde{m}_d(z_d)\}}{\sqrt{Var[\tilde{m}_d(z_d)]}} \xrightarrow{d} N(0,1),$$

where $\mathrm{Bias}\{\tilde{m}_d(z_d)\} = E[G^d_{n,1}]$.

The proof for the supersmooth case follows a similar route as the ordinary smooth case so we only state the differences as follows. First, by Lemmas 8 and 10, for $\eta \geq 0$, we have

$$E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^*K_h(z_d-Z_d)\{m_d(Z_d)+g(X^*)+U-m_d(z_d)\}\right|^{2+\eta}$$
$$= O\left(h^{-\eta}\left(\int \frac{|K^{\text{ft}}(th)|}{|f^{\text{ft}}_\epsilon(t)|}dt\right)^\eta E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^*K_h(z_d-Z_d)\right|^2\right) = O\left(h^{-(2\eta+3)}e^{(\eta+2)\mu h^{-\gamma}}\right).$$

By a similar argument, we have

$$E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)g(x^*)dx^*K_h(z_d-Z_d)\right|^{2+\eta} = O\left(h^{-(2\eta+3)}e^{(\eta+2)\mu h^{-\gamma}}\right).$$

Thus, by Assumption 11, (B.19) and (B.23) hold true.

Also, by Lemma 10, we have

$$\mathbb{A}^d_n - a^d = O(h) + O_p(n^{-1/2}h^{-3/2}e^{\mu h^{-\gamma}}),$$

which implies $\mathbb{A}^d_n - a^d = o_p(1)$ if $h\to 0$ and $nh^3e^{-2\mu h^{-\gamma}}\to\infty$, and the conclusion follows.

Appendix C. Proofs for Section 4

C.1. **Proof of Theorem 5.** Let $\check{M}_\kappa = \Re\check{E}[P_\kappa P_\kappa']$ and $\check{C}_\kappa = \Re\check{E}[YP_\kappa']$. The proof follows from a similar route as that of Theorem 2 except that instead of the orders of $\|\hat{E}[P_\kappa P_\kappa'] - M_\kappa\|^2$ and $\|\hat{E}[P_\kappa Y]-C_\kappa\|^2$, we need to quantify the orders of $\|\check{E}[P_\kappa P_\kappa']-M_\kappa\|^2$ and $\|\check{E}[P_\kappa Y]-C_\kappa\|^2$, which are given by

$$\|\check{E}[P_\kappa P_\kappa'] - M_\kappa\|^2 \leq 2\|\hat{E}[P_\kappa P_\kappa'] - M_\kappa\|^2 + 2\|\check{E}[P_\kappa P_\kappa'] - \hat{E}[P_\kappa P_\kappa']\|^2,$$
$$\|\check{E}[P_\kappa Y] - C_\kappa\|^2 \leq 2\|\hat{E}[P_\kappa Y] - C_\kappa\|^2 + 2\|\check{E}[P_\kappa Y] - \hat{E}[P_\kappa Y]\|^2,$$

and Lemma 4 and 12. Then, the conclusion follows from Assumptions 3 and 14, or Assumptions 4 and 15, which together with Lemma 4 and 12 implies $\|\check{M}_\kappa - M_\kappa\|^2 \leq \|\check{E}[P_\kappa P_\kappa'] - M_\kappa\|^2 \xrightarrow{p} 0$.

Appendix D. Lemmas

For $\zeta > 0$, let $G_{\epsilon,n,\zeta} = \{t\in\mathbb{R} : |f^{\text{ft}}_\epsilon(t)| < n^{-\zeta}\}$ be the region over which the ridge regularization is implemented, and $G^c_{\epsilon,n,\zeta} = \mathbb{R}\backslash G_{\epsilon,n,\zeta}$ be the complement of $G_{\epsilon,n,\zeta}$. First, we introduce Lemmas 1-3 to prepare for the proof of Lemma 4, which is used in the proof of Theorem 2.

Lemma 1. *For $f_1, f_2, f \in L_1(\mathbb{R})\cap L_2(\mathbb{R})$ and $c\in\mathbb{R}$, we have*

(1): $\langle f_1, f_2\rangle = \frac{1}{2\pi}\langle f^{\text{ft}}_1, f^{\text{ft}}_2\rangle$
(2): $\{\int f_1(w-\tilde{w})f_2(\tilde{w})d\tilde{w}\}^{\text{ft}}(t) = f^{\text{ft}}_1(t)f^{\text{ft}}_2(t)$
(3): $\{f_1f_2\}^{\text{ft}}(t) = \frac{1}{2\pi}\int f^{\text{ft}}_1(t-s)f^{\text{ft}}_2(s)ds$
(4): $f^{\text{ft}}(t-s) = \{f(w)e^{-isw}\}^{\text{ft}}(t)$
(5): $f^{\text{ft}}(ct) = \{f(w/c)/c\}^{\text{ft}}(t)$

Proof. Lemma 1 (1) is known as Plancherel's isometry and its proof can be found in Meister (2009, Theorem A.4). One useful special case is when $f_1 = f_2 = f$, which gives Parseval's identity,

$\|f\|_2^2 = \frac{1}{2\pi}\|f^{\mathrm{ft}}\|_2^2$. Lemma 1 (2) is known as the convolution theorem and its proof can be found in Meister (2009, Lemma A.1 (b)). Lemma 1 (3) follows from (2) and $f(w) = \frac{1}{2\pi}\{f^{\mathrm{ft}}\}^{\mathrm{ft}}(-w)$. Lemma 1 (4) follows from the definition of the Fourier transform. Lemma 1 (5) is known as the linear stretching property of the Fourier transform, and its proof can be found in Meister (2009, Lemma A.1 (e)). □

Lemma 2. *Suppose Assumptions 1 and 2 hold true.*

(1): *If f_ϵ is ordinary smooth of order $\beta > 0$, then*

$$\int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right), \qquad \sup_{z_d\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right),$$

$$\sup_{z_d,z_{d'}\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)|^2 dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right).$$

(2): *If f_ϵ is supersmooth of order $\gamma > 0$, then*

$$\int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2 dt = O\left((\log n)^{-\frac{2\alpha}{\gamma}}\right), \qquad \sup_{z_d\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t)|^2 dt = O\left((\log n)^{-\frac{2\alpha}{\gamma}}\right),$$

$$\sup_{z_d,z_{d'}\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}(t)|^2 dt = O\left((\log n)^{-\frac{2\alpha}{\gamma}}\right).$$

Proof. Since the proof is similar, we focus on $\int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2 dt$. For (1), if f_ϵ is ordinary smooth of order β, we have $c_{\mathrm{os},0}(1+|t|)^{-\beta} < n^{-\zeta}$ for $t \in G_{\epsilon,n,\zeta}$. Note that Jensen's inequality $(1+|t|) \leq \sqrt{2}(1+t^2)^{1/2}$ implies $(1+t^2)^{-\alpha} \leq 2^\alpha(1+|t|)^{-2\alpha}$, and it follows $(1+t^2)^{-\alpha} < 2^\alpha c_{\mathrm{os},0}^{-\frac{2\alpha}{\beta}} n^{-\frac{2\alpha\zeta}{\beta}}$. Also note that $\int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha dt \leq \int |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha dt < c_{\mathrm{sob}}$ by $f_{X^*} \in \mathcal{F}_{\alpha,c_{\mathrm{sob}}}$. Then we have

$$\begin{aligned}\int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2 dt &= \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha(1+t^2)^{-\alpha} dt \\ &\leq 2^\alpha c_{\mathrm{os},0}^{-\frac{2\alpha}{\beta}} n^{-\frac{2\alpha\zeta}{\beta}} \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha dt = O(n^{-\frac{2\alpha\zeta}{\beta}}). \qquad \text{(D.1)}\end{aligned}$$

For (2), if f_ϵ is supersmooth of order γ, we have $c_{\mathrm{ss},0}\exp(-\mu|t|^\gamma) < n^{-\zeta}$ for $t \in G_{\epsilon,n,\zeta}$. Note that

$$\begin{aligned} c_{\mathrm{ss},0}\exp(-\mu|t|^\gamma) < n^{-\zeta} \quad &\Rightarrow \quad |t|^\gamma > \mu^{-1}\big[\log(c_{\mathrm{ss},0}) + \zeta\log(n)\big] \\ &\Rightarrow \quad 1+|t|^2 > 1+\mu^{-\frac{2}{\gamma}}\big[\log(c_{\mathrm{ss},0}) + \zeta\log(n)\big]^{\frac{2}{\gamma}} \\ &\Rightarrow \quad (1+|t|^2)^{-\alpha} < \big(1+\mu^{-\frac{2}{\gamma}}\big[\log(c_{\mathrm{ss},0}) + \zeta\log(n)\big]^{\frac{2}{\gamma}}\big)^{-\alpha}, \end{aligned}$$

which implies there exists a constant $C > 0$ such that $(1+t^2)^{-\alpha} \leq C(\log n)^{-\frac{2\alpha}{\gamma}}$ for $t \in G_{\epsilon,n,\zeta}$. Then, similarly to the previous ordinary smooth case, we have

$$\begin{aligned}\int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2 dt &= \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha(1+t^2)^{-\alpha} dt \\ &\leq C(\log n)^{-\frac{2\alpha}{\gamma}} \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*}(t)|^2(1+t^2)^\alpha dt = O((\log n)^{-\frac{2\alpha}{\gamma}}). \qquad \text{(D.2)}\end{aligned}$$

□

Lemma 3. *Suppose Assumptions 1 and 2 hold true.*

(1): *If f_ϵ is ordinary smooth of order β with $\beta > 1/2$, then*

$$\int \frac{|f_\epsilon^{\rm ft}(t)|^{2r+2}}{\{|f_\epsilon^{\rm ft}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = O(n^{\frac{\zeta(2\beta+1)}{\beta}}).$$

(2): *If f_ϵ is supersmooth of order $\gamma > 0$, then*

$$\int \frac{|f_\epsilon^{\rm ft}(t)|^{2r+2}}{\{|f_\epsilon^{\rm ft}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = O(n^{2\zeta(r+2)}).$$

Proof. For (1), decompose

$$\int \frac{|f_\epsilon^{\rm ft}(t)|^{2r+2}}{\{|f_\epsilon^{\rm ft}(t)| \vee n^{-\zeta}\}^{2r+4}} dt = n^{2\zeta(r+2)} \int_{G_{\epsilon,n,\zeta}} |f_\epsilon^{\rm ft}(t)|^{2r+2} dt + \int_{G_{\epsilon,n,\zeta}^c} \frac{1}{|f_\epsilon^{\rm ft}(t)|^2} dt. \tag{D.3}$$

