ESTIMATION OF VARYING COEFFICIENT MODELS WITH MEASUREMENT ERROR

HAO DONG, TAISUKE OTSU, AND LUKE TAYLOR

Abstract. We propose a semiparametric estimator for varying coefficient models when the regressors in the nonparametric components are measured with error. Varying coefficient models are an extension of other popular semiparametric models, including partially linear and nonparametric additive models, and deliver an attractive solution to the curse-of-dimensionality. We use deconvolution kernel estimation in a two-step procedure and show that the estimator is consistent and asymptotically normally distributed. We do not assume that we know the distribution of the measurement error *a priori*. Instead, we suppose we have access to a repeated measurement of the noisy regressor and present results using the approach of Delaigle, Hall and Meister (2008) and, for cases when the measurement error may be asymmetric, the approach of Li and Vuong (1998) based on Kotlarski's (1967) identity. We show that the convergence rate of the estimator is significantly reduced when the distribution of the measurement error is assumed unknown and possibly asymmetric. We study the small sample behavior of our estimator in a simulation study and apply it to a real dataset. In particular, we consider the role of cognitive ability in augmenting the effect of risk preferences on earnings.

1. Introduction

Varying coefficient models, introduced by Hastie and Tibshirani (1993), represent a very general class of semiparametric specification. In its canonical form, the varying coefficient model is given by

$$Y = \beta_0(Z) + X_1\beta_1(Z) + X_2\beta_2(Z) + \cdots + X_k\beta_k(Z) + U, \qquad E[U|X,Z] = 0,$$

where Y is a scalar dependent variable, $X = (X_1, \ldots, X_k)'$ and Z are covariates, $(\beta_0(\cdot), \ldots, \beta_k(\cdot))$ are unknown functions of Z, and U is an error term. Note that X and Z need not necessarily be mutually exclusive sets of variables, and may even coincide. This specification allows the effect of each X_j on Y to depend on Z in a nonparametric manner. As well as nesting nonparametric additive models (Hastie and Tibshirani, 1993), the varying coefficient model is also a generalisation of the partially linear model (Robinson, 1988).

In this paper, we extend the varying coefficient model to allow for Z to be imperfectly measured. Contaminated data exists in almost all areas of the natural and social sciences and is particularly common in economics. The prevalence of the problem is demonstrated by the large - and continually growing - literature aimed at correcting this issue. There are many possible causes for such noisy data, for example: an imperfect measurement instrument, which is particularly widespread in macroeconomics, where variables such as unemployment and inflation can

The authors would like to thank anonymous referees and an associate editor for helpful comments, and acknowledge financial supports from the SMU Dedman College Research Fund (12-412268) (Dong), the ERC Consolidator Grant (SNP 615882) (Otsu), and the Aarhus University Research Fund (AUFF-26852) (Taylor).

only proxy for the truth; reporting errors in survey data; expectations of future variables, such as expected future consumption or inflation; and variables whose very definitions are imprecise, for example, ability or non-cognitive skills. More generally, measurement error is present whenever the variables of the theoretical model do not exactly match the variables in the data. Such error-ridden variables are a well-known source of inconsistency in many estimators, and varying coefficient models are no exception.

We propose a deconvolution based estimator for the varying coefficient model when the covariate in the nonparametric component is contaminated with classical measurement error. We show that the estimator is consistent and asymptotically normally distributed under a range of potential assumptions regarding the measurement error. In particular, we provide results for both ordinary smooth and supersmooth error, and when the error distribution is assumed known or when it must be estimated using auxiliary data. Moreover, in contrast to much of the previous literature, we also prove the asymptotic properties of the estimator when the error distribution is estimated but is not assumed symmetric.

Although seemingly innocuous, allowing the error distribution to be asymmetric poses considerable technical challenges (see, for example, Li and Vuong, 1998, Bonhomme and Robin, 2010, and Kurisu and Otsu, 2020). Indeed, Delaigle, Hall and Meister (2008) showed that nonparametric deconvolution estimators with an unknown - but symmetric - measurement error density can obtain the same convergence rate as the corresponding estimator with a known error density. In contrast, when the symmetry assumption is relaxed, the noise from estimating the error characteristic function dominates the asymptotic properties of the final estimator and results in slower convergence rates than its known (or symmetric) estimator counterparts.

The plethora of recent papers studying the theoretical properties of varying coefficient models highlights their growing popularity (see, for example, Ma and Song, 2015, He, Lian, Ma and Huang, 2018, and Yao, Zhang and Kumbhakar, 2019). However, these models are not only of theoretical interest. They have been put to great use in many applied settings (see, for example, Mamuneas, Savvides and Stengos, 2006, Heshmati, Kumbhakar and Sun, 2014, and Feng, Gao, Peng and Zhang, 2017). Within empirical work, one of the biggest appeals of varying coefficient models is their similarity to conventional linear regression models, which facilitates a straightforward interpretation of the estimation results.

The ability of varying coefficient models to mitigate the 'curse-of-dimensionality' is another source of their popularity in applied work. Typically, a single covariate is used in the nonparametric component; for example, one may be interested in how the return to education changes over the lifecycle of an individual or on their ability, the latter being a prime example of a mismeasured variable. In this case, estimators of these effects converge at the rate $\sqrt{na_n}$, where a_n denotes the bandwidth parameter. This is in contrast to a fully nonparametric model where the convergence rate is $\sqrt{na_n^d}$, where d is the dimension of the full set of regressors. Moreover, in the presence of measurement error, the curse-of-dimensionality is, in general, exacerbated. Hence, the benefits of using varying coefficient models are increased when working with contaminated data.

As a result, varying coefficient models have been extended to allow for measurement error in the covariates in several ways. You and Chen (2006) considered the setting where one of the coefficients is constant, and its associated covariate is contaminated with error from a known distribution. Zhou and Liang (2009) extended this model to allow for an unknown error distribution but where auxiliary information is available to estimate this density. In both cases, $\sqrt{n}$-convergence is obtained for this finite-dimensional parameter using profile least squares estimation. Finally, Li and Greene (2008) suppose that the error-prone covariate has a varying coefficient which depends on correctly measured regressors. They use locally corrected score equations to estimate the nonparametric functions and show that the convergence rate is not affected by the measurement error.

