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; the presence of measurement error in covariates is one such situation. Here, 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 are generally inconsistent to estimate the unknown functions.

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 ϵ is a measurement error and independent of X^*. Furthermore, for the most part, we focus on the case where X is scalar and the density of ϵ is known to the researcher. 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 stage 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. As such, this paper contributes to the literature on nonparametric additive models by developing the first estimator that achieves asymptotic normality in an errors-in-variables case, and dealing with supersmooth measurement error in a covariate.

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. In Section 3.1, we derive the convergence rate of the first stage estimator. In Section 3.2, we establish the limiting distribution of the second stage estimator. Sections 4 and 5 present a simulation study and an empirical application, respectively. Finally, Section 6 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 $\|A\| = [\mathrm{trace}(A^\dagger A)]^{1/2}$ 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 is 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), for which we make the following assumptions.

Assumption 1. **(1):** *ϵ is independent of (Y, X^*, Z).*

(2): *The distribution of (ϵ, X^*, Z) is absolutely continuous with respect to the Lebesgue measure.*

(3): *The density f_ϵ of ϵ is known and $f_\epsilon^{\mathrm{ft}}(t) \neq 0$ for all $t \in \mathbb{R}$.*

(4): *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 $d = 1, \ldots, D$.*

(5): $E[U|X^*, Z] = 0$.

(6): *$g, m_1, \ldots, m_D$ 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 that the measurement error is classical in nature. Assumption 1 (2) guarantees the existence of densities on which the proceeding discussions rely. Assumption 1 (3) is commonly used in the literature on nonparametric estimation with measurement error (see, Meister, 2009, for a review). In particular, the assumption f_ϵ being known is restrictive in economic applications. In Remark 1 below, we discuss how to relax this assumption by using auxiliary information such as repeated measurements so that f_ϵ can be identified from the distribution of observables. Assumptions 1 (5) and (6) are normalizations required for identification.

Assumption 1 (4) requires all covariates to be continuously distributed on their support. 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. It is assumed that the density of X^* is bounded away from zero on $\mathcal{I}$ so that the conditional expectations (on the event $X^* \in \mathcal{I}$) 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 by an application of the marginal integration argument for the nonparametric additive model combined with identification of the joint density of (Y, X^*, Z) based on the deconvolution technique. 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 implemented 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, this method is obviously infeasible because X^* is unobservable. 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 the expectations $E[P_\kappa P_\kappa']$ and $E[YP_\kappa']$. Any moment that does not involve X^* can be estimated by the conventional method of moments. For the moments depending on X^*, we must employ a deconvolution technique.

We first consider estimation of $E[Yp_k(X^*)]$, which appears in $E[YP_\kappa']$. To this end, we first introduce some notation. By Plancherel's isometry (see Lemma 1 (1) in Appendix E), the moment of interest is written as

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

where $m(\cdot) = E[Y|X^* = \cdot]$, and the last equality follows by the law of iterated expectations and independence of ϵ and (Y, X^*) (Assumption 1 (1)). A naive estimator of this moment could be obtained by replacing $E[Ye^{\text{i}tX}]$ by its sample analog $n^{-1}\sum_{j=1}^n Y_j e^{\text{i}tX_j}$. However, it is well known that this estimator is not well-behaved since $f_\epsilon^{\text{ft}}(t) \to 0$ as $|t| \to \infty$. Intuitively, the estimation error of the sample analog can be severely amplified in tails, so that the above integral may not be well-behaved. To deal with such situations, it is common to introduce some form of regularization. Here we employ the ridge approach of Hall and Meister (2007) and suggest estimating $E[Yp_k(X^*)]$ by

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

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.

Similarly, the moments $E[p_k(X^*)q_{0l}(Z_d)]$ and $E[p_k(X^*)p_l(X^*)]$ appearing in $E[P_\kappa P_\kappa']$ can be estimated by

$$\hat{E}[p_k(X^*)q_{0l}(Z_d)] = \frac{1}{2\pi}\int\left(\frac{1}{n}\sum_{j=1}^{n} q_l(Z_{d,j})e^{\mathrm{i}tX_j}\right)\frac{[p_0p_k]^{\mathrm{ft}}(-t)f_\epsilon^{\mathrm{ft}}(-t)|f_\epsilon^{\mathrm{ft}}(t)|^r}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}dt,$$

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

By applying these estimators to each element 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 the 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 only uses 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 $Y_j - \hat{\mu} - \hat{g}(X_j^*) - \sum_{d'\neq d}^{D}\hat{m}_{d'}(Z_{d',j})$ in a nonlinear way, instead of simply acting as a covariate. One immediate 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:

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

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

Remark 1. [Case of unknown f_ϵ] In the next section, we investigate the asymptotic properties of the first- and second-stage estimators of $g, m_1, \ldots, m_D$. Before proceeding further, we comment on a major limitation of our estimator: Assumption 1 (3). This assumption says that the measurement error density, f_ϵ, is known to the researcher, which is unrealistic in many settings. In general, with a single noisy measurement of X^*, f_ϵ cannot be identified. However, identification of f_ϵ can be restored if two or more independent noisy measurements of X^* are available. With repeated measurements of X^*, we can obtain an estimator of f_ϵ by applying existing methods, such as Li and Vuong (1998), Delaigle, Hall and Meister (2008), and Bonhomme and Robin (2010). Then, by replacing f_ϵ with its estimated counterpart, we can adapt our two-stage estimation method for the case of unknown f_ϵ. Although a formal analysis is beyond the scope of this paper, our asymptotic theory in the next section provides a framework to analyze such a plug-in estimator for the case of unknown f_ϵ. In particular, we expect that providing the estimator of f_ϵ converges at a sufficiently fast rate, similar asymptotic results to those of the next section can be established.[1]

[1] For example, if f_ϵ^{ft} is estimated by the methods of Li and Vuong (1998) or Comte and Kappus (2015), then the uniform convergence rate result in Kurisu and Otsu (2020) can be applied. If we additionally assume symmetry of f_ϵ around zero, then we can apply the estimator by Delaigle, Hall and Meister (2008) for f_ϵ^{ft}, and its uniform convergence rate can be obtained by Kato and Sasaki (2018, Lemma 4). Even if the convergence rates of these estimators for f_ϵ^{ft} are not sufficiently fast, we can adapt a subsample-based modification as in Adusumilli *et al.* (2020), i.e., use a subsample to compute our estimator of the nonparametric additive model but insert an estimator for f_ϵ^{ft} using the full sample so that the estimation error of f_ϵ^{ft} becomes asymptotically negligible compared to that of the additive model.

Remark 2. [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 not observable, 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

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

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

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 3. [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.