First, note that $|f_\epsilon^{\rm ft}(t)| \le c_{{\rm os},1}(1+|t|)^{-\beta}$ for $t \in \mathbb{R}$ and $c_{{\rm os},0}(1+|t|)^{-\beta} < n^{-\zeta}$ for $t \in G_{\epsilon,n,\zeta}$, which implies $(1+|t|)^{-2\beta(r+1)+1+\eta} < c_{{\rm os},0}^{-\frac{2\beta(r+1)-1-\eta}{\beta}} n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}}$ for $t \in G_{\epsilon,n,\zeta}$ and for any constant $\eta \in (0, 2\beta(r+1)-1)$. Also note that $\int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-1-\eta} dt \to 0$ as $n \to \infty$ because $\int (1+|t|)^{-1-\eta} dt < \infty$ for any $\eta > 0$ and $|t| > c_{{\rm os},0}^{\frac{1}{\beta}} n^{\frac{\zeta}{\beta}} - 1$ for $t \in G_{\epsilon,n,\zeta}$. Then, for the first term of (D.3), we have

$$\begin{aligned} &\int_{G_{\epsilon,n,\zeta}} |f_\epsilon^{\rm ft}(t)|^{2r+2} dt \le c_{{\rm os},1}^{2r+2} \int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-2\beta(r+1)+1+\eta}(1+|t|)^{-1-\eta} dt \\ \le\ & c_{{\rm os},1}^{2r+2} c_{{\rm os},0}^{-\frac{2\beta(r+1)-1-\eta}{\beta}} n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}} \int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-1-\eta} dt = o(n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}}). \end{aligned} \tag{D.4}$$

For the second term of (D.3), note that $|f_\epsilon^{\rm ft}(t)|^{-2} \le n^{2\zeta}$ and $|t| < c_{{\rm os},1}^{\frac{1}{\beta}} n^{\frac{\zeta}{\beta}}$ for $t \in G_{\epsilon,n,\zeta}^c$, which implies

$$\int_{G_{\epsilon,n,\zeta}^c} |f_\epsilon^{\rm ft}(t)|^{-2} dt \le n^{2\zeta} \int_{G_{\epsilon,n,\zeta}^c} dt \le 2c_{{\rm os},1}^{\frac{1}{\beta}} n^{\frac{\zeta(2\beta+1)}{\beta}} = O(n^{\frac{\zeta(2\beta+1)}{\beta}}). \tag{D.5}$$

Then, (1) follows from combining (D.3), (D.4), and (D.5).

For (2), note that $|f_\epsilon^{\rm ft}(t)| \vee n^{-\zeta} \ge n^{-\zeta}$ and $|f_\epsilon^{\rm ft}(t)| \le c_{{\rm ss},1}\exp(-\mu|t|^\gamma)$, which implies

$$\int \frac{|f_\epsilon^{\rm ft}(t)|^{2r+2}}{\{|f_\epsilon^{\rm ft}(t)| \vee n^{-\zeta}\}^{2r+4}} dt \le 2c_{{\rm ss},1}^{2r+2} n^{2\zeta(r+2)} \int_0^{+\infty} \exp(-(2r+2)\mu|t|^\gamma) dt. \tag{D.6}$$

Also note that t^2 and $\exp((2r+2)\mu|t|^\gamma)$ are strictly increasing and there exists a constant δ such that $\exp((2r+2)\mu|t|^\gamma) > t^2$ for any $t > \delta$. Then, since $\exp(-(2r+2)\mu|t|^\gamma) \le 1$ for $t \ge 0$, we have

$$\begin{aligned} \int_0^{+\infty} \exp(-(2r+2)\mu|t|^\gamma) dt &= \{\int_0^\delta + \int_\delta^{+\infty}\} \exp(-(2r+2)\mu|t|^\gamma) dt \\ &\le \delta + \int_\delta^{+\infty} t^{-2} dt = \delta + \delta^{-1} < \infty. \end{aligned} \tag{D.7}$$

Then, (2) follows from combining (D.6) and (D.7). □

Let $\mathcal{I}_{M_\kappa} = \{(p, Q) : E[p(X^*)Q]$ is an element of $M_\kappa\}$ be the index set characterizing the components of M_κ, where p is a product of $\{p_0, p_1, \ldots, p_\kappa\}$ and Q is a product of $\{1, q_1(Z_1), \ldots, q_\kappa(Z_D)\}$.

To keep notation simple throughout the proof of Lemma 4, let ϱ_n^B denote $n^{-\frac{2\alpha\zeta}{\beta}}$ under Assumption 3 and $(\log n)^{-\frac{2\alpha}{\gamma}}$ under Assumption 4, and ϱ_n^V denote $n^{2\zeta+\frac{\zeta}{\beta}-1}$ under Assumption 3 and $n^{2\zeta(r+2)-1}$ under Assumption 4.

Lemma 4. *Suppose Assumptions 1 and 2 hold true.*

(1): *Under Assumption 3, it holds*

$$|\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2 = O_p\left(\kappa^2 n^{2\zeta+\frac{\zeta}{\beta}-1} + \kappa n^{-\frac{2\alpha\zeta}{\beta}}\right), \qquad |\hat{E}[P_\kappa Y] - C_\kappa|^2 = O_p\left(\kappa n^{2\zeta+\frac{\zeta}{\beta}-1} + n^{-\frac{2\alpha\zeta}{\beta}}\right).$$

(2): *Under Assumption 4 with $r \geq 0$ and $0 < \zeta < \frac{1}{2(r+2)}$, it holds*

$$|\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2 = O_p\left(\kappa(\log n)^{-\frac{2\alpha}{\gamma}}\right), \qquad |\hat{E}[P_\kappa Y] - C_\kappa|^2 = O_p\left((\log n)^{-\frac{2\alpha}{\gamma}}\right).$$

Proof. Since the proof is similar, we focus on the proof for $|\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2$. Let $B_{p,Q} = E\{\hat{E}[p(X^*)Q]\} - E[p(X^*)Q]$ be the bias of the proposed estimator of the element of M_κ characterized by p and Q. Let $V_{p,Q} = \hat{E}[p(X^*)Q] - E\{\hat{E}[p(X^*)Q]\}$, and $V_{p,Q,j}$ be its component associated with the j-th observation, i.e., $V_{p,Q} = \frac{1}{n}\sum_{j=1}^n V_{p,Q,j}$. First, note that random sampling implies

$$\begin{aligned} E|\hat{E}[P_\kappa P_\kappa'] - M_\kappa|^2 &= \frac{1}{n^2}\sum_{j,j'=1}^n \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} E\left[(B_{p,Q}+V_{p,Q,j})\overline{(B_{p,Q}+V_{p,Q,j'})}\right] \\ &= \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} |B_{p,Q}|^2 + \frac{1}{n}\sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} E|V_{p,Q,1}|^2 \equiv B + V. \end{aligned}$$

Using Lemma 1 (1), the law of iterated expectations, and properties of classical measurement error, we have

$$\begin{aligned} E[p(X^*)Q] &= \langle E[Q|X^*]f_{X^*}, p\rangle = \frac{1}{2\pi}\int E[Qe^{\mathrm{i}tX^*}]p^{\mathrm{ft}}(-t)dt, \\ E\{\hat{E}[p(X^*)Q]\} &= \frac{1}{2\pi}\int E[Qe^{\mathrm{i}tX^*}]\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}p^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}}dt. \end{aligned}$$

So, the bias term B can be written as

$$B = \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} \left|\frac{1}{2\pi}\int\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1\right)E[Qe^{\mathrm{i}tX^*}]p^{\mathrm{ft}}(-t)dt\right|^2 \equiv B_1 + \cdots + B_7,$$

where $B_1, \ldots, B_7$ are summations of the terms whose (p,Q) are in the form of $(p_0, 1)$, $(p_k, 1)$, $(p_k p_l, 1)$, $(p_0, q_k(Z_d))$, $(p_k, q_l(Z_d))$, $(p_0, q_k(Z_d)q_l(Z_d))$, and $(p_0, q_k(Z_d)q_l(Z_{d'}))$ separately for $k, l = 1, \ldots, \kappa$ and $d, d' = 1, \ldots, D$ with $d \neq d'$.

Since the proof is similar for B_1, B_2, and B_3, we focus on B_3, for which we have

$$\begin{aligned}
B_3 &= \sum_{k,l=1}^{\kappa} \left| \frac{1}{2\pi} \int \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(t) \{p_k p_l\}^{\mathrm{ft}}(-t) dt \right|^2 \\
&= \sum_{k,l=1}^{\kappa} \left| \frac{1}{4\pi^2} \iint \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(t) p_k^{\mathrm{ft}}(-t-s) p_l^{\mathrm{ft}}(s) ds dt \right|^2 \\
&= \sum_{k,l=1}^{\kappa} \left| \frac{1}{4\pi^2} \iint \left(\frac{|f_\epsilon^{\mathrm{ft}}(u-v)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(u-v)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(u-v) p_k^{\mathrm{ft}}(-u) p_l^{\mathrm{ft}}(v) du dv \right|^2 \\
&\leq \frac{1}{16\pi^4} \int \left\{ \sum_{k=1}^{\kappa} \left| \left\langle \left(\frac{|f_\epsilon^{\mathrm{ft}}(u-v)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(u-v)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(u-v), p_k^{\mathrm{ft}}(u) \right\rangle_u \right|^2 \right\} \sum_{l=1}^{\kappa} |p_l^{\mathrm{ft}}(v)|^2 dv \\
&\leq \frac{\kappa}{4\pi^2} \int_{G_{\epsilon,n,\zeta}} |f_{X^*}^{\mathrm{ft}}(t)|^2 dt = O(\kappa \varrho_n^B),
\end{aligned}$$

where the second step follows from Lemma 1 (3), the third step uses the change of variables $(u, v) = (t + s, s)$, the last step follows from Lemma 2, and the penultimate step follows from Lemma 1 (1), the orthonormality of $\{p_l\}_{l=1}^{\kappa}$ and

$$\begin{aligned}
&\sum_{k=1}^{\kappa} \left| \left\langle \left(\frac{|f_\epsilon^{\mathrm{ft}}(u-v)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(u-v)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(u-v), p_k^{\mathrm{ft}}(u) \right\rangle_u \right|^2 \\
=\;& 4\pi^2 \sum_{k=1}^{\kappa} \left| \left\langle h_1(w) e^{-\mathrm{i}vw}, p_k(w) \right\rangle_w \right|^2 \leq 4\pi^2 \left\| h_1(w) e^{-\mathrm{i}vw} \right\|_2^2 \\
\leq\;& 2\pi \left\| \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(t) \right\|_2^2 \leq 2\pi \int_{G_{\epsilon,n,\zeta}} |f_{X^*}^{\mathrm{ft}}(t)|^2 dt,
\end{aligned}$$

where h_1 denotes the Fourier inverse of $\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) f_{X^*}^{\mathrm{ft}}(t)$, the first step follows from Lemma 1 (1) and (4), the second step follows from the orthonormality of $\{p_k\}_{k=1}^{\kappa}$, and the third step follows from $|e^{-\mathrm{i}vw}| = 1$ and Lemma 1 (1). Similarly, we have $B_1, B_2 = O(\varrho_n^B)$.