We depart from the previous literature by considering measurement error in the nonparametric component. This poses very different problems to those encountered in the aforementioned research. In particular, we require deconvolution techniques to recover the distribution of the latent covariates needed to estimate the smooth coefficient functions. We show that, in contrast to settings where the mismeasured covariates enter the model linearly, the measurement error impacts the rate of convergence of the estimator. Furthermore, the rate of convergence depends sensitively on the degree of smoothness of the measurement error density.

More generally, deconvolution methods were first applied to measurement error problems for density estimation by Carroll and Hall (1988) and Stefanski and Carroll (1990). Following this seminal work, myriad extensions have been developed. Much of this research has focused on relaxing the assumption that the density of the measurement error is known and must be estimated from auxiliary data; for symmetrically distributed error, see, for example, Horowitz and Markatou (1996) and Delaigle, Hall and Meister (2008), among many others, and for non-symmetric error, see, for example, Li and Vuong (1998) and Bonhomme and Robin (2010). With respect to regression estimation, Fan and Truong (1993) proposed a deconvolution method based on the Nadaraya-Watson estimator. A related strand of research considers adaptive estimation of regression functions. Here, estimation methods exist for when the error distribution is assumed known (Comte and Taupin, 2007), when it is unknown but symmetric (Kappus and Mabon, 2014), and when it is unknown and non-symmetric (Comte and Kappus, 2015). Also, as an alternative to kernel methods, wavelets could potentially be used in our setting, see, for example, Pensky and Vidakovic, (1999) and Fan and Koo (2002). Schennach (2016) gives a recent review of the vast measurement error literature.

The paper proceeds as follows. In Section 2, we outline the model setting, discuss our estimator when the density of the measurement error is assumed to be known, and present the asymptotic properties of the estimator. In Section 3, we relax the assumption of a known error distribution and detail the resulting asymptotic properties. Section 4 presents the small sample properties of our estimator in a simulation study. Section 5 considers an empirical application of our estimator, and Section 6 concludes. In the supplementary material, we consider the case of possibly asymmetric measurement error densities and derive the asymptotic properties of the estimator.

2. CASE OF KNOWN MEASUREMENT ERROR DISTRIBUTION

2.1. **Setup and estimator.** Consider the varying coefficient model

$$Y = X'\beta(W^*) + U, \qquad E[U|X, W^*] = 0, \tag{2.1}$$

where $X = (X_1, \ldots, X_k)' \in \mathbb{R}^k$ is a vector of observable covariates, $W^* \in \mathbb{R}$ is an error-free covariate, and $\beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_k(\cdot))'$ is a vector of unknown functions. In this paper, we concentrate on the case where X and W^* are non-overlapping; this negates the need for backfitting algorithms in the estimation procedure. We wish to estimate $\beta(w^*)$ at a given point $w^* \in \mathbb{R}$ using an i.i.d. sample of (Y, X, W), where W is a noisy measurement of W^* generated by

$$W = W^* + \epsilon, \tag{2.2}$$

and ϵ is a measurement error. In this paper, we assume the measurement error is classical; that is, ϵ is independent of (Y, X, W^*). However, we do not require full independence between ϵ and W^*, we only need $f_W^{\text{ft}}(t) = f_{W^*}^{\text{ft}}(t) f_\epsilon^{\text{ft}}(t)$ for all $t \in \mathbb{R}$, where $f^{\text{ft}}(t) = \int e^{\mathrm{i}tx} f(x)dx$ denotes the Fourier transform of a function f with $\mathrm{i} = \sqrt{-1}$, and f_A denotes the density function of a random variable A. As argued in Schennach (2019), this assumption is only as strong as a conditional mean restriction.

Our estimation strategy proceeds as follows. Let $M_{XX}(w^*) = E[XX'|W^* = w^*]$ and $M_{XY}(w^*) = E[XY|W^* = w^*]$. $M_{XX}(w^*)$ is assumed to be invertible (see, Assumption M below). By pre-multiplying (2.1) by X and taking the conditional expectation, the object of interest $\beta(w^*)$ can be written as

$$\beta(w^*) = M_{XX}(w^*)^{-1} M_{XY}(w^*). \tag{2.3}$$

The conditional moments on the right hand side can be estimated by using deconvolution techniques. In particular, we estimate $M_{XX}(w^*)$ and $M_{XY}(w^*)$ by

$$\hat{M}_{XX}(w^*) = \frac{\sum_{j=1}^n X_j X_j' \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right)}{\sum_{j=1}^n \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right)}, \qquad \hat{M}_{XY}(w^*) = \frac{\sum_{j=1}^n X_j Y_j \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right)}{\sum_{j=1}^n \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right)},$$

respectively, where $\mathbb{K}$ is a deconvolution kernel function defined by

$$\mathbb{K}(x) = \frac{1}{2\pi} \int e^{-\mathrm{i}tx} \frac{K^{\text{ft}}(t)}{f_\epsilon^{\text{ft}}(t/a_n)} dt, \tag{2.4}$$

K:$\mathbb{R} \to \mathbb{R}$ is an ordinary kernel function, and a_n is a bandwidth parameter.

Based on these deconvolution estimators for the conditional moments, $\beta(w^*)$ can be estimated by[1]

$$\hat{\beta}(w^*) = \left[\sum_{j=1}^n X_j X_j' \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right)\right]^{-1} \sum_{j=1}^n X_j Y_j \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right). \tag{2.5}$$

Li *et al.* (2002) considered a similar estimator when there is no measurement error and W^* is directly observed. In this paper, we employ the deconvolution kernel $\mathbb{K}$ to deal with the

[1]Our assumptions below guarantee that $\frac{1}{na_n} \sum_{j=1}^n X_j X_j' \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right) \xrightarrow{p} M_{XX}(w^*)$ (see, proof of Theorem 1 (i)). Thus, the matrix $\frac{1}{na_n} \sum_{j=1}^n X_j X_j' \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right)$ is invertible with probability approaching one.

measurement error in W. The deconvolution kernel $\mathbb{K}$ defined as in (2.4), however, requires researchers to know the measurement error distribution, which is an assumption that can rarely hold in practice. In Section 3, we use the repeated measurements on W^* to relax this known error distribution assumption. Also note that even though we focus on the case where W^* is scalar in both Sections 2 and 3 to keep the notation simple, an extension to the case where W^* is multivariate is discussed in Section 6.