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

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 4. [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). 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 measurement error that our method can be applied to. As an example, consider the case when W is a noisy measurement of W^* that 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 the 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^* that 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^*,[2] 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 $\mathcal{F}_{\alpha,c} = \{f \in L_2(\mathbb{R}) : \int |f^{\mathrm{ft}}(t)|^2(1+|t|^2)^\alpha dt \leq c\}$ denote the Sobolev class of order $\alpha > 0$ and $c > 0$.[3] 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) and (1.2).*

- **(2):** *$E[Y^2|X^*, Z] < \infty$.*
- **(3):** *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 $\mathcal{F}_{\alpha,c_{\mathrm{sob}}}$ for all $d, d' = 1, \ldots, D$ and $z_d, z_{d'} \in [-1,1]$.*
- **(4):** *$\{p_k\}_{k=1}^\infty$ is a series of basis functions on $\mathcal{I}$ such that $\int_{\mathcal{I}} p_k(w)dw = 0$ for all k, and $\int_{\mathcal{I}} p_k(w)p_{k'}(w)dw = \delta_{k,k'}$ for all k, k'.*
- **(5):** *$\{q_k\}_{k=1}^\infty$ is a series of basis functions on $[-1,1]$ such that $\int_{-1}^1 q_k(w)dw = 0$ for all k, and $\int_{-1}^1 q_k(w)q_{k'}(w)dw = \delta_{k,k'}$ for all k, k'.*
- **(6):** *$\lambda_{\min}(E[P_\kappa P_\kappa']) \geq \underline{\lambda} > 0$ for all κ.*
- **(7):** *$\sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D} \|P_\kappa(x^*, z)\| = O(\kappa^{1/2})$ as $\kappa \to \infty$.*
- **(8):** *There exists $\theta_0 = (\mu_0, \theta_0^0, \theta_0^1, \ldots, \theta_0^D)$ such that*

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

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 $d = 1, \ldots, D$.
- **(9):** *$r \geq 0$, $\zeta > 0$, and $\kappa \to \infty$ as $n \to \infty$.*

Assumption 2 (1) and (2) are standard for cross-section data. Extensions to more general data environments are beyond the scope of this paper. Assumption 2 (3) lists the Sobolev conditions

---

[2] 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 as 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^* decided 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).

[3] 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 $\int |f^{\mathrm{ft}}(t)|^2|t|^{2\alpha}dt < c$. First, it is easy to see that $\int |f^{\mathrm{ft}}(t)|^2(1+|t|^2)^\alpha dt < c$ implies $\int |f^{\mathrm{ft}}(t)|^2|t|^{2\alpha}dt < c$. For the other direction, we have $\int |f^{\mathrm{ft}}(t)|^2(1+|t|^2)^\alpha dt \leq 2^\alpha \int_{|t|\leq 1} |f^{\mathrm{ft}}(t)|^2 dt + 2^\alpha \int |f^{\mathrm{ft}}(t)|^2|t|^{2\alpha}dt < c'$, 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)).

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 (4) and (5) are conditions on the basis functions $\{p_k\}_{k=1}^{\infty}$ and $\{q_k\}_{k=1}^{\infty}$. Similar conditions are adopted by Horowitz and Mammen (2004) for the first-stage estimator without measurement errors. Assumption 2 (6)-(8) are commonly used for series-based estimation; see, e.g., Newey (1997, Assumptions 2 and 3). Assumption 2 (9) contains mild requirements of 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 Assumption 2 (3) and (8) are both smoothness conditions, they focus on different objects for different purposes: Assumption 2 (3) imposes smoothness of densities and products of densities and regression functions to control the estimation bias, while Assumption 2 (8) 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 (8) essentially requires that g and m_d are twice continuously differentiable, which is consistent with Assumption 2 (3) 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 $\beta > 1/2$.*

Assumption 4. **(1):** *f_ϵ is supersmooth of order $\gamma > 0$.*
(2): *$r = 0$ and $0 < \zeta < \frac{1}{4}$.*

Assumption 4 (2) contains further conditions on smoothing parameters r and ζ. In the supersmooth case, by setting $r = 0$, we can maximize flexibility in the choice of ζ. Given $r = 0$, Assumption 4 (2) 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 magnitudes of the estimation bias, while the variance of the estimation error is dominated under Assumption 4 (2). 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 5. [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 $\beta > 1/2$,[4] 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 $\hat{g}$.[5] 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).[6]

Remark 6. [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 $\alpha = 2$,[7] the optimal convergence rate of our $\hat{g}$ and $\hat{m}_d$ is slower than $n^{-\frac{1}{4}}$; thus, it is slower than $n^{-\frac{3}{10}}$, which is the optimal convergence rate obtained in Horowitz and Mammen (2004). 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 7. [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, 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

[4]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.

[5]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.

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

[7]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.

$r = 0$ 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 $\int |\widehat{m(x)f_{X^*}}(x) - m(x)f_{X^*}(x)|^2dx$ 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 collects regularity conditions used to derive the asymptotic distributions of $\tilde{g}$ and $\tilde{m}$. Assumption 5 (1) contains smoothness conditions regarding 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) is 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): *$\int |s|^\beta\{|K^{\mathrm{ft}}(s)| + |K^{\mathrm{ft}\prime}(s)|\}ds < \infty$, $\int |s|^{2\beta}|K^{\mathrm{ft}}(s)|^2ds < \infty$.*
(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 smoothness, 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), 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 C.*

Assumption 9 (1) requires the bandwidth h to decay at most at a logarithmic rate, which is due to the fact that the error characteristic function in the denominator decays 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 that 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 C, $Var[\tilde{g}(x^*)]$ 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 C.

As 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): *$\sup_s\left|g^{\mathrm{ft}}\left(-\frac{s}{h}\right)\frac{s}{h^2}\right|\to 0$ 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 we have an extra layer of integration of x^* over $\mathcal{I}$ in the definition of $\tilde{m}_d$, and must be specific for the smoothness of the integration. In Assumption 10 (2), we require $\beta\geq 2$, which is a technical assumption to guarantee $\int|K^{\mathrm{ft}}(s)||s|^{\beta-2}ds<\infty$. 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 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 extra layer of 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): *$nh^3e^{-2\mu h^{-\gamma}}\to\infty$ as $n\to\infty$.*

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

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 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 that 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 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 D, $Var[\tilde{m}_d(z_d)]$ diverges at the rate $h^{-2\beta}$, which is slower than that of $\tilde{g}$ due to the extra layer of 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 D. 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. 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 7 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, $n = 500$ 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.

5. Empirical application

In this section, we use our estimator to analyse 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 was the Armed Services Vocational Aptitude Battery (ASVAB), a series of multiple-choice tests designed to measure cognitive ability. The individuals were then periodically interviewed in the following years, with the most recent interview being conducted in 2016. Among other variables, information on education, tenure, and wages was 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 it is 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). As stated previously in this paper, our estimation strategy assumes the distribution and standard deviation of the measurement error to be known, thus, we must specify choices. Here, 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 4. In Figures 5.1 - 5.3, 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).