Since the proof is similar for B_4 and B_5, we focus on B_5, for which we have

$$\begin{aligned}
B_5 &= 2 \sum_{d=1}^{D} \sum_{k,l=1}^{\kappa} \left| \frac{1}{2\pi} \int \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[q_l(Z_d) e^{\mathrm{i}tX^*}] p_k^{\mathrm{ft}}(-t) dt \right|^2 \\
&= \frac{1}{2\pi^2} \sum_{d=1}^{D} \sum_{k,l=1}^{\kappa} \left| \left\langle \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[q_l(Z_d) e^{\mathrm{i}tX^*}], p_k^{\mathrm{ft}}(t) \right\rangle_t \right|^2 \\
&\leq \frac{1}{\pi} \sum_{d=1}^{D} \int_{G_{\epsilon,n,\zeta}} \left\{ \sum_{l=1}^{\kappa} \left| \int f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t) f_{Z_d}(z_d) q_l(z_d) dz_d \right|^2 \right\} dt \\
&\leq \frac{1}{\pi} \sum_{d=1}^{D} \int_{G_{\epsilon,n,\zeta}} \left\{ \int |f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2 |f_{Z_d}(z_d)|^2 dz_d \right\} dt \\
&\leq \frac{1}{\pi} \left\{ \begin{array}{c} \max_{d \in \{1,\cdots,D\}} \sup_{z_d \in [-1,1]} \int_{G_{\epsilon,n,\zeta}} |f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)|^2 dt \\ \times \sum_{d=1}^{D} \int |f_{Z_d}(z_d)|^2 dz_d \end{array} \right\} = O(\varrho_n^B),
\end{aligned}$$

where the fourth step follows from the orthonormality of $\{q_l\}_{l=1}^{\kappa}$, the last step follows from Lemma 2 and the boundedness of f_{Z_d}, and the third step follows from $E[q_l(Z_d)e^{\mathrm{i}tX^*}] = \int f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t) f_{Z_d}(z_d) q_l(z_d) dz_d$ and

$$\begin{aligned}
&\sum_{k=1}^{\kappa} \left| \left\langle \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[q_l(Z_d)e^{\mathrm{i}tX^*}], p_k^{\mathrm{ft}}(t) \right\rangle_t \right|^2 = 4\pi^2 \sum_{k=1}^{\kappa} \left| \langle h_{2,l,d}, p_k \rangle \right|^2 \le 4\pi^2 \|h_{2,l,d}\|_2^2 \\
= \;& 2\pi \left\| \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[q_l(Z_d)e^{\mathrm{i}tX^*}] \right\|_2^2 \le 2\pi \int_{G_{\epsilon,n,\zeta}} |E[q_l(Z_d)e^{\mathrm{i}tX^*}]|^2 dt,
\end{aligned}$$

where $h_{2,l,d}$ denotes the Fourier inverse of $\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[q_l(Z_d)e^{\mathrm{i}tX^*}]$, the first step follows from Lemma 1 (1), the second step follows from the orthonormality of $\{p_k\}_{k=1}^{\kappa}$, and the third step follows from Lemma 1 (1). Similarly, we have $B_4 = O(\varrho_n^B)$.

Since the proof is similar for B_6 and B_7, we focus on B_6, for which we have

$$\begin{aligned}
B_6 \;=\;& \sum_{d=1}^{D} \sum_{k,l=1}^{\kappa} \left| \frac{1}{2\pi} \int \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[q_k(Z_d)q_l(Z_d)e^{\mathrm{i}tX^*}] p_0^{\mathrm{ft}}(-t) dt \right|^2 \\
\le\;& \frac{A(\mathcal{I})}{2\pi} \sum_{d=1}^{D} \sum_{k,l=1}^{\kappa} \int \left| \left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} - 1 \right) E[q_k(Z_d)q_l(Z_d)e^{\mathrm{i}tX^*}] \right|^2 \\
\le\;& \frac{A(\mathcal{I})}{2\pi} \sum_{d=1}^{D} \sum_{k=1}^{\kappa} \int_{G_{\epsilon,n,\zeta}} \sum_{l=1}^{\kappa} \left| \left\langle f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t) f_{Z_d}(z_d) q_k(z_d), q_l(z_d) \right\rangle_{z_d} \right|^2 dt \\
\le\;& \frac{A(\mathcal{I})}{2\pi} \sum_{d=1}^{D} \sum_{k=1}^{\kappa} \int_{G_{\epsilon,n,\zeta}} \left\{ \int |f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t) f_{Z_d}(z_d) q_k(z_d)|^2 dz_d \right\} dt \\
\le\;& \frac{A(\mathcal{I})}{2\pi} \left\{ \begin{array}{c} \max_{d\in\{1,\cdots,D\}} \sup_{z_d\in[-1,1]} \int_{G_{\epsilon,n,\zeta}} |f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t)|^2 dt \\ \times \sum_{d=1}^{D} \sum_{k=1}^{\kappa} \int |f_{Z_d}(z_d)|^2 |q_k(z_d)|^2 dz_d \end{array} \right\} = O(\kappa \varrho_n^B),
\end{aligned}$$

where the second step follows from the Cauchy-Schwarz inequality and Lemma 1 (1), the third step follows from $E[q_k(Z_d)q_l(Z_d)e^{\mathrm{i}tX^*}] = \int_{z_d} f^{\mathrm{ft}}_{X^*|Z_d=z_d}(t) f_{Z_d}(z_d) q_k(z_d) q_l(z_d) dz_d$, the fourth follows from the orthonormality of $\{q_l\}_{l=1}^{\kappa}$, and the last step follows from Lemma 2, the boundedness of f_{Z_d}, and the unity of q_k. Similarly, we have $B_7 = O(\varrho_n^B)$.

Combining these results, we obtain

$$B = O(\kappa \varrho_n^B) = \begin{cases} O(\kappa n^{-\frac{2\alpha\zeta}{\beta}}) & \text{under Assumption 3,} \\ O(\kappa (\log n)^{-\frac{2\alpha}{\gamma}}) & \text{under Assumption 4.} \end{cases}$$

We now consider the variance term V. Similarly to the bias term, we decompose

$$V \;\le\; \frac{1}{n} \sum_{(p,Q)\in\mathcal{I}_{M_\kappa}} E \left| \frac{1}{2\pi} \int Q e^{\mathrm{i}tX} \frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r p^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)| \vee n^{-\zeta}\}^{r+2}} \right|^2 \equiv V_1 + \cdots + V_7,$$

where $V_1, \ldots, V_7$ are summations of non-central second moment terms with (p, Q) in the forms of $(p_0, 1)$, $(p_k, 1)$, $(p_k p_l, 1)$, $(p_0, q_k(Z_d))$, $(p_k, q_l(Z_d))$, $(p_0, q_k(Z_d)q_l(Z_d))$, and $(p_0, q_k(Z_d)q_l(Z_{d'}))$ separately for $k, l = 1, \ldots, \kappa$ and $d, d' = 1, \ldots, D$ with $d \neq d'$.

Since the proof is similar for V_1, V_2, and V_3, we focus on V_3, for which we have

$$\begin{aligned}
V_3 &= \frac{1}{n}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r\{p_k p_l\}^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt\right|^2 \\
&= \frac{1}{4\pi^2 n}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\iint e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}p_k^{\mathrm{ft}}(-t-s)p_l^{\mathrm{ft}}(s)dsdt\right|^2 \\
&= \frac{1}{4\pi^2 n}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\iint e^{\mathrm{i}(u-v)X}\frac{f_\epsilon^{\mathrm{ft}}(-u+v)|f_\epsilon^{\mathrm{ft}}(u-v)|^r}{\{|f_\epsilon^{\mathrm{ft}}(u-v)|\vee n^{-\zeta}\}^{r+2}}p_k^{\mathrm{ft}}(-u)p_l^{\mathrm{ft}}(v)dudv\right|^2 \\
&\leq \frac{1}{16\pi^4 n}\iint\left\{\sum_{k=1}^{\kappa}\left|\left\langle e^{\mathrm{i}(u-v)x}\frac{f_\epsilon^{\mathrm{ft}}(-u+v)|f_\epsilon^{\mathrm{ft}}(u-v)|^r}{\{|f_\epsilon^{\mathrm{ft}}(u-v)|\vee n^{-\zeta}\}^{r+2}}, p_k^{\mathrm{ft}}(u)\right\rangle_u\right|^2\right\} f_X(x)dx\sum_{l=1}^{\kappa}|p_l^{\mathrm{ft}}(v)|^2 dv \\
&\leq \frac{\kappa}{4\pi^2 n}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt = O(\kappa\varrho_n^V),
\end{aligned}$$

where the second step follows from Lemma 1 (3), the third step uses the change of variables $(u, v) = (t + s, s)$, the last step follows from Lemma 3, and the penultimate step follows from Lemma 1 (1), the unity of $\{p_l\}_{l=1}^{\kappa}$, and

$$\begin{aligned}
&\sum_{k=1}^{\kappa}\left|\left\langle e^{\mathrm{i}(u-v)x}\frac{f_\epsilon^{\mathrm{ft}}(-u+v)|f_\epsilon^{\mathrm{ft}}(u-v)|^r}{\{|f_\epsilon^{\mathrm{ft}}(u-v)|\vee n^{-\zeta}\}^{r+2}}, p_k^{\mathrm{ft}}(u)\right\rangle_u\right|^2 \\
&= 4\pi^2\sum_{k=1}^{\kappa}\left|\left\langle h_{3,x}(w)e^{-\mathrm{i}vw}, p_k(w)\right\rangle_w\right|^2 \leq 4\pi^2\|h_{3,x}(w)e^{-\mathrm{i}vw}\|_2^2 \\
&\leq \int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt,
\end{aligned}$$

where $h_{3,x}$ denotes the Fourier inversion of $e^{\mathrm{i}tx}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}$ for every x in the support of X, the first step follows from Lemma 1 (1) and (4), the second step follows from the orthonormality of $\{p_k\}_{k=1}^{\kappa}$, and the last step follows from $|e^{-ivw}| = |e^{itx}| = 1$, Lemma 1 (1) and $|f_\epsilon^{\mathrm{ft}}(-t)| = |f_\epsilon^{\mathrm{ft}}(t)|$. Similarly, we have $V_1, V_2 = O(\varrho_n^V)$.