2.2. **Asymptotic properties.** In this section, we focus on the estimator $\hat{\beta}(w^*)$ defined in (2.5) and study its asymptotic properties. As in the majority of the deconvolution literature, to investigate the asymptotic properties of $\hat{\beta}(w^*)$, we consider two separate cases based on the tail behavior of the characteristic function of the measurement error. The first is known as the case of ordinary smooth error, or the ordinary smooth case for short, and is characterized by the characteristic function of the measurement error decaying to zero at some polynomial rate. The second is known as the case of supersmooth error, or the supersmooth case for short, and is defined by an exponentially decaying error characteristic function.

Let $\lambda_{\max}(B)$ and $\lambda_{\min}(B)$ denote the maximum and minimum eigenvalues, respectively, of a matrix B. We impose the following assumptions for both the ordinary smooth and supersmooth cases.

Assumption M.

- **(1):** $\{Y_j, X_j, W_j\}_{j=1}^n$ *is an i.i.d. sample of* (Y, X, W) *satisfying (2.1) and (2.2), and* ϵ *is independent of* (Y, X, W^*).
- **(2):** $\beta(\cdot)$, $f_{W^*}(\cdot)$, *and* $E[X_{k_1}X_{k_2}|W^* = \cdot]$ *for* $k_1, k_2 = 1, \ldots, k$ *have* $p \in \mathbb{N}$ *continuous, bounded, and integrable derivatives.* $E[U^2|X, W^*]$ *and* $E[X_{k_1}^2 X_{k_2}^2|W^*]$ *for* $k_1, k_2 = 1, \ldots, k$ *are bounded. Also* $\lambda_{\min}(E[XX'|W^*]) > 0$ *holds almost everywhere.*
- **(3):** K *satisfies* $\int K(x)dx = 1$, $\int x^p K(x)dx \neq 0$ *for some* $p \in \mathbb{N}$, *and* $\int x^q K(x)dx = 0$ *for all positive integers satisfying* $q < p$. *Also* K^{ft} *is supported on* $[-1, 1]$ *and bounded.*
- **(4):** $E[|U|^{2+\eta}|X, W^*]$ *and* $E[|X_{k_1}|^{2+\eta}|W^*]$ *for* $k_1 = 1, \ldots, k$ *are bounded for some* $\eta > 0$.

Assumption M (1) requires random sampling and classical measurement error. Assumption M (2) constitutes mild assumptions on the smoothness and boundedness of densities and conditional moments, and the last condition guarantees identification of $\beta(w^*)$. The p-th order continuous differentiability of the functions $\beta(\cdot)$, $f_{W^*}(\cdot)$, and $E[X_{k_1}X_{k_2}|W^* = \cdot]$ are imposed to characterize the order of the bias term, i.e., the component a_n^{2p} in Theorem 1 (i) below. By inspection of the proof of Li *et al.* (2002, Theorem 2.1), we can see that analogous differentiability conditions are required to characterize the order of the bias term even for the error-free case. Assumption M (3) concerns the kernel function K. In particular, we require K to be a p-th order kernel to control the estimation bias. In addition to the high-order property, we also require K^{ft} to be compactly supported for regularisation, which is necessary in the deconvolution problem. Assumption M (4) contains additional assumptions on the boundedness of conditional moments, which are used to apply Lyapunov's central limit theorem for the asymptotic distribution of $\hat{\beta}(w^*)$.

We begin with the ordinary smooth case, for which the following assumptions are also imposed. Let $|\cdot|$ denote the Euclidean norm.

Assumption O.

(1): *There exist positive constants* α *and* c_0^{os} *such that*

$$|f_\epsilon^{\mathrm{ft}}(t)|(1+|t|)^\alpha \geq c_0^{\mathrm{os}} \text{ for all } t \in \mathbb{R}.$$

(2): $a_n \to 0$ *and* $na_n^{1+2\alpha} \to \infty$ *as* $n \to \infty$.

(3): *There exists a positive constant* $c^{\mathrm{os}} \geq c_0^{\mathrm{os}}$ *such that*

$$f_\epsilon^{\mathrm{ft}}(t)|t|^\alpha \to c^{\mathrm{os}} \text{ as } |t| \to \infty.$$

(4): $E[X_{k_1}X_{k_2}U^2|W^* = \cdot]$ *and* $E[X_{k_1}^2 X_{k_2}^2|W^* = \cdot]$ *for* $k_1, k_2 = 1, \ldots, k$ *are continuous.*

(5): $na_n^{1+2\alpha+2p} \to 0$ *as* $n \to \infty$.

Assumption O (1) says that f_ϵ is ordinary smooth of order α. Popular examples of ordinary smooth densities include the Laplace and gamma densities. The traditional ordinary smooth assumption, as in (2.31) of Meister (2009), involves two bounds on $|f_\epsilon^{\mathrm{ft}}(t)|$. Assumption O (1) only imposes the lower bound as this is sufficient to study the upper bound of the risk as in Theorem 1 (i). Assumption O (2) gives conditions on the bandwidth a_n. In particular, $a_n \to 0$ is required for a vanishing bias, and $na_n^{1+2\alpha} \to \infty$ is needed to control the estimation variance. We emphasise that only Assumption O (1)-(2) are needed to derive the convergence rate of $\hat{\beta}(w^*)$, and Assumption O (3)-(5) are additional conditions to derive the asymptotic distribution of $\hat{\beta}(w^*)$. Assumption O (3) characterizes the exact tail behavior of f_ϵ^{ft}, which is typically needed to derive the distributional result for the deconvolution-based estimators. Assumption O (4) contains some smoothness conditions for conditional moments. Assumption O (5) gives an additional restriction on the bandwidth, where we undersmooth so that the estimation bias is asymptotically negligible.

Under these assumptions, the asymptotic properties of $\hat{\beta}(w^*)$ are obtained as follows.

Theorem 1.