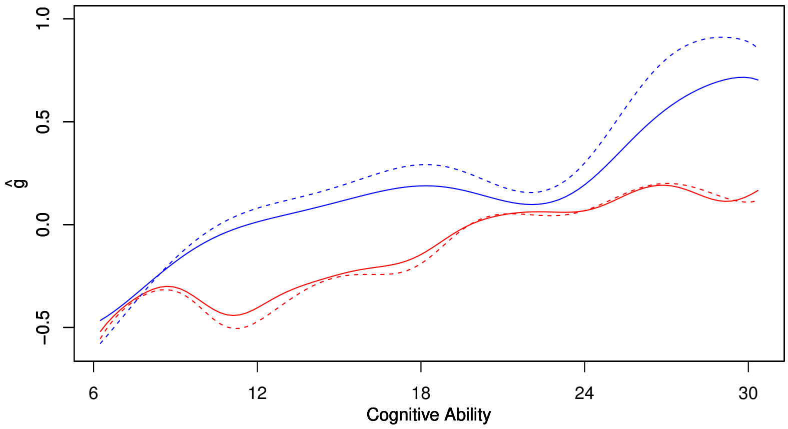

FIGURE 5.1. Plot of $\hat{g}$ between the 1% and 99% quantile of the ASVAB test score using the NLSY79 dataset. The blue lines give results for black individuals and the red lines for white individuals. The solid lines use a standard deviation of 0.1 for the measurement error and the dashed lines use 0.3.

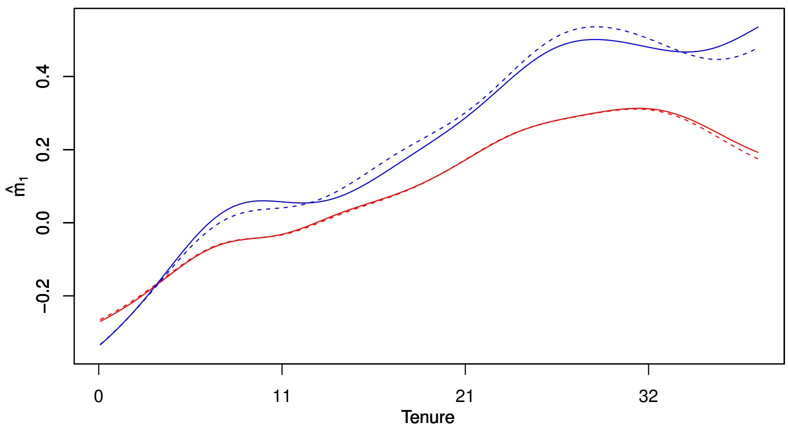

FIGURE 5.2. Plot of $\hat{m}_1$ between the 1% and 99% quantile of tenure (in years) using the NLSY79 dataset. The blue lines give results for black individuals and the red lines for white individuals. The solid lines use a standard deviation of 0.1 for the measurement error and the dashed lines use 0.3.

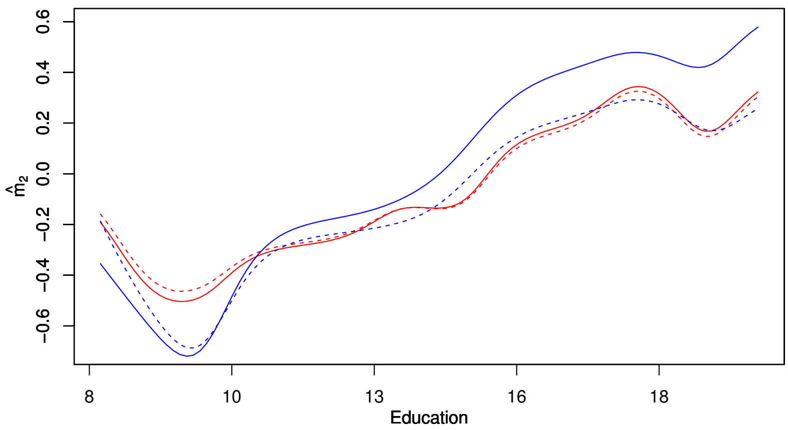

FIGURE 5.3. Plot of $\hat{m}_2$ between the 1% and 99% quantile of education (in years) using the NLSY79 dataset. The blue lines give results for black individuals and the red lines for white individuals. The solid lines use a standard deviation of 0.1 for the measurement error and the dashed lines use 0.3.

6. 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 the 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 has not been addressed by Han and Park (2018) yet, 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 and situations where the measurement error distribution is unknown but auxiliary information such as repeated measurements are available.

Appendix A. 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 $f^{\text{ft}}_{Y,X,Z}(y, \cdot, z)(t) = \int f_{Y,X,Z}(y, x, z)e^{\text{i}tx}dx$. By Assumption 1 and Lemma 1 (2), the joint density $f_{Y,X^*,Z}$ 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. Thus, by Assumption 1, $g, m_1, \ldots, m_D$, and μ are identified as

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

for $d = 1, \ldots, D$. ■

Appendix B. 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 first inequality follows by Jensen's inequality, and the second inequality follows by $\lambda_{\max}(A) = \sup_{\|\delta\|=1} \delta' A\delta$ and $\lambda_{\max}(A'A) \leq \|A\|^2$.

Note $\|\hat{M}_\kappa - M_\kappa\|^2 \leq \|\hat{E}[P_\kappa P'_\kappa] - M_\kappa\|^2$ and $\|\hat{C}_\kappa - C_\kappa\|^2 \leq \|\hat{E}[P_\kappa Y] - C_\kappa\|^2$. So the orders of $\|\hat{M}_\kappa - M_\kappa\|^2$ and $\|\hat{C}_\kappa - C_\kappa\|^2$ are obtained by Lemma 4. We also note that $\lambda_{\max}(\hat{M}_\kappa^{-2}) = \lambda_{\min}^{-2}(\hat{M}_\kappa)$, and $\lambda_{\min}(A) = \inf_{\|\delta\|=1} \delta' A\delta$. Thus, the upper bound of $\lambda_{\max}(\hat{M}_\kappa^{-2})$ follows by

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

$$\left(\inf_{\|\delta\|=1} \delta'(\hat{M}_\kappa - M_\kappa)\delta\right)^2 \leq \|\hat{M}_\kappa - M_\kappa\|^2 \xrightarrow{p} 0,$$

and $\lambda_{\min}(M_\kappa) \geq \underline{\lambda} > 0$. Moreover, we note $C_\kappa = E[P_\kappa E[Y|X^*, Z]]$ and

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

where the first inequality follows by the property of the maximum eigenvalue, and the second inequality follows by the matrix Cauchy-Schwarz inequality in Tripathi (1999, Theorem 1), and the last inequality is due to the fact that $g, m_1, \cdots, m_D$ are all bounded and are supported on

$\mathcal{I}$ and $[-1,1]$, respectively. Combining these results, we have

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

Since $\theta^* = \theta_0 + M_\kappa^{-1}E[P_\kappa r_k]$, we have

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

where the last equality follows by Assumption 2 (8). Therefore, the convergence rate of $\|\hat{\theta}-\theta_0\|$ follows by the triangle inequality.