Since the proof is similar for other terms, we focus on V_5, for which we have

$$\begin{aligned}
V_5 &= \frac{2}{n}\sum_{d=1}^{D}\sum_{k,l=1}^{\kappa} E\left|\frac{1}{2\pi}\int q_l(Z_d)e^{\mathrm{i}tX}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r p_k^{\mathrm{ft}}(-t)}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt\right|^2 \\
&= \frac{1}{2\pi^2 n}\sum_{d=1}^{D}\sum_{l=1}^{\kappa}\iint|q_l(z_d)|^2\sum_{k=1}^{\kappa}\left|\left\langle e^{\mathrm{i}tx}\frac{f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}, p_k^{\mathrm{ft}}(t)\right\rangle_t\right|^2 f_{X,Z_d}(x,z_d)dxdz_d \\
&\leq \frac{1}{2\pi^2 n}\int\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{2r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{2r+4}}dt\sum_{d=1}^{D}\sum_{l=1}^{\kappa}\int|q_l(z_d)|^2 f_{Z_d}(z_d)dz_d = O(\kappa\varrho_n^V),
\end{aligned}$$

where the last step follows from Lemma 3, the boundedness of f_{Z_d}, and the unity of $\{q_l\}_{l=1}^{\kappa}$. Similarly, we have $V_4 = O(\kappa\varrho_n^V)$ and $V_6, V_7 = O(\kappa^2\varrho_n^V)$.

Combining these results, we obtain

$$V = O(\kappa^2\varrho_n^V) = \begin{cases} O(\kappa^2 n^{2\zeta+\frac{\zeta}{\beta}-1}) & \text{under Assumption 3,} \\ O(\kappa^2 n^{2\zeta(r+2)-1}) & \text{under Assumption 4.}\end{cases}$$

Under Assumption 4, κ can only diverge at a logarithmic rate so that $\kappa(\log n)^{-\frac{2\alpha}{\gamma}}$ converges to zero. Therefore, for $0 < \zeta < \frac{1}{2(r+2)}$ and n large enough, we have $\kappa^2 n^{2\zeta(r+2)-1} \ll \kappa(\log n)^{-\frac{2\alpha}{\gamma}}$, and the conclusion follows. $\square$

Lemma 5. *Under Assumptions 3 and 6, there exists* $\psi \in L_1(\mathbb{R})$ *such that*

$$\sup_n h^\beta \frac{|K^{\rm ft}(s)|}{|f_\epsilon^{\rm ft}(s/h)|} \leq \psi(s),$$

which implies that there exists a constant $c > 0$ *such that* $h^{\beta+1}\int \frac{|K^{\rm ft}(th)|}{|f_\epsilon^{\rm ft}(t)|}dt \leq c$.

Proof. Since $\lim_{|t|\to\infty} |t|^\beta |f_\epsilon^{\rm ft}(t)| = c_\epsilon$, there exists a constant c_F such that $|t|^\beta |f_\epsilon^{\rm ft}(t)| > c_\epsilon/2$ for all $t \geq c_F$. Then for constants $c_1, c_2 > 0$ such that $c_1 > h^\beta$ and $c_2 > c_F h$ for all n, we have

$$\begin{aligned} h^\beta \frac{|K^{\rm ft}(s)|}{|f_\epsilon^{\rm ft}(s/h)|} &\leq h^\beta \frac{\max_{|s|\leq c_F h} |K^{\rm ft}(s)|}{\min_{|s|\leq c_F} |f_\epsilon^{\rm ft}(s)|} \mathbb{1}\{|s| \leq c_F h\} + \frac{|K^{\rm ft}(s)||s|^\beta}{(|s|/h)^\beta |f_\epsilon^{\rm ft}(s/h)|} \mathbb{1}\{|s| > c_F h\} \\ &\leq c_1 c_{\rm os,0}^{-1} (1 + c_F)^\beta \|K^{\rm ft}\|_\infty \mathbb{1}\{|s| \leq c_2\} + \frac{2|K^{\rm ft}(s)||s|^\beta}{c_\epsilon} \equiv \psi(s), \qquad \text{(D.8)} \end{aligned}$$

where integrability of $\psi(s)$ follows by $\|K^{\rm ft}\|_\infty < \infty$, the ordinary smoothness of f_ϵ, and $\int |K^{\rm ft}(s)||s|^\beta ds < \infty$. The second statement immediately follows by the change of variables $t = s/h$. $\square$

The following lemma is an extension of Fan (1991a, Lemma 2.1) to the multivariate case.

Lemma 6. *Suppose* $K_n : \mathbb{R}^d \to \mathbb{C}$ *is a sequence of functions satisfying*

$$K_n(x) \to K(x) \text{ and } \sup_n |K_n(x)| \leq K^*(x),$$

where K^* *satisfies* $\int |K^*(x)|dx < \infty$. *If* f *is bounded and* c *is a continuity point of* f, *then for any sequence* $h \to 0$ *as* $n \to \infty$,

$$\int h^{-d} K_n(h^{-1}(c - x)) f(x) dx = f(c) \int K(x) dx + o(1).$$

Proof. Note that

$$\begin{aligned} &\left| \int h^{-d} K_n(h^{-1}(c - x)) f(x) dx - f(c) \int K(x) dx \right| \\ \leq &\left| \int K_n(z) \big[f(c - zh) - f(c) \big] dz \right| + |f(c)| \left| \int \big[K_n(z) - K(z) \big] dz \right|, \end{aligned}$$

where the inequality follows by the change of variables $z = \frac{c-x}{h}$. The second term converges to zero, which follows by $K_n \to K$, $\sup_n |K_n| \leq K^*$, $\int |K^*(x)|dx < \infty$, and the dominated convergence theorem. For the first term,

$$\left| \int K_n(z)\{f(c - zh) - f(c)\} dz \right| \leq \sup_{\|z\|\leq\delta} |f(c - z) - f(c)| \int |K^*(z)| dz + (\|f\|_\infty + |f(c)|) \int_{\|z\|>\delta/h} |K^*(z)| dz,$$

where $\delta \to 0$ and $\delta/h \to \infty$ as $n \to \infty$. The first term on the right-hand side converges to zero because f is continuous at c and $\int |K^*(x)|dx < \infty$, and the second term also converges to zero because f is bounded and $\int |K^*(x)|dx < \infty$. $\square$

Lemma 7. *Suppose f is continuous at x^*, f_ϵ is ordinary smooth of order β, $\|f_\epsilon^{\mathrm{ft}'}\|_\infty < \infty$, $|s|^\beta \big|f_\epsilon^{\mathrm{ft}}(s)\big| \to c_\epsilon$, $|s|^{\beta+1}\big|f_\epsilon^{\mathrm{ft}'}(s)\big| \to \beta c_\epsilon$, $\|K^{\mathrm{ft}}\|_\infty < \infty$, $\|K^{\mathrm{ft}'}\|_\infty < \infty$, $\int |s|^\beta |K^{\mathrm{ft}}(s)|ds < \infty$, and $\int |s|^\beta |K^{\mathrm{ft}'}(s)|ds < \infty$. Then*

$$\lim_{n\to\infty} h^{2\beta+1} \int_x \frac{1}{4\pi^2} \left| \int_t \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} e^{-\mathrm{i}t(x^*-x)} dt \right|^2 f(x)dx = \frac{f(x^*)}{2\pi c_\epsilon^2} \int |s|^{2\beta} |K^{\mathrm{ft}}(s)|^2 ds.$$

Proof. First, observe that

$$\begin{aligned} & \lim_{n\to\infty} \frac{h^\beta}{2\pi} \int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx} ds = \lim_{n\to\infty} \frac{1}{2\pi} \int \frac{K^{\mathrm{ft}}(s)|s|^\beta}{(|s|/h)^\beta f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx} ds \\ = & \frac{1}{2\pi} \int \left\{ \lim_{n\to\infty} \frac{K^{\mathrm{ft}}(s)|s|^\beta}{(|s|/h)^\beta f_\epsilon^{\mathrm{ft}}(s/h)} \mathbb{1}\{|s| > c_F h\} \right\} e^{-\mathrm{i}sx} ds = \frac{1}{2\pi c_\epsilon} \int K^{\mathrm{ft}}(s)|s|^\beta e^{-\mathrm{i}sx} ds, \end{aligned}$$

where the second and last equalities follow by Lemma 5 and the dominated convergence theorem. Then it follows

$$\frac{h^{2\beta}}{4\pi^2} \left| \int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx} ds \right|^2 \to \frac{1}{4\pi^2 c_\epsilon^2} \left| \int K^{\mathrm{ft}}(s)|s|^\beta e^{-\mathrm{i}sx} ds \right|^2. \tag{D.9}$$

Moreover, using integration by parts, we have

$$\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx} ds = \frac{1}{\mathrm{i}x} \int \frac{K^{\mathrm{ft}'}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx} ds + \frac{1}{\mathrm{i}xh} \int \frac{K^{\mathrm{ft}}(s) f_\epsilon^{\mathrm{ft}'}(s/h)}{f_\epsilon^{\mathrm{ft}^2}(s/h)} e^{-\mathrm{i}sx} ds. \tag{D.10}$$

Since $|s|^\beta |f_\epsilon^{\mathrm{ft}}(s)| \to c_\epsilon$ and $|s|^{\beta+1}|f_\epsilon^{\mathrm{ft}'}(s)| \to \beta c_\epsilon$ as $s \to \infty$, there exists a constant $c_F > 0$ such that $|s|^\beta |f_\epsilon^{\mathrm{ft}}(s)| > c_\epsilon/2$ and $|s|^{\beta+1}|f_\epsilon^{\mathrm{ft}'}(s)| < 5\beta c_\epsilon/4$ for any s satisfying $|s| > c_F$. Then we have

$$\begin{aligned} & \left| \frac{1}{\mathrm{i}x} \int \frac{K^{\mathrm{ft}'}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx} ds \right| \le \frac{1}{|x|} \int \frac{|K^{\mathrm{ft}'}(s)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|} ds \\ \le & \frac{h}{|x|} \left(\frac{2c_F \max_{|s|\le c_F h} |K^{\mathrm{ft}'}(s)|}{\min_{|s|\le c_F} |f_\epsilon^{\mathrm{ft}}(s)|} \right) + \frac{h^{-\beta}}{|x|} \int_{|s|>c_F h} \frac{|K^{\mathrm{ft}'}(s)||s|^\beta}{(|s|/h)^\beta |f_\epsilon^{\mathrm{ft}}(s/h)|} ds \\ \le & \frac{h}{|x|} 2c_F c_{\mathrm{os},0}^{-1} (1+c_F)^\beta \|K^{\mathrm{ft}'}\|_\infty + \frac{h^{-\beta}}{|x|} \left(\frac{2}{c_\epsilon} \right) \int |K^{\mathrm{ft}'}(s)||s|^\beta ds = O(h^{-\beta}|x|^{-1}), \end{aligned} \tag{D.11}$$

and

$$\begin{aligned} & \left| \frac{1}{\mathrm{i}xh} \int \frac{K^{\mathrm{ft}}(s) f_\epsilon^{\mathrm{ft}'}(s/h)}{f_\epsilon^{\mathrm{ft}^2}(s/h)} ds \right| \le \frac{h^{-1}}{|x|} \int \frac{|K^{\mathrm{ft}}(s)||f_\epsilon^{\mathrm{ft}'}(s/h)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|^2} ds \\ \le & \frac{1}{|x|} \left(\frac{2c_F \max_{|s|\le c_F h} |K^{\mathrm{ft}}(s)| \max_{|s|\le c_F} |f_\epsilon^{\mathrm{ft}'}(s)|}{\min_{|s|\le c_F} |f_\epsilon^{\mathrm{ft}}(s)|^2} \right) + \frac{h^{-\beta}}{|x|} \int_{|s|>c_F h} \frac{|K^{\mathrm{ft}}(s)||s|^{\beta-1} (|s|/h)^{\beta+1} |f_\epsilon^{\mathrm{ft}'}(s/h)|}{(|s|/h)^{2\beta} |f_\epsilon^{\mathrm{ft}}(s/h)|^2} ds \\ \le & \frac{h}{|x|} 2c_F c_{\mathrm{os},0}^{-2} (1+c_F)^{2\beta} \|K^{\mathrm{ft}}\|_\infty \|f_\epsilon^{\mathrm{ft}'}\|_\infty + \frac{h^{-\beta}}{|x|} \left(\frac{5\beta}{c_\epsilon} \right) \int |K^{\mathrm{ft}}(s)||s|^{\beta-1} ds = O(h^{-\beta}|x|^{-1}). \end{aligned} \tag{D.12}$$