(i): *Under Assumptions M (1)-(3) and O (1)-(2), it holds*

$$|\hat{\beta}(w^*) - \beta(w^*)|^2 = O_p(n^{-1}a_n^{-(1+2\alpha)} + a_n^{2p}).$$

(ii): *Under Assumptions M (1)-(4) and O (1)-(5), it holds*

$$\sqrt{na_n^{1+2\alpha}}\{\hat{\beta}(w^*) - \beta(w^*)\} \xrightarrow{d} N(0, \Omega(w^*)),$$

where $\Omega(w^*) = S(w^*)^{-1}\Sigma(w^*)S(w^*)^{-1}$ *with*

$$\begin{aligned} S(w^*) &= E[XX'|W^* = w^*]f_{W^*}(w^*), \\ \Sigma(w^*) &= C\int E[XX'(U + X'\{\beta(W^*) - \beta(w^*)\})^2|W^* = w^* - v]f_{W^*}(w^* - v)f_\epsilon(v)dv, \end{aligned}$$

and $C = \frac{1}{2\pi c^{\mathrm{os}2}}\int |K^{\mathrm{ft}}(t)|^2|t|^{2\alpha}dt$ *is a constant that depends on both* K *and* f_ϵ.

Theorem 1 (i) characterizes the L_2-risk property of our deconvolution estimator $\hat{\beta}(w^*)$. The second term, a_n^{2p}, in the convergence rate characterizes the magnitude of the estimation bias, which is identical to that of the error-free case. Note that when the class of densities for W^*

is restricted to satisfy $\int |f^{\mathrm{ft}}_{W^*}(t)|^2 \exp(2c_1|t|^{c_2})dt \le 2\pi c_3$ for some positive constants c_1, c_2, and c_3, exponential convergence rates for the estimation bias are possible, see, for example, Pensky and Vidakovic (1999), Butucea and Tsybakov (2008), and Comte and Lacour (2013). The first term, $n^{-1}a_n^{-(1+2\alpha)}$, characterizes the magnitude of the estimation variance. Compared to that of the error-free case, the estimation variance of $\hat{\beta}(w^*)$ decays more slowly due to the term $a_n^{-2\alpha}$; a smoother error distribution, which is characterized by a larger α, would lead to a larger estimation variance, hence a slower convergence rate. Similar convergence rates have been observed in other nonparametric measurement error problems, such as, Dong and Otsu (2019) for nonparametric additive models with errors-in-variables, Adusumilli and Otsu (2018) for nonparametric instrumental variable regressions with errors-in-variables, and Otsu and Taylor (2020) for specification testing in errors-in-variables regressions.

Theorem 1 (ii) says that the estimator $\hat{\beta}(w^*)$ is asymptotically normal, centred at the true value, and has variance $S(w^*)^{-1}\Sigma(w^*)S(w^*)^{-1}$. It is worthy to note that f_ϵ can be set as the Dirac delta function when there is no measurement error. So in the error-free context, $\Sigma(w^*) = \int K^2(x)dx\, E[XX'U^2|W^* = w^*]f_{W^*}(w^*)$ and the asymptotic variance $\Omega(w^*)$ would degenerate to the error-free asymptotic variance as in Li *et al.* (2002, Theorem 2.1).

In the supersmooth case, we impose the following additional assumptions.

Assumption S.

(1): *There exist positive constants* μ, c_0^{ss}, *and* $1/3 < \gamma \le 2$ *such that*

$$|f^{\mathrm{ft}}_\epsilon(t)|e^{\mu|t|^\gamma} \ge c_0^{\mathrm{ss}} \text{ for all } t \in \mathbb{R}.$$

(2): $a_n \to 0$ *and* $na_n e^{-2\mu a_n^{-\gamma}} \to \infty$ *as* $n \to \infty$.

(3): *There exists a positive constant* $c^{ss} \ge c_0^{ss}$ *such that*

$$f^{\mathrm{ft}}_\epsilon(t)e^{\mu|t|^\gamma} \to c^{\mathrm{ss}} \text{ as } |t| \to \infty.$$

(4): $K^{\mathrm{ft}}(1-t) = At^\theta + o(t^\theta)$ *as* $t \to 0$ *for some constants* A *and* $\theta \ge 0$.

(5): $na_n^{2p-2(2+\theta)}e^{-2\mu a_n^{-\gamma}} \to 0$ *as* $n \to \infty$ *if* $1/3 < \gamma < 1$ *and* $na_n^{2p-2\gamma(\theta+1)+2}e^{-2\mu a_n^{-\gamma}} \to 0$ *as* $n \to \infty$ *if* $1 \le \gamma \le 2$.

Assumption S (1) says that f_ϵ is supersmooth. Popular examples of supersmooth densities include the Gaussian and Cauchy densities. Similar to the ordinary smooth case, Assumption S (1) is different from the traditional supersmooth assumption, as in (2.32) of Meister (2009), in the sense that it only imposes the lower bound on $|f^{\mathrm{ft}}_\epsilon(t)|$ as this is sufficient to study the upper bound of the risk as in Theorem 2 (i). We restrict γ to be less than or equal to 2 to ensure that f_ϵ is a density (Chung, 1974, Theorem 6.5.4), while the lower bound of 1/3 guarantees that an approximation error is of smaller order than the asymptotic variance (the same assumption is imposed in van Es and Uh, 2004). Assumption S (2) concerns the bandwidth a_n; similar comments to the ordinary smooth case apply here. Equally, Assumptions S (3) and (5) are analogous to Assumptions O (3) and (5) in the ordinary smooth case. Assumption S (4) is an additional condition on the kernel function K. Examples of popular kernel functions which satisfy this extra constraint include the Sinc kernel, $K(x) = \sin(x)/(\pi x)$, where $\theta = 0$, and the kernel proposed in Fan (1992),

$$K(x) = \frac{48x(x^2-15)\cos(x) - 144(2x^2-5)\sin(x)}{\pi x^7},$$

where $\theta = 3$.

Under these assumptions, our deconvolution estimator $\hat{\beta}(w^*)$ has the following asymptotic properties.

Theorem 2.