Next, we prove the convergence rates of $\hat{g}$ and $\hat{m}_d$. Let $\hat{\theta} = (\hat{\mu}, \hat{\theta}^0, \ldots, \hat{\theta}^D)$, where $\hat{\theta}^0$ is the vector of estimated coefficients corresponding to $P_{\kappa,0}$, and $\hat{\theta}^d$ is the vector of estimated coefficients corresponding to $P_{\kappa,d}$ for $d=1,\ldots,D$. Note $\sup_{x^*\in\mathcal{I}}\|P_{\kappa,0}(x^*)\| \leq \sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D}\|P_\kappa(x^*,z)\|$, $\sup_{z_d\in[-1,1]}\|P_{\kappa,d}(z_d)\| \leq \sup_{(x^*,z)\in\mathcal{I}\times[-1,1]^D}\|P_\kappa(x^*,z)\|$, $\|\hat{\theta}^0-\theta_0^0\| \leq \|\hat{\theta}-\theta_0\|$, and $\|\hat{\theta}^d-\theta_0^d\| \leq \|\hat{\theta}-\theta_0\|$ for $d=1,\ldots,D$. Then the convergence rate of $\hat{g}$ is given by

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

where the last inequality is obtained using the Cauchy-Schwartz inequality and Assumption 2 (8), and the last equality follows by Assumption 2 (7) and Theorem 2. Similarly, the uniform convergence rate of $\hat{m}_d$ for $d=1,\ldots,D$ follows by

$$\begin{aligned} \sup_{z_d\in[-1,1]}|\hat{m}_d(z_d)-m_d(z_d)| &\leq \sup_{z_d\in[-1,1]}|P_{\kappa,d}(z_d)'(\hat{\theta}^d-\theta_0^d)| + \sup_{z_d\in[-1,1]}|r_{\kappa,m_d}(z_d)| \\ &\leq \sup_{z_d\in[-1,1]}\|P_{\kappa,d}(z_d)\|\cdot\|\hat{\theta}^d-\theta_0^d\| + O(\kappa^{-2}) \\ &= \begin{cases} O_p\left(\kappa^{\frac{3}{2}} n^{\zeta+\frac{\zeta}{2\beta}-\frac{1}{2}} + \kappa n^{-\frac{\alpha\zeta}{\beta}} + \kappa^{-\frac{3}{2}}\right) & \text{under Assumption 3,} \\ O_p\left(\kappa(\log n)^{-\frac{\alpha}{\gamma}} + \kappa^{-\frac{3}{2}}\right) & \text{under Assumption 4,} \end{cases} \end{aligned}$$

where the last inequality is obtained by the Cauchy-Schwartz inequality and Assumption 2 (8), and the last equality follows by Assumption 2 (7) and Theorem 2. ■

Appendix C. 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

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

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

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

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

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

If $\eta > 0$, we have

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

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

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

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

Step 2: Show

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

For the numerator, we note

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

and

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

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

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

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

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

Thus, (C.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

$$
\frac{\sum_{j=1}^{n}G_{k,n,j}-nE[G_{k,n,1}]}{\sqrt{nVar[G_{1,n,1}]}}\xrightarrow{p}0, \qquad \text{(C.11)}
$$

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

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

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

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

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 (C.12), by Lemma 7, we have $E|\mathbb{K}_h(x^*-X)|^2 = O(h^{-(2\beta+1)})$ and thus

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

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

Combining (C.1), (C.6), and (C.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 $Var[\tilde{g}(x^*)] = \frac{1}{n}Var\big[\sum_{k=1}^4 G_{k,n,1}\big]$. 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 (C.8), (C.10), and (C.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 (C.2), by (C.3), parallel to (C.4), for $\eta > 0$, we have

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

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 (C.15), we have

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

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

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

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

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

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

To verify Lyapunov's condition (C.2) and to check that the first-stage estimation error is negligible as in (C.6), besides (C.15), we also need the parallel result to (C.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 (C.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 the 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 (C.2) and (C.6) hold true, and the conclusion follows. ■

Appendix D. 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 $a^d = \int_{x^*\in\mathcal{I}} f_{X^*,Z_d}(x^*,z_d)dx^*\big(\int K(w)dw\big)^2$. 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

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

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

Step 1: Show

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

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

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

For the first term, we have

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

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

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

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, (D.1) holds true if $nh^4\to\infty$ as $n\to\infty$.

Step 2: Show

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

For the numerator, we have

$$\begin{aligned}E|G^d_{2,n,1}|^2 \;\leq\;& \frac{2}{a^{2d}}E\left|\int_{x^*\in\mathcal{I}}\mathbb{K}_h(x^*-X)dx^*K_h(z_d-Z_d)\big[\mu+\sum_{d'\neq d}m_{d'}(Z_{d'})-\hat{\mu}-\sum_{d'\neq d}\hat{m}_{d'}(Z_{d'})\big]\right|^2\\ &+\frac{2}{a^{2d}}E\left|\int_{x^*\in\mathcal{I}}\mathbb{K}_h(x^*-X)\{\hat{g}(x^*)-g(x^*)\}dx^*K_h(z_d-Z_d)\right|^2.\end{aligned} \tag{D.3}$$

For the first term, we have

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

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

$$\begin{aligned}
a^d E[G_{1,n,1}^d] &= \int_{x^*\in\mathcal{I}} \frac{1}{2\pi}\int e^{-\mathrm{i}tx^*}K^{\mathrm{ft}}(th)\left\{E[\{m_d(Z_d)+g(X^*)\}K_h(z_d-Z_d)|X^*]f_{X^*}\right\}^{\mathrm{ft}}(t)dtdx^* \\
&\quad -\int_{x^*\in\mathcal{I}} \frac{m_d(z_d)+g(x^*)}{2\pi}\int e^{-\mathrm{i}tx^*}K^{\mathrm{ft}}(th)\left\{E[K_h(z_d-Z_d)|X^*]f_{X^*}\right\}^{\mathrm{ft}}(t)dtdx^* \\
&= \int_{x^*\in\mathcal{I}} E[\{m_d(Z_d)+g(X^*)\}K_h(x^*-X^*)K_h(z_d-Z_d)]dx^* \\
&\quad -\int_{x^*\in\mathcal{I}} \{m_d(z_d)+g(x^*)\}E[K_h(x^*-X^*)K_h(z_d-Z_d)]dx^* = O(h^2),
\end{aligned} \tag{D.4}$$

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

$$\frac{1}{Var[G_{1,n,1}^d]} = O(h^{2\beta}). \tag{D.5}$$

Thus, (D.2) 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

$$\mathbb{A}_n^d - a^d = o_p(1). \tag{D.6}$$

To see this, we note

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

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

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

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

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

which implies (D.6) holds true if $h \to 0$ and $nh^{2\beta} \to \infty$ as $n \to \infty$.

Combining (D.1), (D.2), and (D.6), by a similar argument as in the proof of Theorem 3, we have

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

where $\text{Bias}\{\tilde{m}_d(z_d)\} = E[G_{n,1}^d]$.