Thus, Lemma 5, (D.10), (D.11), and (D.12) imply there are a pair of constants $c_1, c_2 > 0$ such that

$$\sup_n h^{2\beta} \left| \int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)} e^{-\mathrm{i}sx} ds \right|^2 \le \min\{c_1, c_2|x|^{-2}\}. \tag{D.13}$$

Therefore, the conclusion follows by

$$\begin{aligned}
& \lim_{n\to\infty} h^{2\beta+1} \int \frac{1}{4\pi^2} \left| \int_t \frac{K^{\rm ft}(th)}{f_\epsilon^{\rm ft}(t)} e^{-{\rm i}t(x^*-x)} dt \right|^2 f(x)dx \\
= & \lim_{n\to\infty} \int_x \frac{h^{2\beta-1}}{4\pi^2} \left| \int_s \frac{K^{\rm ft}(s)}{f_\epsilon^{\rm ft}(s/h)} e^{-\frac{{\rm i}s(x^*-x)}{h}} ds \right|^2 f(x)dx \\
= & \frac{f(x^*)}{c_\epsilon^2} \int_x \left| \frac{1}{2\pi} \int_s K^{\rm ft}(s)|s|^\beta e^{-{\rm i}sx} ds \right|^2 dx = \frac{f(x^*)}{2\pi c_\epsilon^2} \int |K^{\rm ft}(s)|^2 |s|^{2\beta} ds, \qquad \text{(D.14)}
\end{aligned}$$

where the first equality follows by the change of variables $s = th$, the second equality follows by Lemma 6 with $K_n(x) = \frac{h^{2\beta}}{4\pi^2} \left| \int \frac{K^{\rm ft}(s)}{f_\epsilon^{\rm ft}(s/h)} e^{-{\rm i}sx} ds \right|^2$ and $K^*(x) = \min\{c_1, c_2|x|^{-2}\}$, and the third equality follows by Lemma 1 (1). □

Lemma 8. *Suppose Assumptions 4 and 8 hold true. Then there exists a constant $c > 0$ such that*

$$h e^{-\mu h^{-\gamma}} \int \frac{|K^{\rm ft}(th)|}{|f_\epsilon^{\rm ft}(t)|} dt \le c, \qquad h e^{-2\mu h^{-\gamma}} \int_x \left| \int_t e^{-{\rm i}t(x^*-x)} \frac{K^{\rm ft}(th)}{f_\epsilon^{\rm ft}(t)} dt \right|^2 dx \le c.$$

Proof. The first statement follows by

$$\int \frac{|K^{\rm ft}(th)|}{|f_\epsilon^{\rm ft}(t)|} dt = h^{-1} \int \frac{|K^{\rm ft}(s)|}{|f_\epsilon^{\rm ft}(s/h)|} ds \le c_{{\rm ss},0}^{-1} h^{-1} \int_{|s|\le 1} |K^{\rm ft}(s)| e^{\mu(|s|/h)^\gamma} ds = O(h^{-1} e^{\mu h^{-\gamma}}),$$

where the first equality follows by the change of variables $s = th$, the inequality follows by the supersmoothness of f_ϵ and the fact that $K^{\rm ft}$ is supported on $[-1, 1]$, and the last equality uses $\|K^{\rm ft}\|_\infty < \infty$.

The second statement follows by

$$\begin{aligned}
\int_x \left| \int_t e^{-{\rm i}t(x^*-x)} \frac{K^{\rm ft}(th)}{f_\epsilon^{\rm ft}(t)} dt \right|^2 dx &= 2\pi \int \frac{|K^{\rm ft}(th)|^2}{|f_\epsilon^{\rm ft}(t)|^2} dt = 2\pi h^{-1} \int \frac{|K^{\rm ft}(s)|^2}{|f_\epsilon^{\rm ft}(s/h)|^2} ds \\
&\le 2\pi c_{{\rm ss},0}^{-2} h^{-1} \int_{|s|\le 1} |K^{\rm ft}(s)|^2 e^{2\mu(|s|/h)^\gamma} ds = O(h^{-1} e^{2\mu h^{-\gamma}}),
\end{aligned}$$

where the first equality follows by Lemma 1 (1), the second equality follows by the change of variables $s = th$, the inequality follows by the supersmoothness of f_ϵ and the fact that $K^{\rm ft}$ is supported on $[-1, 1]$, and the last equality uses $\|K^{\rm ft}\|_\infty < \infty$. □

Lemma 9. *Under Assumptions 5, 6 and 10, there exist constants $c_2 \ge c_1 > 0$ such that*

$$c_1 \le h^{2\beta} E \left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^* - X) dx^* K_h(z_d - Z_d) \right|^2 \le c_2,$$

$$c_1 \le h^{2\beta} E \left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^* - X) dx^* K_h(z_d - Z_d) \big[Y - \mu - \sum_{d'\neq d} m_{d'}(Z_{d'}) - m_d(z_d) \big] \right|^2 \le c_2,$$

for all n large enough. Moreover, if $\operatorname{supp} g = \mathcal{I} = [b_1, b_2]$ *and* $\sup_s \left| g^{\rm ft}(-\frac{s}{h}) \frac{s}{h^2} \right| \to 0$ *as $n \to \infty$, then*

$$\lim_{n\to\infty} h^{2\beta} E \left| \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^* - X) g(x^*) dx^* K_h(z_d - Z_d) \right|^2 = 0,$$

$$\lim_{n\to\infty} h^{2\beta} E\left\{\begin{array}{c}\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)g(x^*)dx^* \int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* \\ \times|K_h(z_d-Z_d)|^2\left[Y-\mu-\sum_{d'\neq d} m_{d'}(Z_{d'})-m_d(z_d)\right]\end{array}\right\}=0.$$

Proof. By $\mathcal{I}=[b_1,b_2]$, decompose

$$\begin{aligned} & h^{2\beta} E\left|\int_{x^*\in\mathcal{I}} \mathbb{K}_h(x^*-X)dx^* K_h(z_d-Z_d)\right|^2 \\ = \;& \frac{h^{2\beta}}{4\pi^2}\int_{u,v}\left|\int_t e^{\mathrm{i}tu}\left[\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right]\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\, K_h(z_d-v)\right|^2 f_{X,Z_d}(u,v)dudv \equiv J_1+J_2+J_3,\end{aligned}$$

where

$$\begin{aligned} J_1 &= \frac{h^{2\beta}}{4\pi^2}\int_{u,v}\left|\int_{|t|<M} e^{\mathrm{i}tu}\left[\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right]\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\, K_h(z_d-v)\right|^2 f_{X,Z_d}(u,v)dudv, \\ J_2 &= \frac{h^{2\beta}}{4\pi^2}\int_{u,v}\left|\int_{|t|\geq M} e^{\mathrm{i}tu}\left[\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right]\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\, K_h(z_d-v)\right|^2 f_{X,Z_d}(u,v)dudv, \\ J_3 &= \frac{h^{2\beta}}{2\pi^2}\int_{u,v}\Re\left\{\begin{array}{c}\int_{|t|<M} e^{\mathrm{i}tu}\left[\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right]\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt \\ \times\int_{|t|\geq M} e^{\mathrm{i}tu}\left[\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right]\frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)}dt\end{array}\right\}|K_h(z_d-v)|^2 f_{X,Z_d}(u,v)dudv,\end{aligned}$$

and M is a constant such that $|f_\epsilon^{\mathrm{ft}}(t)||t|^\beta > c_\epsilon/2$ and $|f_\epsilon^{\mathrm{ft}'}(t)||t|^{\beta+1} < 5\beta c_\epsilon/4$ for any t satisfying $|t|>M$. For J_1, note that

$$|J_1| \leq \frac{h^{2\beta}}{4\pi^2}\left(\int_{|t|<M}\left|\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right|\frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|}dt\right)^2 E|K_h(z_d-Z_d)|^2 = O(h^{2\beta-1}),$$

where the second equality follows by $\left|\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right| \leq |b_2-b_1|$, $\|K^{\mathrm{ft}}\|_\infty<\infty$, ordinary smoothness of f_ϵ, and $hE|K_h(z_d-Z_d)|^2 = f_{Z_d}(z_d)\int K^2(v)dv+o(h)$. Also, for J_3,

$$\begin{aligned} |J_3| &\leq \frac{h^{2\beta}}{\pi^2}\int_{|t|<M}\left|\frac{e^{-\mathrm{i}tb_1}-e^{-\mathrm{i}tb_2}}{\mathrm{i}t}\right|\frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|}dt\int_{|t|\geq M}\frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)||t|}dtE|K_h(z_d-Z_d)|^2 \\ &= O\left(h^{\beta-1}\int_{|t|\geq M}\frac{|K^{\mathrm{ft}}(s)||s|^{\beta-1}}{|f_\epsilon^{\mathrm{ft}}(s/h)||s/h|^\beta}ds\right) = O(h^{\beta-1}),\end{aligned}$$

where the second equality follows by the choice of M and $\int |K^{\mathrm{ft}}(s)||s|^{\beta-1}ds<\infty$.