(i): *Under Assumptions M (1)-(3) and S (1)-(2), it holds*

$$|\hat{\beta}(w^*) - \beta(w^*)|^2 = O_p(n^{-1}a_n^{-1}e^{2\mu a_n^{-\gamma}} + a_n^{2p}).$$

(ii): *Under Assumptions M (1)-(4) and S (1)-(5), it holds*

$$\Omega_n(w^*)^{-1/2}\{\hat{\beta}(w^*) - \beta(w^*)\} \xrightarrow{d} N(0, I_k),$$

where $\Omega_n(w^*) = n^{-1}S(w^*)^{-1}Var(\xi_1)S(w^*)^{-1}$ *and* ξ_1 *is defined in (A.2) in the Appendix.*

Similar comments to Theorem 1 apply here. Compared to the ordinary smooth case, the convergence rate of $\hat{\beta}(w^*)$ is considerably slower in the supersmooth case, reflecting the more difficult task of deconvolution in the presence of supersmooth contamination. In particular, rather than the polynomial rate obtained in Theorem 1 (i), by setting $a_n = c_s^{-1/\gamma}(\log(n))^{-1/\gamma}$ with $0 < c_s < 1/2\mu$, the variance term is $O\left(n^{2\mu c_s - 1}(\log(n))^{1/\gamma}\right)$, the bias term is $O\left((\log(n))^{-2p/\gamma}\right)$, and $\hat{\beta}(w^*)$ then converges at the rate of $(\log(n))^{-2p/\gamma}$, i.e., the rate presented in Theorem 2 (i) is logarithmic. Comparably slow logarithmic rates have been observed in other supersmooth settings, such as Fan (1992) and Schennach (2004).

3. CASE OF UNKNOWN MEASUREMENT ERROR DISTRIBUTION

In many applications, it is unrealistic to assume the measurement error distribution is known. In this section, we consider the situation where f_ϵ is unknown but repeated measurements on W^* are available.[2] In particular, we have two independent noisy measurements of the error-free covariate W^*, i.e.,

$$W = W^* + \epsilon \quad \text{and} \quad W^r = W^* + \epsilon^r. \tag{3.1}$$

Depending on whether the error density f_ϵ is symmetric around zero or not determines how these repeated measurements on W^* can be used to estimate f_ϵ. In this section, we use the

[2] Delaigle and Hall (2016) proposed an intriguing alternative approach to estimate the distributions of W^* and ϵ, which does not require additional data. They suppose ϵ has a symmetric density but the distribution of W^* is asymmetric and cannot be represented by a convolution of a density with another symmetric density. In our notation, their estimator for f_ϵ^{ft} can be written as $\hat{f}_{\epsilon,DH}^{\text{ft}}(t) = \left(\frac{1}{n}\sum_{j=1}^n e^{\text{i}tW_j}\right) / \left(\sum_{j=1}^n \hat{p}_j e^{\text{i}tW_j}\right)$, where $(\hat{p}_1, \ldots, \hat{p}_n)$ solves

$$\min_{p_1,\ldots,p_n} \int_{-\infty}^{\infty} \left| \frac{1}{n}\sum_{j=1}^n e^{\text{i}tW_j} - |\hat{\psi}(t)|^{1/2} \frac{\sum_{j=1}^n p_j e^{\text{i}tW_j}}{\left|\sum_{j=1}^n p_j e^{\text{i}tW_j}\right|} \right|^2 w(t)dt,$$

for some weight function $w(\cdot)$ and $\hat{\psi}(t) = \frac{1}{n(n-1)}\sum_{j_1<j_2} e^{\text{i}t(W_{j_1} - W_{j_2})}$. By replacing f_ϵ^{ft} with $\hat{f}_{\epsilon,DH}^{\text{ft}}$ in the deconvolution kernel (2.4), we can construct a feasible estimator for $\beta(w^*)$. Since technical arguments are substantially different from the present paper, we leave the analysis of such an estimator for future research.

repeated measurements on W^* to construct the estimator of $\beta(w^*)$ for the case of symmetric f_ϵ and derive the asymptotic properties of the proposed estimator. In the supplementary material, we consider the case of possibly asymmetric f_ϵ and again derive the asymptotic properties of the estimator.

When f_ϵ is symmetric around zero, to identify the distribution of ϵ, we impose the following additional assumption.

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

Assumption RS requires that ϵ^r is an independent copy of ϵ, and f_ϵ^{ft} is real-valued when f_ϵ is symmetric around zero. Under Assumption RS, the error distribution can be identified by $f_\epsilon^{\text{ft}}(t) = |E[\cos\{t(W - W^r)\}]|^{1/2}$. Based on an i.i.d. sample $\{W_j, W_j^r\}_{j=1}^n$ of (W, W^r), f_ϵ^{ft} can be estimated by (Delaigle, Hall and Meister, 2008)

$$\hat{f}_{\epsilon,s}^{\text{ft}}(t) = \left| \frac{1}{n} \sum_{j=1}^{n} \cos\{t(W_j - W_j^r)\} \right|^{1/2}.$$

Based on this estimator, we propose to estimate $\beta(w^*)$ by

$$\tilde{\beta}_s(w^*) = \left[\sum_{j=1}^{n} X_j X_j' \hat{\mathbb{K}}_s \left(\frac{w^* - W_j}{a_n} \right) \right]^{-1} \sum_{j=1}^{n} X_j Y_j \hat{\mathbb{K}}_s \left(\frac{w^* - W_j}{a_n} \right), \tag{3.2}$$

where $\hat{\mathbb{K}}_s$ is the deconvolution kernel function obtained by replacing f_ϵ^{ft} in (2.4) with $\hat{f}_{\epsilon,s}^{\text{ft}}$.

To analyse the asymptotic properties of $\tilde{\beta}_s(w^*)$, we focus on the following class of functions introduced by Schennach (2004).

Definition. *Let $\mathbb{W}$ be the set of all functions $\psi : \mathbb{R} \to \mathbb{R}$ such that (i) $\psi(t)$ is absolutely integrable in every finite interval, and (ii) $\int_{|t| \geq T} |\psi(t) - \Psi(t)| dt < \infty$ for some $T > 0$ and some function $\Psi(t)$ that can be written as a finite linear combination of finite products of functions of the forms $|t|^c$, $\text{sgn}(t)|t|^c$, $\log |t|$, $\sin(c_1 t)$, $\cos(c_1 t)$, and $\exp(c_1 t^a)$ with $c > 0$, $c_1 \in \mathbb{R}$, and $a \in \mathbb{N}$.*

As Schennach (2004, p. 1062) argued, this class $\mathbb{W}$ characterizes functions that are well behaved at infinity, and is useful to derive a lower bound for the asymptotic variance of our estimator. First, we impose the following assumptions for both the ordinary smooth and supersmooth error cases. Let $b(\cdot) = E[XX'|W^* = \cdot]\{\beta(\cdot) - \beta(w^*)\} f_{W^*}(\cdot)$.