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

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

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, (D.1) and (D.5) hold true.

Also, by Lemma 10, we have

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

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

Appendix E. Lemmas

For $\zeta > 0$, let $G_{\epsilon,n,\zeta} = \{t \in \mathbb{R} : |f_\epsilon^{\text{ft}}(t)| < n^{-\zeta}\}$ be the region over which the ridge regularization is implemented, and $G_{\epsilon,n,\zeta}^c = \mathbb{R}\backslash G_{\epsilon,n,\zeta}$. 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) $\langle f_1, f_2\rangle = \frac{1}{2\pi}\langle f_1^{\text{ft}}, f_2^{\text{ft}}\rangle$,

(2) $\left(\int f_1(w-w')f_2(w')dw'\right)^{\text{ft}}(t) = f_1^{\text{ft}}(t)f_2^{\text{ft}}(t)$,

(3) $\left(f_1f_2\right)^{\text{ft}}(t) = \frac{1}{2\pi}\int f_1^{\text{ft}}(t-s)f_2^{\text{ft}}(s)ds$,

(4) $f^{\text{ft}}(t-s) = \{f(w)e^{-isw}\}^{\text{ft}}(t)$,

(5) $f^{\text{ft}}(ct) = \left[f(\cdot/c)/c\right]^{\text{ft}}(t)$.

Proof: Lemma 1 (1) is known as Plancherel's isometry and its proof can be found in Meister (2009, Theorem A.4). One of its useful special cases is when $f_1 = f_2 = f$, which gives Parseval's identity, $\|f\|_2^2 = \frac{1}{2\pi}\|f^{\text{ft}}\|_2^2$. Lemma 1 (2) is known as the convolution theorem and its proof can be found in Meister (2009, Theorem A.5). Lemma 1 (3) can be understood as the convolution theorem with respect to the inverse Fourier transform, which will be used in the following discussion, and its proof is attached as follows. Lemma 1 (4) immediately follows by the definition of the Fourier transform. Lemma 1 (5) is known as the linear stretching property of the Fourier transform, and its proof is in Meister (2009, Lemma A.1 (e)).

We now prove Lemma 1 (3). Let $\delta(w)$ be the Dirac delta function. Then we have

$$\frac{1}{2\pi}\int f_1^{\text{ft}}(t-s)f_2^{\text{ft}}(s)ds = \frac{1}{2\pi}\int_s\int_w f_1(w)e^{\mathrm{i}(t-s)w}dw\int_{w'} f_2(w')e^{\mathrm{i}sw'}dw'ds$$
$$= \int_w f_1(w)e^{\mathrm{i}tw}\int_{w'}\left\{\frac{1}{2\pi}\int_s e^{\mathrm{i}s(w'-w)}ds\right\}f_2(w')dw'$$
$$= \int_w f_1(w)e^{\mathrm{i}tw}\int_{w'}\delta(w'-w)f_2(w')dw' = \int f_1(w)f_2(w)e^{\mathrm{i}tw}dw,$$

where the third equality follows by $\delta(w) = \frac{1}{2\pi}\int e^{\mathrm{i}tw}dt$ and the last equality follows by the property of the Dirac delta function, that is $\int\delta(w'-w)f(w')dw' = f(w)$. ■

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

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

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

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

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

Proof of Lemma 2 (1): If f_ϵ is ordinary smooth of order β, $c_{\text{os},0}(1+|t|)^{-\beta} < |f_\epsilon^{\text{ft}}(t)|$ for $t\in\mathbb{R}$, and it follows $(1+|t|)^{-\beta} < c_{\text{os},0}^{-1}n^{-\zeta}$ for $t\in G_{\epsilon,n,\zeta}$. Note that Jensen's inequality $(1+|t|) \le \sqrt{2}(1+|t|^2)^{1/2}$ implies $(1+t^2)^{-\alpha} < 2^\alpha(1+|t|)^{-2\alpha}$, and it follows $(1+t^2)^{-\alpha} < 2^\alpha c_{\text{os},0}^{-\frac{2\alpha}{\beta}}n^{-\frac{2\alpha\zeta}{\beta}}$. Also note that $\int_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\text{ft}}(t)|^2(1+t^2)^\alpha dt \le \int|f_{X^*}^{\text{ft}}(t)|^2(1+t^2)^\alpha dt < c_{\text{sob}}$ by $f_{X^*}\in\mathcal{F}_{\alpha,c_{\text{sob}}}$. Then we have

$$\begin{aligned}\int_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\text{ft}}(t)|^2dt &= \int_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\text{ft}}(t)|^2(1+t^2)^\alpha(1+t^2)^{-\alpha}dt \\ &\le 2^\alpha c_{\text{os},0}^{-\frac{2\alpha}{\beta}}n^{-\frac{2\alpha\zeta}{\beta}}\int_{G_{\epsilon,n,\zeta}}|f_{X^*}^{\text{ft}}(t)|^2(1+t^2)^\alpha dt = O\left(n^{-\frac{2\alpha\zeta}{\beta}}\right). \qquad \text{(E.1)}\end{aligned}$$

By a similar argument, using $f_{X^*|Z_d=z_d}\in\mathcal{F}_{\alpha,c_{\text{sob}}}$ and $f_{X^*|Z_d=z_d,Z_{d'}=z_{d'}}\in\mathcal{F}_{\alpha,c_{\text{sob}}}$ for any $z_d, z_{d'}\in[-1,1]$, we obtain the second and third statements.

Proof of Lemma 2 (2): If f_ϵ is supersmooth of order γ, $c_{\text{ss},0}\exp(-\mu|t|^\gamma) < n^{-\zeta}$ for $t \in G_{\epsilon,n,\zeta}$, which implies there exists some constant $C > 0$ such that $(1+t^2)^{-\alpha} \le C(\log n)^{-\frac{2\alpha}{\gamma}}$ for $t \in G_{\epsilon,n,\zeta}$, which follows by

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

Then, similarly to the previous ordinary smooth case, we have

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

By a similar argument, using $f_{X^*|Z_d=z_d} \in \mathcal{F}_{\alpha,c_{\text{sob}}}$ and $f_{X^*|Z_d=z_d,Z_{d'}=z_{d'}} \in \mathcal{F}_{\alpha,c_{\text{sob}}}$ separately for any $z_d, z_{d'} \in [-1,1]$, we have the second and third statements. ■

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

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

$$\int \frac{|f^{\text{ft}}_\epsilon(t)|^{2r+2}}{\{|f^{\text{ft}}_\epsilon(t)| \vee n^{-\zeta}\}^{2r+4}}dt = O\left(n^{\frac{\zeta(2\beta+1)}{\beta}}\right).$$

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

$$\int \frac{|f^{\text{ft}}_\epsilon(t)|^{2r+2}}{\{|f^{\text{ft}}_\epsilon(t)| \vee n^{-\zeta}\}^{2r+4}}dt = O\left(n^{2\zeta(r+2)}\right).$$