So, J_2 is the dominant term and can decomposed as $J_2 = J_{2,1}+J_{2,2}+J_{2,3}$, where

$$\begin{aligned} J_{2,1} &= \frac{h^{2\beta}}{4\pi^2}\int_{u,v}\left|\int_{|s|\geq Mh} e^{\frac{\mathrm{i}s(u-b_1)}{h}}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}dsK_h(z_d-v)\right|^2 f_{X,Z_d}(u,v)dudv, \\ J_{2,2} &= \frac{h^{2\beta}}{4\pi^2}\int_{u,v}\left|\int_{|s|\geq Mh} e^{\frac{\mathrm{i}s(u-b_2)}{h}}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}dsK_h(z_d-v)\right|^2 f_{X,Z_d}(u,v)dudv, \\ J_{2,3} &= \frac{h^{2\beta}}{2\pi^2}\int_{u,v}\Re\left\{\begin{array}{c}\int_{|s|\geq Mh} e^{\frac{\mathrm{i}s(u-b_1)}{h}}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}ds \\ \times\int_{|s|\geq Mh} e^{\frac{\mathrm{i}s(u-b_2)}{h}}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}ds\end{array}\right\}|K_h(z_d-v)|^2 f_{X,Z_d}(u,v)dudv.\end{aligned}$$

For $J_{2,1}$ and $J_{2,2}$, we show

$$\begin{aligned} J_{2,1} &\to \frac{f_{X,Z_d}(b_1,z_d)}{2\pi c_\epsilon^2}\int |K^{\mathrm{ft}}(s)|^2|s|^{2\beta-2}ds\int K^2(v)dv, \\ J_{2,2} &\to \frac{f_{X,Z_d}(b_2,z_d)}{2\pi c_\epsilon^2}\int |K^{\mathrm{ft}}(s)|^2|s|^{2\beta-2}ds\int K^2(v)dv. \end{aligned} \tag{D.15}$$

In particular, letting $K_n(u,v)=\frac{h^{2\beta}}{4\pi^2}\left|\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}dsK(v)\right|^2$, we have

$$\begin{aligned} J_{2,1} &= \int h^{-2}K_n\left(\frac{b_1-u}{h},\frac{z_d-v}{h}\right)f_{X,Z_d}(u,v)dudv, \\ J_{2,2} &= \int h^{-2}K_n\left(\frac{b_2-u}{h},\frac{z_d-v}{h}\right)f_{X,Z_d}(u,v)dudv. \end{aligned}$$

Note $K_n(u,v)\to K(u,v)=\frac{1}{4\pi^2c_\epsilon^2}\left|e^{-\mathrm{i}su}K^{\mathrm{ft}}(s)s^{\beta-1}dsK(v)\right|^2$ and $\int K(u,v)dudv=\frac{1}{2\pi c_\epsilon^2}\int|K^{\mathrm{ft}}(s)|^2|s|^{2\beta-2}ds\int K^2(v)dv$ by Plancherel's isometry. Then by Lemma 6, if there exists K^* such that $\sup_n|K_n|\leq|K^*|$ and $\int K^*(u,v)dudv<\infty$, (D.15) would follow. To see this, using integration by parts, we have

$$h^\beta\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}ds=\left.\frac{h^\beta e^{-\mathrm{i}su}K^{\mathrm{ft}}(s)}{\mathrm{i}uf_\epsilon^{\mathrm{ft}}(s/h)s}\right|_{-Mh}^{Mh}+\frac{h^\beta}{\mathrm{i}u}\int_{|s|\geq Mh}e^{-\mathrm{i}su}\left(\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}\right)'ds,$$

where $\left|\frac{h^\beta e^{-\mathrm{i}Mhu}K^{\mathrm{ft}}(Mh)}{\mathrm{i}uf_\epsilon^{\mathrm{ft}}(M)Mh}\right|\to 0$ and $\left|\frac{h^\beta e^{\mathrm{i}Mhu}K^{\mathrm{ft}}(-Mh)}{\mathrm{i}uf_\epsilon^{\mathrm{ft}}(-M)Mh}\right|\to 0$ if $\beta>1$, and

$$\begin{aligned} h^\beta\int_{|s|\geq Mh}e^{-\mathrm{i}su}\left(\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}\right)'ds &= \int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}\prime}(s)s^{\beta-1}}{f_\epsilon^{\mathrm{ft}}(s/h)(s/h)^\beta}ds+\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}}(s)s^{\beta-2}}{f_\epsilon^{\mathrm{ft}}(s/h)(s/h)^\beta}ds \\ &\quad+\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}}(s)s^{\beta-1}f_\epsilon^{\mathrm{ft}\prime}(s/h)(s/h)^{\beta+1}}{\left[f_\epsilon^{\mathrm{ft}}(s/h)(s/h)^\beta\right]^2}ds, \end{aligned}$$

with

$$\begin{aligned} \left|\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}\prime}(s)s^{\beta-1}}{f_\epsilon^{\mathrm{ft}}(s/h)(s/h)^\beta}ds\right| &\leq \int_{|s|\geq Mh}\frac{|K^{\mathrm{ft}\prime}(s)||s|^{\beta-1}}{|f_\epsilon^{\mathrm{ft}}(s/h)||s/h|^\beta}ds\leq\frac{2}{c_\epsilon}\int|K^{\mathrm{ft}\prime}(s)||s|^{\beta-1}ds, \\ \left|\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}}(s)s^{\beta-2}}{f_\epsilon^{\mathrm{ft}}(s/h)(s/h)^\beta}ds\right| &\leq \int_{|s|\geq Mh}\frac{|K^{\mathrm{ft}}(s)||s|^{\beta-2}}{|f_\epsilon^{\mathrm{ft}}(s/h)||s/h|^\beta}ds\leq\frac{2}{c_\epsilon}\int|K^{\mathrm{ft}}(s)||s|^{\beta-2}ds, \end{aligned}$$

and

$$\begin{aligned} \left|\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}}(s)s^{\beta-1}f_\epsilon^{\mathrm{ft}\prime}(s/h)(s/h)^{\beta+1}}{\left[f_\epsilon^{\mathrm{ft}}(s/h)(s/h)^\beta\right]^2}ds,\right| &\leq \int_{|s|\geq Mh}\frac{|K^{\mathrm{ft}}(s)||s|^{\beta-1}|f_\epsilon^{\mathrm{ft}\prime}(s/h)||s/h|^{\beta+1}}{\left[|f_\epsilon^{\mathrm{ft}}(s/h)||s/h|^\beta\right]^2}ds \\ &\leq \frac{5\beta}{c_\epsilon}\int|K^{\mathrm{ft}\prime}(s)||s|^{\beta-1}ds. \end{aligned}$$

By $\int|K^{\mathrm{ft}\prime}(s)||s|^{\beta-1}ds<\infty$ and $\int|K^{\mathrm{ft}}(s)||s|^{\beta-2}ds<\infty$, there exists a constant $c_2>0$ such that $\sup_n|K_n(u,v)|<\frac{c_2|K(v)|^2}{u^2}$. Also, we note

$$h^\beta\left|\int_{|s|\geq Mh}e^{-\mathrm{i}su}\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)s}ds\right|\leq\int_{|s|\geq Mh}\frac{|K^{\mathrm{ft}}(s)||s|^{\beta-1}}{|f_\epsilon^{\mathrm{ft}}(s/h)||s/h|^\beta}ds\leq\frac{2}{c_\epsilon}\int|K^{\mathrm{ft}}(s)||s|^{\beta-1}ds<\infty.$$

Then we can choose $K^*(u,v) = \min\left(c_1|K(v)|^2, \frac{c_2|K(v)|^2}{u^2}\right)$, and it is easy to verify that K^* satisfies the required conditions and (D.15) is obtained.

For the cross-product term $J_{2,3}$, by the Cauchy-Schwarz inequality, we have

$$|J_{2,3}| \le 2\sqrt{J_{2,1}J_{2,2}} \to \frac{\sqrt{f_{X,Z_d}(b_1,z_d)f_{X,Z_d}(b_2,z_d)}}{\pi c_\epsilon^2} \int |K^{\text{ft}}(s)|^2|s|^{2\beta-2}ds \int K^2(v)dv.$$

Thus, by $J_{2,1} + J_{2,2} - |J_{2,3}| \le J_2 \le J_{2,1} + J_{2,2} + |J_{2,3}|$, if $\{f_{X,Z_d}(b_1,z_d) + f_{X,Z_d}(b_2,z_d)\} > 2\sqrt{f_{X,Z_d}(b_1,z_d)f_{X,Z_d}(b_2,z_d)}$, there exist constants $c_2 \ge c_1 > 0$ such that $c_1 \le J_2 \le c_2$ as $n \to \infty$, and the first statement follows by $J_1 = o(1)$ and $J_3 = o(1)$.

By replacing f_{X,Z_d} with $E[|g(X^*) + m_d(Z_d) + U - m_d(z_d)|^2|X,Z_d]f_{X,Z_d}$, a similar argument yields the second statement.

The proofs of the next two statements are similar, so we focus on the third statement. If $\operatorname{supp} g = [b_1,b_2]$, we have

$$\begin{aligned}
& h^{2\beta}E\left|\int \mathbb{K}_h(x^*-X)g(x^*)dx^*K_h(z_d-Z_d)\right|^2 \\
= {} & \frac{h^{2\beta}}{4\pi^2}\int_{u,v}\left|\int_t e^{itu}g^{\text{ft}}(-t)\frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)}dtK_h(z_d-v)\right|^2 f_{X,Z_d}(u,v)dudv \\
= {} & \frac{h^{2\beta}}{4\pi^2}\int_{u,v}\left|\int_{|t|\ge M} e^{itu}g^{\text{ft}}(-t)\frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)}dtK_h(z_d-v)\right|^2 f_{X,Z_d}(u,v)dudv + o(1),
\end{aligned}$$

where the last equality follows by a similar argument as in the proof of the first statement. Also,

$$\begin{aligned}
& h^{\beta}\left|\int_{|t|\ge M} e^{itu}g^{\text{ft}}(-t)\frac{K^{\text{ft}}(th)}{f_\epsilon^{\text{ft}}(t)}dtK_h(z_d-v)\right| \\
\le {} & \int_{|s|\ge Mh}\left|g^{\text{ft}}(-s/h)(s/h^2)\right|\frac{|K^{\text{ft}}(s)||s|^{\beta-1}}{|f_\epsilon^{\text{ft}}(s/h)||s/h|^{\beta}}dtK\left(\frac{z_d-v}{h}\right) \\
\le {} & \frac{2\sup_{|s|\ge Mh}|g^{\text{ft}}(-s/h)s/h^2|\,\|K\|_\infty}{c_\epsilon}\int |K^{\text{ft}}(s)||s|^{\beta-1}ds,
\end{aligned}$$

and the conclusion follows because $\sup_s |g^{\text{ft}}(-s/h)s/h^2|$ can be arbitrarily small for all n large enough. The last statement can be shown in the same manner. □

Lemma 10. *Under Assumptions 4, 5, 8 and 11, there exist constants $c, c' > 0$ such that*

$$h^3e^{-2\mu h^{-\gamma}}E\left|\int_{x^*\in\mathcal{I}}\mathbb{K}_h(x^*-X)dx^*K_h(z_d-Z_d)\right|^2 \le c,$$

$$h^3e^{-2\mu h^{-\gamma}}E\left|\int_{x^*\in\mathcal{I}}\mathbb{K}_h(x^*-X)g(x^*)dx^*K_h(z_d-Z_d)\right|^2 \le c',$$

for all n large enough.