Assumption MS.

- **(1):** *$E|\epsilon|^{2+\varsigma} < \infty$ for some $\varsigma > 0$.*
- **(2):** *f_ϵ is continuous and non-vanishing everywhere. $E[XX'|W^* = \cdot] f_{W^*}(\cdot)$, $\beta(\cdot)$, and $K(\cdot)$ are symmetric around zero.*
- **(3):** *$\frac{d}{dt} \left\{ \{\omega' b(\cdot)\}^{\text{ft}}(t) \{f_\epsilon^{\text{ft}}(t)\}^{-2} \right\} \in \mathbb{W}$ for any $\omega \in \mathbb{R}^k$ with $|\omega| = 1$.*
- **(4):** *For any $\omega \in \mathbb{R}^k$ with $|\omega| = 1$, there exists some constant $\bar{c}_s > 0$ such that*

$$Var(\omega'(\xi_1 + \xi_{s,1}^*)) \geq \bar{c}_s \max\{Var(\omega'\xi_1), Var(\omega'\xi_{s,1}^*)\},$$

where $\xi^*_{s,1}$ is defined in (B.16) in Appendix.

Assumption MS (1) is a regularity condition required by Lemma 3 in Appendix, which is used to characterize the uniform convergence rate of the empirical characteristic function of $\epsilon - \epsilon^r$ over an expanding region. Assumptions MS (2)-(4) contain further conditions used to derive the distribution of our estimator when the error distribution is unknown and symmetric around zero. We emphasise that the convergence rate result given below does not require these three assumptions.

Assumptions MS (2)-(3) are regularity conditions used to derive the lower bound for the variance of $\tilde{\beta}_s(w^*)$; similar assumptions have been used in Schennach (2004, Assumptions 12 and 13). In particular, Assumption MS (3) characterizes the tail behavior of the dominant component of $\tilde{\beta}_s(w^*)$'s asymptotic representation brought by the estimation error of $f^{\mathrm{ft}}_{\epsilon,s}$, and is used together with Assumption O (3) in the ordinary smooth case or Assumption S (3) in the supersmooth case to establish the lower bound of the variance of $\tilde{\beta}_s(w^*)$. Assumption MS (2) also includes conditions on the symmetry of the density, conditional moment, coefficient function, and kernel, which are imposed here to be compatible with the symmetry of f_ϵ. We emphasise that these additional symmetry conditions are only imposed to allow a cleaner expression for the lower bound for the variance of $\tilde{\beta}_s(w^*)$, which allows for more compact conditions on the bandwidth. The same results on the convergence rate and asymptotic normality of $\tilde{\beta}_s(w^*)$ can be established without these additional symmetry conditions, but at the cost of more complex conditions on the bandwidth. Assumption MS (4) states that the variance of $\tilde{\beta}_s(w^*)$ is of an order no less than any term in its asymptotic representation; a similar assumption is used in Schennach (2004, Assumption 14).

When f_ϵ is ordinary smooth, we further impose the following assumptions.

Assumption OS.

(1): $n^{-1/2}a_n^{-(1+3\alpha)}\log(1/a_n) \to 0$ *as* $n \to \infty$.

(2): *For any* $\omega \in \mathbb{R}^k$ *with* $|\omega| = 1$ *and some* $\eta > 0$, *as* $n \to \infty$,

$$\left\{ \begin{array}{c} \min\{n^{-1}a_n^{-2p}, na_n^{2(1+5\alpha)}\log(1/a_n)^{-4}, n^{\eta/(\eta+2)}a_n^{2(1+2\alpha)}\} \\ \times \max\left\{a_n^{-(1+2\alpha)}, \int \left|\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)\cos(tw^*)K^{\mathrm{ft}}(ta_n)}{\{f^{\mathrm{ft}}_\epsilon(t)\}^2}\right|^2 dt\right\} \end{array} \right\} \to \infty.$$

Assumption OS (1) is imposed to control the magnitude of the estimation error from the estimated error characteristic function $\hat{f}^{\mathrm{ft}}_{\epsilon,s}$ when the measurement error is ordinary smooth. Assumption OS (2) imposes additional bandwidth conditions to derive the asymptotic normality of $\tilde{\beta}_s(w^*)$. The component $\max\left\{a_n^{-(1+2\alpha)}, \int \left|\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)\cos(tw^*)K^{\mathrm{ft}}(ta_n)}{\{f^{\mathrm{ft}}_\epsilon(t)\}^2}\right|^2 dt\right\}$ characterizes (up to $1/n$) the magnitude of the lower bound of the variance of $\tilde{\beta}_s(w^*)$. In particular, the first term characterizes the estimation variance if f_ϵ is known and depends on the smoothness of f_ϵ as in the standard deconvolution literature, and the second term, which can be further simplified as $\int_{|t|\leq a_n^{-1}} \left|\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)\cos(tw^*)}{\{f^{\mathrm{ft}}_\epsilon(t)\}^2}\right|^2 dt$ if the Sinc kernel is used, characterizes the estimation error brought by using $\hat{f}^{\mathrm{ft}}_{\epsilon,s}$ and depends on the relative smoothness to f_ϵ of the density f_{W^*}, conditional moment $E[XX'|W^*]$, and coefficient function β. The component $n^{-1}a_n^{-2p}$ is related to the estimation bias when f_ϵ is known, the component $na_n^{2(1+5\alpha)}\log(1/a_n)^{-4}$ is related to the estimation error brought by using $\hat{f}^{\text{ft}}_{\epsilon,s}$, and the component $n^{\eta/(\eta+2)}a_n^{2(1+2\alpha)}$ is related to Lyapunov's condition to apply the central limit theorem. So, Assumption OS (2) contains further restrictions on the bandwidth in order to use Lyapunov's central limit theorem and to ensure the asymptotic negligibility of the estimation bias and higher order terms in the estimation error from using $\hat{f}^{\text{ft}}_{\epsilon,s}$.