Proof of Lemma 3 (1): By the definition of $G_{\epsilon,n,\zeta}$, we have

$$\int \frac{|f^{\text{ft}}_\epsilon(t)|^{2r+2}}{\{|f^{\text{ft}}_\epsilon(t)| \vee n^{-\zeta}\}^{2r+4}}dt = n^{2\zeta(r+2)}\int_{G_{\epsilon,n,\zeta}} |f^{\text{ft}}_\epsilon(t)|^{2r+2}dt + \int_{G^c_{\epsilon,n,\zeta}} \frac{1}{|f^{\text{ft}}_\epsilon(t)|^2}dt. \qquad \text{(E.3)}$$

If f_ϵ is ordinary smooth of order β, $c_{\text{os},0}(1+|t|)^{-\beta} \le |f^{\text{ft}}_\epsilon(t)| \le c_{\text{os},1}(1+|t|)^{-\beta}$ for $t \in \mathbb{R}$. For $t \in G_{\epsilon,n,\zeta}$, we have $c_{\text{os},0}(1+|t|)^{-\beta} \le |f^{\text{ft}}_\epsilon(t)| < n^{-\zeta}$, which implies $(1+|t|)^{-\beta} < c^{-1}_{\text{os},0}n^{-\zeta}$. Thus, there exists some constant $0 < \eta < 2\beta(r+1) - 1$ such that $(1+|t|)^{-2\beta(r+1)+1+\eta} < c_{\text{os},0}^{-\frac{2\beta(r+1)-1-\eta}{\beta}} n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}}$ for $t \in G_{\epsilon,n,\zeta}$ if $\beta > 1/2$. Also note $\int_{G_{\epsilon,n,\zeta}}(1+|t|)^{-1-\eta}dt \to 0$ as $n \to \infty$ because $1 + |t| > c_{\text{os},0}^{\frac{1}{\beta}} n^{\frac{\zeta}{\beta}}$ for $t \in G_{\epsilon,n,\zeta}$ and $\int(1+|t|)^{-1-\eta}dt < \infty$ for any $\eta > 0$. Thus, we have the following result:

$$
\begin{aligned}
\int_{G_{\epsilon,n,\zeta}} |f^{\text{ft}}_\epsilon(t)|^{2r+2}dt &\le c^2_{\text{os},1}\int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-2\beta(r+1)+1+\eta}(1+|t|)^{-1-\eta}dt \\
&\le c_{\text{os},1}c_{\text{os},0}^{-\frac{2\beta(r+1)-1-\eta}{\beta}} n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}} \int_{G_{\epsilon,n,\zeta}} (1+|t|)^{-1-\eta}dt \\
&= O\left(n^{-\frac{\zeta(2\beta(r+1)-1-\eta)}{\beta}}\right). \qquad \text{(E.4)}
\end{aligned}
$$

For $t \in G^c_{\epsilon,n,\zeta}$, $|f^{\text{ft}}_\epsilon(t)|^{-2} \leq n^{2\zeta}$. If f_ϵ is ordinary smooth of order $\beta > 0$, $c_{\text{os},1}(1+|t|)^{-\beta} \geq |f^{\text{ft}}_\epsilon(t)| \geq n^{-\zeta}$ for $t \in G^c_{\epsilon,n,\zeta}$, which implies $|t| < c_{\text{os},1}^{\frac{1}{\beta}} n^{\frac{\zeta}{\beta}}$. Then it follows

$$\int_{G^c_{\epsilon,n,\zeta}} |f^{\text{ft}}_\epsilon(t)|^{-2} dt \leq n^{2\zeta} \int_{G^c_{\epsilon,n,\zeta}} dt \leq 2c_{\text{os},1}^{\frac{1}{\beta}} n^{\frac{\zeta(2\beta+1)}{\beta}} = O\left(n^{\frac{\zeta(2\beta+1)}{\beta}}\right). \tag{E.5}$$

Combining (E.3), (E.4), and (E.5), the conclusion follows.

Proof of Lemma 3 (2): For $t \in G^c_{\epsilon,n,\zeta}$, $|f^{\text{ft}}_\epsilon(t)| \geq n^{-\zeta}$, which implies $|f^{\text{ft}}_\epsilon(t)|^{-2r-4} \leq n^{2\zeta(r+2)}$. Then, we have

$$\int \frac{|f^{\text{ft}}_\epsilon(t)|^{2r+2}}{\{|f^{\text{ft}}_\epsilon(t)| \vee n^{-\zeta}\}^{2r+4}} dt \leq n^{2\zeta(r+2)} \int |f^{\text{ft}}_\epsilon(t)|^{2r+2} dt. \tag{E.6}$$

If f_ϵ is supersmooth of order $\gamma > 0$, we have

$$\int |f^{\text{ft}}_\epsilon(t)|^{2r+2} dt \leq 2c_{\text{ss},1}^{2r+2} \int_0^{+\infty} \exp\big(-(2r+2)\mu|t|^\gamma\big) dt, \tag{E.7}$$

where the inequality follows by the smoothness of f_ϵ and the symmetry of the integration. Note $t^2 \exp(-(2r+2)\mu|t|^\gamma) \to 0$ as $t \to \infty$. Due to the strict monotonicity of t^2 and $\exp((2r+2)\mu|t|^\gamma)$, there exists a constant δ such that $\exp((2r+2)\mu|t|^\gamma) > t^2$ for any $t > \delta$. Then, we have

$$\begin{aligned} \int_0^{+\infty} \exp(-(2r+2)\mu|t|^\gamma) dt &= \int_0^\delta + \int_\delta^{+\infty} \exp(-(2r+2)\mu|t|^\gamma) dt \\ &\leq \delta + \int_\delta^{+\infty} t^{-2} dt = \delta + \delta^{-1} < \infty. \end{aligned} \tag{E.8}$$

Combining (E.6), (E.7), and (E.8), the conclusion follows. ■

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

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 of Lemma 4: 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

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

where the second equality follows by Assumption 2 (1).

For the bias term B, Lemma 1 (1) and the law of iterated expectations imply

$$E[p(X^*)Q] = \langle E[Q|X^*]f_{X^*}, p\rangle = \frac{1}{2\pi}\int E[Qe^{\mathrm{i}tX^*}]p^{\mathrm{ft}}(-t)dt,$$

and

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

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) has the form $(p_0,1)$, $(p_k,1)$, $(p_kp_l,1)$, $(p_0,q_k(Z_d))$, $(p_k,q_l(Z_d))$, $(p_0,q_k(Z_d)q_l(Z_d))$, 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 the proof of B_3. Note

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

where $\varrho_n^B = n^{-\frac{2\alpha\zeta}{\beta}}$ under Assumption 3 and $(\log n)^{-\frac{2\alpha}{\gamma}}$ under Assumption 4, the second equality follows by Lemma 1 (2), the third equality follows by the change of variables $(u,v)=(t+s,s)$, the last equality follows by Lemma 2, and the last inequality follows by Lemma 1 (1), the orthonormality of $\{p_l\}_{l=1}^{\kappa}$, and the fact that

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

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 equality follows by Lemma 1 (1) and (4), the first inequality follows by the orthonormality of $\{p_k\}_{k=1}^{\kappa}$, the second inequality follows by $|e^{-\mathrm{i}vw}|=1$ and Lemma 1 (1), and the last equality follows by the definition of $G_{\epsilon,n,\zeta}$. By similar arguments, we have $B_1, B_2 = O(\varrho_n^B)$.