Proof. Let $\mathcal{I} = [b_1, b_2]$. For the first statement, we have

$$
\begin{aligned}
& h^3 e^{-2\mu h^{-\gamma}} E \left| \int_{x^* \in \mathcal{I}} \mathbb{K}_h(x^* - X) dx^* K_h(z_d - Z_d) \right|^2 \\
= \quad & \frac{h^3 e^{-2\mu h^{-\gamma}}}{4\pi^2} \int_{u,v} \left| \int e^{\mathrm{i}tu} \left[\frac{e^{-\mathrm{i}tb_1} - e^{-\mathrm{i}tb_2}}{\mathrm{i}t} \right] \frac{K^{\mathrm{ft}}(th)}{f_\epsilon^{\mathrm{ft}}(t)} dt K_h(z_d - v) \right|^2 f_{X,Z_d}(u,v) du dv \\
\leq \quad & \frac{(b_2 - b_1)^2}{4\pi^2} \left(h e^{-\mu h^{-\gamma}} \int \frac{|K^{\mathrm{ft}}(th)|}{|f_\epsilon^{\mathrm{ft}}(t)|} dt \right)^2 h E |K_h(z_d - Z_d)|^2,
\end{aligned}
$$

where the inequality follows by Lemma 8. The conclusion follows by Lemma 8 and $hE|K_h(z_d - Z_d)|^2 = f_{Z_d}(z_d) \int K^2(v) dv + o(h)$. The second statement is shown in the same manner by using $\|g^{\mathrm{ft}}\|_\infty < \infty$. □

Lemma 11. *Under Assumptions 12 and 13 (1), for any $\tau_n \to \infty$, we have*

$$
\sup_{|t| \leq \tau_n} |\{\check{f}_\epsilon^{\mathrm{ft}}(t)\}^2 - \{f_\epsilon^{\mathrm{ft}}(t)\}^2| = O_p(n^{-1/2} \log(\tau_n)).
$$

Proof. Since $|\{\check{f}_\epsilon^{\mathrm{ft}}(t)\}^2 - \{f_\epsilon^{\mathrm{ft}}(t)\}^2| \leq |\frac{1}{n} \sum_{j=1}^n e^{\mathrm{i}t(\epsilon_j - \epsilon_j^r)} - \{f_\epsilon^{\mathrm{ft}}(t)\}^2|$, the conclusion follows from Assumption 13 (1) and Kurisu and Otsu (2020, Lemma 1). □

Lemma 12. *Suppose Assumptions 12 and 13 hold true. Under Assumption 3 or 4, we have*

$$
\begin{aligned}
|\check{E}[P_\kappa P_\kappa'] - \hat{E}[P_\kappa P_\kappa']|^2 &= O_p\left(\kappa^2 n^{-1/2 + 3\zeta} \log(\rho_{\epsilon,n,\zeta,1}) + \kappa^2 n^{3\zeta} \max_{1 \leq k \leq \kappa} \int_{G_{\epsilon,n,\zeta,0}} |p_k^{\mathrm{ft}}(t)| dt \right), \\
|\check{E}[P_\kappa Y] - \hat{E}[P_\kappa Y]|^2 &= O_p\left(\kappa n^{-1/2 + 3\zeta} \log(\rho_{\epsilon,n,\zeta,1}) + \kappa n^{3\zeta} \max_{1 \leq k \leq \kappa} \int_{G_{\epsilon,n,\zeta,0}} |p_k^{\mathrm{ft}}(t)| dt \right),
\end{aligned}
$$

where $\rho_{\epsilon,n,\zeta,\iota}$ denotes $c_{os,\iota}^{1/\beta} n^{\zeta/\beta} - 1$ under Assumption 3 and $\mu^{-1/\gamma} \log(c_{\mathrm{ss},\iota} n^\zeta)^{1/\gamma}$ under Assumption 4, and $G_{\epsilon,n,\zeta,\iota} = \{t \in \mathbb{R} : |t| > \rho_{\epsilon,n,\zeta,\iota}\}$ for $\iota = 0, 1$.

Proof. Since the proof is similar, we focus on $|\check{E}[P_\kappa P_\kappa'] - \hat{E}[P_\kappa P_\kappa']|^2$. Elements of $\check{E}[P_\kappa P_\kappa'] - \hat{E}[P_\kappa P_\kappa']$ are in the form of $\check{E}[p(X^*)Q] - \hat{E}[p(X^*)Q]$, which for $r = 0$ can be expressed as

$$
\check{E}[p(X^*)Q] - \hat{E}[p(X^*)Q] = R_{1,p,Q} + R_{2,p,Q},
$$

where

$$
\begin{aligned}
R_{1,p,Q} &= \int \Xi_{p,Q}(t) \frac{\check{f}_\epsilon^{\mathrm{ft}}(t) - f_\epsilon^{\mathrm{ft}}(t)}{\{\check{f}_\epsilon^{\mathrm{ft}}(t) \vee n^{-\zeta}\}^2} dt, \qquad \Xi_{p,Q}(t) = \frac{p^{\mathrm{ft}}(-t)}{2\pi n} \sum_{j=1}^n Q_j e^{\mathrm{i}tX_j}, \\
R_{2,p,Q} &= \int \Xi_{p,Q}(t) f_\epsilon^{\mathrm{ft}}(t) \left\{ \frac{1}{\{\check{f}_\epsilon^{\mathrm{ft}}(t) \vee n^{-\zeta}\}^2} - \frac{1}{\{f_\epsilon^{\mathrm{ft}}(t) \vee n^{-\zeta}\}^2} \right\} dt.
\end{aligned}
$$

Here, we use the fact that $f_\epsilon^{\mathrm{ft}}(-t) = f_\epsilon^{\mathrm{ft}}(t)$ and $f_\epsilon^{\mathrm{ft}}(t) > 0$ for all $t \in \mathbb{R}$ under Assumption 12.

First, decompose $R_{1,p,Q} = R_{11,p,Q} + R_{12,p,Q}$, where $R_{11,p,Q}$ and $R_{12,p,Q}$ are the integration over $G^c_{\epsilon,n,\zeta}$ and $G_{\epsilon,n,\zeta}$, respectively. For $R_{11,p,Q}$, we have

$$
\begin{aligned}
|R_{11,p,Q}| &\leq \frac{\sum_{j=1}^n |Q_j|}{2\pi n} \int_{G^c_{\epsilon,n,\zeta}} |p^{\rm ft}(-t)| \left\{ \frac{|\{\check{f}^{\rm ft}_\epsilon(t)\}^2 - \{f^{\rm ft}_\epsilon(t)\}^2|}{\{\check{f}^{\rm ft}_\epsilon(t) + f^{\rm ft}_\epsilon(t)\}\{\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta}\}^2} \right\} dt \\
&\leq \frac{\sum_{j=1}^n |Q_j|}{2\pi n^{1-3\zeta}} \int |p^{\rm ft}(t)| dt \sup_{t \in G^c_{\epsilon,n,\zeta,1}} |\{\check{f}^{\rm ft}_\epsilon(t)\}^2 - \{f^{\rm ft}_\epsilon(t)\}^2| \\
&= O_p(n^{-1/2+3\zeta} \log(\rho_{\epsilon,n,\zeta,1})), \qquad \text{(D.16)}
\end{aligned}
$$

where the second step follows from the fact that $f^{\rm ft}_\epsilon(t) > n^{-\zeta}$ for $t \in G^c_{\epsilon,n,\zeta}$, $\check{f}^{\rm ft}_\epsilon(t) \geq 0$, $\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta} \geq n^{-\zeta}$, and $G^c_{\epsilon,n,\zeta} \subseteqq G^c_{\epsilon,n,\zeta,1}$, and the last step follows from $\frac{1}{n}\sum_{j=1}^n |Q_j| = O_p(1)$ and $\int |p^{\rm ft}(t)| dt < \infty$ (Assumption 13 (2)), and Lemma 11. For $R_{12,p,Q}$, we have

$$
\begin{aligned}
|R_{12,p,Q}| &\leq \frac{\sum_{j=1}^n |Q_j|}{2\pi n} \int_{G_{\epsilon,n,\zeta}} |p^{\rm ft}(-t)| \left\{ \frac{|\check{f}^{\rm ft}_\epsilon(t) - f^{\rm ft}_\epsilon(t)|}{\{\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta}\}^2} \right\} dt \\
&\leq \frac{\sum_{j=1}^n |Q_j|}{\pi n^{1-2\zeta}} \int_{G_{\epsilon,n,\zeta,0}} |p^{\rm ft}(t)| dt = O_p\left(n^{2\zeta} \int_{G_{\epsilon,n,\zeta,0}} |p^{\rm ft}(t)| dt \right), \qquad \text{(D.17)}
\end{aligned}
$$

where the second step follows from the fact that $\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta} \geq n^{-\zeta}$, $|\check{f}^{\rm ft}_\epsilon(t) - f^{\rm ft}_\epsilon(t)| \leq 2$, and $G_{\epsilon,n,\zeta} \subseteqq G_{\epsilon,n,\zeta,0}$, and the last step follows from $\frac{1}{n}\sum_{j=1}^n |Q_j| = O_p(1)$ (Assumption 13 (2)).