Theorem 3.

(i): *Under Assumptions M (1)-(4), O (1)-(2), RS, MS (1), and OS (1), it holds*

$$|\tilde{\beta}_s(w^*) - \beta(w^*)|^2 = O_p(n^{-1}a_n^{-2(1+3\alpha)}\log(1/a_n)^2 + a_n^{2p}).$$

(ii): *Under Assumptions M (1)-(5), O (1)-(5), RS, MS (1)-(4), and OS (1)-(2), it holds*

$$\hat{\Omega}_{n,s}(w^*)^{-1/2}\{\tilde{\beta}_s(w^*) - \beta(w^*)\} \xrightarrow{d} N(0, I_k),$$

where $\hat{\Omega}_{n,s}(w^*) = n^{-1}S(w^*)^{-1}Var(\xi_1 + \xi^*_{s,1})S(w^*)^{-1}$.

Theorem 3 (i) shows the L_2-risk of our estimator $\tilde{\beta}_s(w^*)$ when the measurement error is ordinary smooth and symmetric around zero. The first term results from the estimation error of $\hat{f}^{\text{ft}}_{\epsilon,s}$, while the second term is the usual bias term from an error-free nonparametric estimator. Theorem 3 (ii) shows that the estimator retains its asymptotic normality when the measurement error characteristic function is estimated using the approach of Delaigle, Hall and Meister (2008), and the extra estimation error brought by using $\hat{f}^{\text{ft}}_{\epsilon,s}$ affects both the convergence rate and the asymptotic variance. Compared to the rate result of $\hat{\beta}(w^*)$ established in Theorem 1 when the measurement error distribution is known, $\tilde{\beta}_s(w^*)$ has a slower convergence rate due to the estimation error from using $\hat{f}^{\text{ft}}_{\epsilon,s}$.

When f_ϵ is supersmooth, we impose the following assumptions.

Assumption SS.

(1): $n^{-1/2}e^{3\mu a_n^{-\gamma}}a_n^{-1}\log(1/a_n) \to 0$ *as* $n \to 0$.

(2): *For any* $\omega \in \mathbb{R}^k$ *with* $|\omega| = 1$ *and for some* $\eta > 0$, *as* $n \to \infty$,

$$\left\{\begin{array}{c} \min\{n^{-1}a_n^{-2p}, ne^{-10\mu a_n^{-\gamma}}a_n^2\log(1/a_n)^{-4}, n^{\eta/(\eta+2)}e^{-4\mu a_n^{-\gamma}}a_n^2 \\ \times \max\left\{e^{2\mu a_n^{-\gamma}}a_n^{2(2+\theta)}, \int\left|\frac{\{\omega' b(\cdot)\}^{\text{ft}}(t)\cos(tw^*)K^{\text{ft}}(ta_n)}{\{f^{\text{ft}}_\epsilon(t)\}^2}\right|^2 dt\right\}\end{array}\right\} \to \infty,$$

if $1/3 < \gamma < 1$ *and*

$$\left\{\begin{array}{c} \min\{n^{-1}a_n^{-2p}, ne^{-10\mu a_n^{-\gamma}}a_n^2\log(1/a_n)^{-4}, n^{\eta/(\eta+2)}e^{-4\mu a_n^{-\gamma}}a_n^2 \\ \times \max\left\{e^{2\mu a_n^{-\gamma}}a_n^{2(\gamma\theta+\gamma-1)}, \int\left|\frac{\{\omega' b(\cdot)\}^{\text{ft}}(t)\cos(tw^*)K^{\text{ft}}(ta_n)}{\{f^{\text{ft}}_\epsilon(t)\}^2}\right|^2 dt\right\}\end{array}\right\} \to \infty,$$

if $1 \leq \gamma \leq 2$.

Assumption SS (1) is imposed to control the magnitude of the estimation error from using the estimated error characteristic function $\hat{f}^{\text{ft}}_{\epsilon,s}$ when the measurement error is supersmooth. Assumption SS (2) imposes additional conditions on the bandwidth to derive the asymptotic normality of $\tilde{\beta}_s(w^*)$. Similar comments to Assumption OS (2) apply here, except that the bandwidth conditions are separately imposed for the cases with $1/3 < \gamma < 1$ and $1 \leq \gamma \leq 2$, which is compatible with Assumption S (5) as in the case when f_ϵ is known.

Theorem 4.

(i): *Under Assumptions M (1)-(4), S (1)-(2), RS, MS (1), and SS (1), it holds*

$$|\tilde{\beta}_s(w^*) - \beta(w^*)|^2 = O_p(n^{-1} e^{6\mu a_n^{-\gamma}} a_n^{-2} \log(1/a_n)^2 + a_n^{2p}).$$

(ii): *Under Assumptions M (1)-(5), S (1)-(5), RS, MS (1)-(4), and SS (1)-(2), it holds*

$$\hat{\Omega}_{n,s}(w^*)^{-1/2}\{\tilde{\beta}_s(w^*) - \beta(w^*)\} \xrightarrow{d} N(0, I_k),$$

where $\hat{\Omega}_{n,s}(w^*) = n^{-1} S(w^*)^{-1} Var(\xi_1 + \xi^*_{s,1}) S(w^*)^{-1}$.

Similar comments to Theorem 3 apply here. Theorem 4 (i) shows the L_2-risk of our estimator $\tilde{\beta}_s(w^*)$ when the measurement error is supersmooth and symmetric around zero. By similar arguments to the case when the error distribution is known, the rate of $\tilde{\beta}_s(w^*)$ presented in Theorem 4 (i) can be shown to be logarithmic, which is considerably slower than the polynomial rate obtained for $\tilde{\beta}_s(w^*)$ in the ordinary smooth case. Theorem 4 (ii) shows that the estimator retains its asymptotic normality when the measurement error characteristic function is estimated using the approach of Delaigle, Hall and Meister (2008). The estimation error from using $\hat{f}^{\text{ft}}_{\epsilon,s}$ affects both the convergence rate and the asymptotic variance of $\tilde{\beta}_s(w^*)$. Again $\tilde{\beta}_s(w^*)$ converges more slowly than $\hat{\beta}(w^*)$ due to the estimation error from using $\hat{f}^{\text{ft}}_{\epsilon,s}$.