Since the proof is similar for B_4 and B_5, we focus on the proof of B_5. Note

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

where the first inequality follows by

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

where $h_{2,l,d}$ denotes the Fourier inverse of $\left(\frac{|f_\epsilon^{\mathrm{ft}}(t)|^{r+2}}{\{|f_\epsilon^{\mathrm{ft}}(t)|\vee n^{-\zeta}\}^{r+2}}-1\right)E[q_l(Z_d)e^{\mathrm{i}tX^*}]$, and

$$
\begin{aligned}
E[q_l(Z_d)e^{\mathrm{i}tX^*}] &= \int_{z_d}\int_{x^*}e^{\mathrm{i}tx^*}q_l(z_d)f_{X^*,Z_d}(x^*,z_d)dx^*dz_d\\
= &\int_{z_d}\left\{\int_{x^*}e^{\mathrm{i}tx^*}f_{X^*|Z_d=z_d}(x^*)dx^*\right\}f_{Z_d}(z_d)q_l(z_d)dz_d = \int_{z_d}f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)f_{Z_d}(z_d)q_l(z_d)dz_d,
\end{aligned}
$$

the second inequality follows by the orthonormality of $\{q_l\}_{l=1}^{\kappa}$, the third inequality follows by the fact that f_{Z_d} is supported on $[-1,1]$ and $\max_{d\in\{1,\ldots,D\}}\sup_{z_d\in[-1,1]}|f_{Z_d}(z_d)|\leq c_{z,1}$, and the last equality follows by Lemma 2. Similarly, we have $B_4 = O(\varrho_n^B)$.

For B_6, we have

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

where the first inequality follows by the Cauchy-Schwarz inequality and

$$\begin{aligned}
E[q_k(Z_d)q_l(Z_d)e^{\mathrm{i}tX^*}] &= \int_{z_d}\int_{x^*}e^{\mathrm{i}tx^*}q_k(z_d)q_l(z_d)f_{X^*,Z_d}(x^*,z_d)dx^*dz_d \\
= \int_{z_d}\left\{\int_{x^*}e^{\mathrm{i}tx^*}f_{X^*|Z_d=z_d}(x^*)dx^*\right\}f_{Z_d}(z_d)dz_d &= \int_{z_d}f_{X^*|Z_d=z_d}^{\mathrm{ft}}(t)f_{Z_d}(z_d)q_k(z_d)q_l(z_d)dz_d,
\end{aligned}$$

the second inequality follows by the orthonormality of $\{q_l\}_{l=1}^{\kappa}$, the third inequality follows by $\max_{d\in\{1,\cdots,D\}}\sup_{z_d\in[-1,1]}|f_{Z_d}(z_d)|\leq c_Z$, the second equality follows by the unity of q_k, and the last equality follows by Lemma 2. By a similar argument, we have $B_7 = O(\varrho_n^B)$.

Combining these results, we obtain

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

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

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

where $V_1,\ldots,V_7$ are summations of non-central second moment terms with (p,Q) in the forms of $(p_0,1)$, $(p_k,1)$, $(p_kp_l,1)$, $(p_0,q_k(Z_d))$, $(p_k,q_l(Z_d))$, $(p_0,q_k(Z_d)q_l(Z_d))$, 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 the proof of V_3. Note

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

where $\varrho_n^V = n^{2\zeta+\frac{\zeta}{\beta}-1}$ under Assumption 3 and $n^{2\zeta(r+2)-1}$ under Assumption 4, the second equality follows by Lemma 1 (2), the third equality follows by the change of variables $(u,v) = (t+s,s)$, the last equality follows by Lemma 3, and the last inequality follows by Lemma 1 (1), the unity of $\{p_l\}_{l=1}^{\kappa}$, and the following fact

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

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

Since the proof is similar to other terms, we focus on the proof of V_5:

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

By applying similar arguments, we obtain $V_4 = O(\kappa\varrho_n^V)$ and $V_6, V_7 = O(\kappa^2\varrho_n^V)$.

Combining these results,

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

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. ■

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

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

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

Proof of Lemma 5: Since $\lim_{|t|\to\infty} |t|^\beta |f_\epsilon^{\mathrm{ft}}(t)| = c_\epsilon$, there exists a constant c_F such that $|t|^\beta |f_\epsilon^{\mathrm{ft}}(t)| > c_\epsilon/2$ 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

$$\begin{aligned} h^\beta \frac{|K^{\mathrm{ft}}(s)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|} &\leq h^\beta \frac{\max_{|s|\leq c_F h} |K^{\mathrm{ft}}(s)|}{\min_{|s|\leq c_F} |f_\epsilon^{\mathrm{ft}}(s)|} \mathbb{1}\{|s| \leq c_F h\} + \frac{|K^{\mathrm{ft}}(s)||s|^\beta}{(|s|/h)^\beta |f_\epsilon^{\mathrm{ft}}(s/h)|} \mathbb{1}\{|s| > c_F h\} \\ &\leq c_1 c_{\mathrm{os},0}^{-1} (1 + c_F)^\beta \|K^{\mathrm{ft}}\|_\infty \mathbb{1}\{|s| \leq c_2\} + \frac{2|K^{\mathrm{ft}}(s)||s|^\beta}{c_\epsilon} \equiv \psi(s), \end{aligned} \tag{E.9}$$

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

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 of Lemma 6: 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 $\int |K^*(x)| dx < \infty$. ■

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

$$\frac{h^{2\beta}}{4\pi^2}\left|\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}e^{-\mathrm{i}sx}ds\right|^2 \to \frac{1}{4\pi^2 c_\epsilon^2}\left|\int K^{\mathrm{ft}}(s)|s|^\beta e^{-\mathrm{i}sx}ds\right|^2. \tag{E.10}$$

Moreover, using integration by parts, we have

$$\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}e^{-\mathrm{i}sx}ds = \frac{1}{\mathrm{i}x}\int \frac{K^{\mathrm{ft}'}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}e^{-\mathrm{i}sx}ds + \frac{1}{\mathrm{i}xh}\int \frac{K^{\mathrm{ft}}(s)f_\epsilon^{\mathrm{ft}'}(s/h)}{f_\epsilon^{\mathrm{ft}^2}(s/h)}e^{-\mathrm{i}sx}ds. \tag{E.11}$$