Similarly, decompose $R_{2,p,Q} = R_{21,p,Q} + R_{22,p,Q}$, where $R_{21,p,Q}$ and $R_{22,p,Q}$ are the integration over $G^c_{\epsilon,n,\zeta}$ and $G_{\epsilon,n,\zeta}$, respectively. For $R_{21,p,Q}$, we have

$$
\begin{aligned}
|R_{21,p,Q}| &\leq \frac{\sum_{j=1}^n |Q_j|}{2\pi n} \int_{G^c_{\epsilon,n,\zeta}} |p^{\rm ft}(-t)| \left\{ \frac{|\{\check{f}^{\rm ft}_\epsilon(t)\}^2 - \{f^{\rm ft}_\epsilon(t)\}^2|}{f^{\rm ft}_\epsilon(t)\{\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta}\}^2} \right\} dt \\
&\leq \frac{\sum_{j=1}^n |Q_j|}{2\pi n^{1-3\zeta}} \int |p^{\rm ft}(t)| dt \sup_{t \in G^c_{\epsilon,n,\zeta,1}} |\{\check{f}^{\rm ft}_\epsilon(t)\}^2 - \{f^{\rm ft}_\epsilon(t)\}^2| \\
&= O_p(n^{-1/2+3\zeta} \log(\rho_{\epsilon,n,\zeta,1})), \qquad \text{(D.18)}
\end{aligned}
$$

where the first step follows from the fact that $\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta} \geq \check{f}^{\rm ft}_\epsilon(t)$, the second step follows from $f^{\rm ft}_\epsilon(t) > n^{-\zeta}$, $\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta} \geq n^{-\zeta}$, and $G_{\epsilon,n,\zeta,1} \subseteq G_{\epsilon,n,\zeta}$, and the last step follows from $\frac{1}{n}\sum_{j=1}^n |Q_j| = O_p(1)$ and $\int |p^{\rm ft}(t)| dt < \infty$ (Assumption 13 (2)), and Lemma 11. For $R_{22,p,Q}$, we have

$$
\begin{aligned}
|R_{22,p,Q}| &\leq \frac{\sum_{j=1}^n |Q_j|}{2\pi n^{1-\zeta}} \int_{G_{\epsilon,n,\zeta}} |p^{\rm ft}(-t)| \left\{ \frac{|n^{-2\zeta} - \{\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta}\}^2|}{\{\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta}\}^2} \right\} dt \\
&\leq \frac{\sum_{j=1}^n |Q_j|}{2\pi n^{1-3\zeta}} \int_{G_{\epsilon,n,\zeta}} |p^{\rm ft}(-t)| |\{\check{f}^{\rm ft}_\epsilon(t)\}^2 - n^{-2\zeta}| dt \\
&\leq \frac{\sum_{j=1}^n |Q_j|}{\pi n^{1-3\zeta}} \int_{G_{\epsilon,n,\zeta,0}} |p^{\rm ft}(t)| dt = O_p\left(n^{3\zeta} \int_{G_{\epsilon,n,\zeta,0}} |p^{\rm ft}(t)| dt \right), \qquad \text{(D.19)}
\end{aligned}
$$

where the second step follows from $n^{-2\zeta} - \{\check{f}^{\rm ft}_\epsilon(t) \vee n^{-\zeta}\}^2 = 0$ if $\check{f}^{\rm ft}_\epsilon(t) \leq n^{-\zeta}$, the third step follows from $|\{\check{f}^{\rm ft}_\epsilon(t)\}^2 - n^{-2\zeta}| \leq |\{\check{f}^{\rm ft}_\epsilon(t)\}^2 - \{f^{\rm ft}_\epsilon(t)\}^2| \leq 2$ for $t \in G_{\epsilon,n,\zeta}$ and $G_{\epsilon,n,\zeta} \subseteqq G_{\epsilon,n,\zeta,0}$, and the last step follows from $\frac{1}{n}\sum_{j=1}^n |Q_j| = O_p(1)$ (Assumption 13 (2)). The conclusion then follows by combining (D.16), (D.17), (D.18), and (D.19). □

References

[1] An, Y., Wang, L. and R. Xiao (2020) A Nonparametric Nonclassical Measurement Error Approach to Estimating Intergenerational Mobility Elasticities, *Journal of Business & Economic Statistics*, 1-17.

[2] Blinder, A.S. (1973) Wage discrimination: reduced form and structural estimates, *Journal of Human Resources*, 436-455.

[3] Bound, J., Brown, C. and N. Mathiowetz (2001) Measurement error in survey data, in Heckman, J. J. and E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, 3705-3843, Elsevier.

[4] Bonhomme, S. and J. M. Robin (2010) Generalized non-parametric deconvolution with an application to earnings dynamics, *The Review of Economic Studies*, 77, 491-533.

[5] Buja, A., Hastie, T. and R. Tibshirani (1989) Linear smoothers and additive models, *Annals of Statistics*, 17, 453-510.

[6] Card, D. and T. Lemieux (1996) Wage dispersion, returns to skill, and black-white wage differentials, *Journal of Econometrics*, 74, 319-361.

[7] Carroll, R. J. and P. Hall (1988) Optimal rates of convergence for deconvolving a density, *Journal of the American Statistical Association*, 83, 1184-1186.

[8] Chay, K.Y. and D. S. Lee (2000) Changes in relative wages in the 1980s: Returns to observed and unobserved skills and black-white wage differentials, *Journal of Econometrics*, 99, 1-38.

[9] Deaton, A. and J. Muellbauer (1980) *Economics and consumer behavior,* Cambridge university press.

[10] Delaigle, A., Fan, J. and R. J. Carroll (2009) A design-adaptive local polynomial estimator for the errors-in-variables problem, *Journal of the American Statistical Association*, 104, 348-359.

[11] Delaigle, A. and P. Hall (2008) Using SIMEX for smoothing-parameter choice in errors-in-variables problems, *Journal of the American Statistical Association*, 103, 280-287.

[12] Delaigle, A., Hall, P. and A. Meister (2008) On deconvolution with repeated measurements, *Annals of Statistics*, 36, 665-685.

[13] Dominici, F., McDermott, A., Zeger, S.L. and J.M. Samet (2002) On the use of generalized additive models in time-series studies of air pollution and health, *American Journal of Epidemiology*, 156, 193-203.

[14] Dong, H., Otsu, T. and L. Taylor (2020) Estimation of varying coefficient models with measurement error, Working paper.

[15] Fan, J. (1992) Deconvolution with supersmooth distributions, *Canadian Journal of Statistics*, 20, 155-169.

[16] Fan, J. (1991a) Asymptotic normality for deconvolution kernel density estimators, *Sankhyā*, A 53, 97-110.

[17] Fan, J. (1991b) On the optimal rates of convergence for nonparametric deconvolution problems, *Annals of Statistics*, 19, 1257-1272.

[18] Fan, J. and E. Masry (1992) Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes, *Journal of Multivariate Analysis*, 43, 237-271.

[19] Fan, J. and Y. K. Truong (1993) Nonparametric regression with errors in variables, *Annals of Statistics*, 21, 1900-1925.

[20] Gottschalk, P. and M. Huynh (2010) Are earnings inequality and mobility overstated? The impact of non-classical measurement error, *Review of Economics and Statistics*, 92, 302-315.

[21] Hall, P. and S. N. Lahiri (2008) Estimation of distributions, moments and quantiles in deconvolution problems, *Annals of Statistics*, 36, 2110-2134.

[22] Hall, P. and A. Meister (2007) A ridge-parameter approach to deconvolution, *Annals of Statistics*, 35, 1535-1558.

[23] Han, K. and B. U. Park (2018) Smooth backfitting for errors-in-variables additive models, *Annals of Statistics*, 46, 2216-2250.

[24] Horowitz, J. L. (2014) Nonparametric additive models, in Racine, J. S., Su, L. and A. Ullah (eds.), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 129-148, Oxford University Press.

[25] Horowitz, J. L. and S. Lee (2005) Nonparametric estimation of an additive quantile, *Journal of the American Statistical Association*, 100, 1238-1249.

[26] Horowitz, J. L. and E. Mammen (2004) Nonparametric estimation of an additive model with a link function, *Annals of Statistics*, 32, 2412-2443.

[27] Hu, Y. (2017) The econometrics of unobservables: applications of measurement error models in empirical industrial organization and labor economics, *Journal of Econometrics*, 200, 154-168.

[28] Hu, Y. and Y. Sasaki (2015) Closed-form estimation of nonparametric models with non-classical measurement errors, *Journal of Econometrics*, 185, 392-408.

[29] Hu, Y. and S. M. Schennach (2008) Instrumental variable treatment of nonclassical measurement error models, *Econometrica*, 76, 195-216.

[30] Hyslop, D. R. and G. W. Imbens (2001) Bias from classical and other forms of measurement error, *Journal of Business & Economic Statistics*, 19, 475-481.

[31] Kurisu, D. and T. Otsu (2020) On the uniform convergence of deconvolution estimators from repeated measurements, forthcoming in *Econometric Theory*.

[32] Lang, K. and J. Y. K. Lehmann (2012) Racial discrimination in the labor market: Theory and empirics, *Journal of Economic Literature*, 50, 959-1006.

[33] Lang, K. and M. Manove (2011) Education and labor market discrimination, *American Economic Review*, 101, 1467-96.

[34] Leontief, W. (1947) Introduction to a theory of the internal structure of functional relationships, *Econometrica*, 361-373.

[35] Linton, O. B. and J. P. Nielsen (1995) A kernel method of estimating structured nonparametric regression based on marginal integration, *Biometrika*, 82, 93-100.

[36] Linton, O. B. and W. Härdle (1996) Estimation of additive regression models with known links, *Biometrika*, 83, 529-540.

[37] Mammen, E. and Linton, O. B. and J. P. Nielsen (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions, *Annals of Statistics*, 27, 1443-1490.

[38] Meister, A (2009) *Deconvolution Problems in Nonparametric Statistics*, Springer.

[39] Neal, D. (2004) The measured black-white wage gap among women is too small, *Journal of political Economy*, 112, 1-28.

[40] Newey, W. K. (1997) Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics*, 79, 147-168.

[41] Opsomer, J. D. (2000) Asymptotic properties of backfitting estimators, *Journal of Multivariate Analysis*, 73, 166-179.

[42] Opsomer, J. D. and D. Ruppert (1997) Fitting a bivariate additive model by local polynomial regression, *Annals of Statistics*, 25, 186-211.

[43] Ozabaci, D., Henderson, D. J. and L. Su (2014) Additive nonparametric regression in the presence of endogenous regressors, *Journal of Business & Economic Statistics*, 32, 555-575.

[44] Schennach, S. M. (2007) Instrumental variable estimation of nonlinear errors-in-variables models, *Econometrica*, 75, 201-239.

[45] Schennach, S. M. (2019) Convolution without independence, *Journal of Econometrics*, 211, 308-318.

[46] Schennach, S. M. (2020) Mismeasured and unobserved variables, in Durlauf, S. N., Hansen, L. P., Heckman, J. J. and Matzkin, R. L. (eds.), *Handbook of Econometrics*, vol. 7A, 487-565.

[47] Stefanski, L. A. and R. J. Carroll (1990) Deconvolving kernel density estimators, *Statistics*, 21, 169-184.

[48] Stone, C. J. (1985) Additive regression and other nonparametric models, *Annals of Statistics*, 13, 689-705.

[49] Stone, C. J. (1986) The dimensionality reduction principle for generalized additive models, *Annals of Statistics*, 14, 590-606.

[50] Tripathi, G. (1999) A matrix extension of the Cauchy-Schwarz inequality, *Economics Letters*, 63, 1-3.

[51] Xu, B. and B. Lin (2015) How industrialization and urbanization process impacts on CO2 emissions in China: evidence from nonparametric additive regression models. *Energy Economics*, 48, 188-202.

[52] Xu, B. and B. Lin (2016) Reducing CO2 emissions in China's manufacturing industry: Evidence from nonparametric additive regression models. *Energy*, 101, 161-173.

Department of Economics, Southern Methodist University, 3300 Dyer Street, Dallas, TX 75275, US.

Email address: haod@mail.smu.edu

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: t.otsu@lse.ac.uk

Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4 building 2631, 12 8210 Aarhus V, Denmark

Email address: lntaylor@econ.au.dk