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

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

and

$$\begin{aligned}
&\left|\frac{1}{\mathrm{i}xh}\int \frac{K^{\mathrm{ft}}(s)f_\epsilon^{\mathrm{ft}'}(s/h)}{f_\epsilon^{\mathrm{ft}^2}(s/h)}ds\right| \le \frac{h^{-1}}{|x|}\int \frac{|K^{\mathrm{ft}}(s)||f_\epsilon^{\mathrm{ft}'}(s/h)|}{|f_\epsilon^{\mathrm{ft}}(s/h)|^2}ds\\
\le \;& \frac{1}{|x|}\left(\frac{2c_F\max_{|s|\le c_F h}|K^{\mathrm{ft}}(s)|\max_{|s|\le c_F}|f_\epsilon^{\mathrm{ft}'}(s)|}{\min_{|s|\le c_F}|f_\epsilon^{\mathrm{ft}}(s)|^2}\right) + \frac{h^{-\beta}}{|x|}\int_{|s|>c_F h}\frac{|K^{\mathrm{ft}}(s)||s|^{\beta-1}(|s|/h)^{\beta+1}|f_\epsilon^{\mathrm{ft}'}(s/h)|}{(|s|/h)^{2\beta}|f_\epsilon^{\mathrm{ft}}(s/h)|^2}ds\\
\le \;& \frac{h}{|x|}2c_F c_{\mathrm{os},0}^{-2}(1+c_F)^{2\beta}\|K^{\mathrm{ft}}\|_\infty\|f_\epsilon^{\mathrm{ft}'}\|_\infty + \frac{h^{-\beta}}{|x|}\left(\frac{5\beta}{c_\epsilon}\right)\int |K^{\mathrm{ft}}(s)||s|^{\beta-1}ds = O(h^{-\beta}|x|^{-1}).
\end{aligned}\tag{E.13}$$

Thus, Lemma 5, (E.11), (E.12), and (E.13) imply there are a pair of constants $c_1, c_2 > 0$ such that

$$\sup_n h^{2\beta}\left|\int \frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/h)}e^{-\mathrm{i}sx}ds\right|^2 \le \min\{c_1, c_2|x|^{-2}\}. \tag{E.14}$$

Therefore, the conclusion follows by

$$
\begin{aligned}
&\lim_{n\to\infty} h^{2\beta+1}\int \frac{1}{4\pi^2}\left|\int_t \frac{K^{\mathrm{ft}}(th)}{f^{\mathrm{ft}}_\epsilon(t)}e^{-\mathrm{i}t(x^*-x)}dt\right|^2 f(x)dx \\
=\;& \lim_{n\to\infty}\int_x \frac{h^{2\beta-1}}{4\pi^2}\left|\int_s \frac{K^{\mathrm{ft}}(s)}{f^{\mathrm{ft}}_\epsilon(s/h)}e^{-\frac{\mathrm{i}s(x^*-x)}{h}}ds\right|^2 f(x)dx \\
=\;& \frac{f(x^*)}{c_\epsilon^2}\int_x\left|\frac{1}{2\pi}\int_s K^{\mathrm{ft}}(s)|s|^\beta e^{-\mathrm{i}sx}ds\right|^2 dx = \frac{f(x^*)}{2\pi c_\epsilon^2}\int |K^{\mathrm{ft}}(s)|^2|s|^{2\beta}ds, \qquad \text{(E.15)}
\end{aligned}
$$

where the first equality follows by the change of variables $s = th$, the second equality follows by Lemma 6 with $K_n(x) = \frac{h^{2\beta}}{4\pi^2}\left|\int \frac{K^{\mathrm{ft}}(s)}{f^{\mathrm{ft}}_\epsilon(s/h)}e^{-\mathrm{i}sx}ds\right|^2$ 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*

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

Proof of Lemma 8: The first statement follows by

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

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^{ft} is supported on $[-1, 1]$, and the last equality uses $\|K^{\mathrm{ft}}\|_\infty < \infty$.

The second statement follows by

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

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^{ft} is supported on $[-1, 1]$, and the last equality uses $\|K^{\mathrm{ft}}\|_\infty < \infty$. ■

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* $\sup_s\left|g^{\mathrm{ft}}(-\frac{s}{h})\frac{s}{h^2}\right| \to 0$ *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 of Lemma 9: 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

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

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

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

and M is a constant such that $|f_\epsilon^{\mathrm{ft}}(t)||t|^\beta > c_\epsilon/2$ and $|f_\epsilon^{\mathrm{ft}\prime}(t)||t|^{\beta+1} < 5\beta c_\epsilon/4$ 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,

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

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

$$
\begin{aligned}
J_{2,1} &\to \frac{f_{X,Z_d}(b_1,z_d)}{2\pi c_\epsilon^2}\int |K^{\mathrm{ft}}(s)|^2|s|^{2\beta-2}ds\int K^2(v)dv,\\
J_{2,2} &\to \frac{f_{X,Z_d}(b_2,z_d)}{2\pi c_\epsilon^2}\int |K^{\mathrm{ft}}(s)|^2|s|^{2\beta-2}ds\int K^2(v)dv.
\end{aligned} \tag{E.16}
$$

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 $\int K^*(u,v)dudv<\infty$, (E.16) 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 (E.16) is obtained.

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

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

Thus, by $J_{2,1} + J_{2,2} - |J_{2,3}| \leq J_2 \leq J_{2,1} + J_{2,2} + |J_{2,3}|$, 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 $c_2 \geq c_1 > 0$ such that $c_1 \leq J_2 \leq c_2$ 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

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

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

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

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*

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

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

for all n large enough.

Proof of Lemma 10: Let $\mathcal{I} = [b_1, b_2]$. For the first statement, we have

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

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 $\|g^{\mathrm{ft}}\|_\infty < \infty$. ■

References

[1] Adusumilli, K., Kurisu, D., Otsu, T. and Y. J. Whang (2020) Inference on distribution functions under measurement error, *Journal of Econometrics*, 215, 131-164.

[2] 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.

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

[4] 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.

[5] 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.

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

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

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

[9] 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.

[10] Comte, F. and J. Kappus (2015) Density deconvolution from repeated measurements without symmetry assumption on the errors, *Journal of Multivariate Analysis*, 140, 31-46.

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

[12] 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.

[13] 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.

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

[15] 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.

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

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

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

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

[20] 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.

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

[22] 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.

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

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

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

[26] 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.

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

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

[29] 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.

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

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

[32] 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.

[33] Kato, K. and Y. Sasaki (2018) Uniform confidence bands in deconvolution with unknown error distribution, *Journal of Econometrics*, 207, 129-161.

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

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

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

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

[38] Li, T. and Q. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators, *Journal of Multivariate Analysis*, 65, 139-165.

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

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

[41] 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.

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

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

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

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

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

[47] 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.

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

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

[50] 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.

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

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

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

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

[55] 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.

[56] 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.

E-mail address: haod@mail.smu.edu

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

E-mail address: t.otsu@lse.ac.uk

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

E-mail address: lntaylor@econ.au.dk