4. Simulation

In this section, the small sample properties of our deconvolution estimator are investigated using a Monte Carlo study. The following data generating process is considered

$$Y = \beta_0(W^*) + X_1\beta_1(W^*) + X_2\beta_2(W^*) + U,$$

where (X_1, X_2) are drawn from $U[0, 1]$ with correlation of 0.2 and independent of (W^*, U), and U is drawn from $N(0, 1)$ and is independent of W^*. While W^* is assumed unobservable, we suppose $W = W^* + \epsilon_1$ and $W^r = W^* + \epsilon_2$ are observed, where (ϵ_1, ϵ_2) is mutually independent and independent of (X_1, X_2, W^*, U).

For the densities of W^* and (ϵ_1, ϵ_2), we consider two cases. First, for the ordinary smooth setting, (ϵ_1, ϵ_2) have a zero mean Laplace distribution with standard deviation of 1/3, and W^* also has a Laplace distribution with zero mean and standard deviation of 1. Second, for the supersmooth case, (ϵ_1, ϵ_2) have a normal distribution with zero mean and standard deviation of 1/3, and W^* has a standard normal distribution.

We take $\beta_0(w) = \cos(w\pi/2)$, $\beta_1(w) = 1 + w + w^2$, and consider three specifications for $\beta_2(w)$:

$$\begin{aligned}
\text{DGP1} :& \beta_2(w) = 1 + w, \\
\text{DGP2} :& \beta_2(w) = 1 + w + w^2, \\
\text{DGP3} :& \beta_2(w) = 1 + w + w^2 - w^3.
\end{aligned}$$

Note that each varying coefficient is further standardised by its respective standard deviation $\sqrt{Var(\beta_j(W^*))}$ for $j = 0, 1, 2$ so that each component adds the same explanatory power to the model.

Throughout this simulation study, we use the infinite-order flat-top kernel proposed by McMurry and Politis (2004) whose Fourier transform is

$$K^{\text{ft}}(t) = \begin{cases} 1 & \text{if } |t| \leq 0.05, \\ \exp\left\{\frac{-\exp(-(|t|-0.05)^2)}{(|t|-1)^2}\right\} & \text{if } 0.05 < |t| < 1, \\ 0 & \text{if } |t| \geq 1, \end{cases}$$

and satisfies all necessary assumptions given previously in this paper. In preliminary simulations (not reported) more stable estimates are found using this kernel in comparison to either the sinc kernel or the kernel of Fan (1992). Results for two sample sizes, $n = 250$ and 500, are provided, and all results are based on 500 Monte Carlo replications.

4.1. **Bandwidth choice.** As with any nonparametric kernel estimation method, the bandwidth choice is critical for the performance of our estimator. Many data-driven methods exist to select the bandwidth in kernel deconvolution density estimation, see, for example, Delaigle and Gijbels (2004), Lepski (2018), and Lepski and Willer (2019). However, the analogous problem in a regression setting has received less attention. We choose to adapt the 'out-of-bag' approach of Dong, Otsu and Taylor (2020), which we briefly describe here in the context of our varying coefficient estimator.

Consider the following criterion function to determine the optimal bandwidth

$$S(a_n) = E[\{Y_{n+1} - X'_{n+1}\hat{\beta}(W^*_{n+1}; a_n)\}^2],$$

where $(Y_{n+1}, X_{n+1}, W^*_{n+1})$ are independent of the original sample used to compute $\hat{\beta}(\cdot; a_n)$. In the absence of measurement error, a popular method to estimate this criterion is the leave-one-out cross validation estimator

$$\hat{S}_{CV}(a_n) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - X'_i\hat{\beta}_{-i}(W^*_i; a_n)\}^2,$$

where $\hat{\beta}_{-i}(\cdot; a_n)$ is the estimator for β using all observations except the i-th data point. This leave-one-out approach removes the dependence between the estimator and the data used to evaluate the estimator. However, when W^* is measured with error, we do not have access to its true value, making this approach infeasible.

Instead, $S(a_n)$ can be estimated by

$$\begin{aligned} \tilde{S}(a_n) &= \iiint \{y - x'\hat{\beta}(w^*; a_n)\}^2 \hat{f}_{Y,X,W^*}(y, x, w^*)dydxdw^* \\ &= \frac{1}{nh_n}\sum_{i=1}^{n}\int \{Y_i - X'_i\hat{\beta}(w^*; a_n)\}^2 \hat{\mathbb{K}}\left(\frac{W_i - w^*}{h_n}\right) dw^*, \end{aligned}$$

where the equality follows from Y and X being observable, and $\int u^j K(u)du = 0$ for $j = 0, 1, 2$ for a (higher-order) conventional kernel function K.

Two points are important to recognize at this stage. First, an auxiliary bandwidth, h_n, must be selected; we return to this point below. Second, in the same spirit as the leave-one-out cross validation approach, there should be no overlap in the data used to estimate $\hat{\beta}$ as used to evaluate it. Although a leave-one-out approach is feasible here, the number of $\hat{\beta}_{-i}(w^*; a_n)$ to be calculated can become overwhelming when the sample size is relatively large and a sensibly sized grid of candidate bandwidths is considered.

To circumvent this problem, an out-of-bag approach can be used. Specifically, take a bootstrap sample of size n (with replacement) from the original data, and estimate $\hat{\beta}$ using this bootstrap sample. Evaluate $\tilde{S}(a_n)$ using the observations which were not selected as part of the bootstrap sample, i.e., the out-of-bag observations. On average, these out-of-bag observations make up approximately 37% of the total sample size (Breiman, 2001). Repeat this process for bootstrap samples $b = 1, \ldots, B$, giving B bootstrap versions of $\hat{S}_b(a_n)$ defined as

$$\hat{S}_b(a_n) = \frac{1}{\#I_b^c h_{\#I_b^c}} \sum_{i \in I_b^c} \int \{Y_i - X_i' \hat{\beta}_{I_b}(w^*; a_n)\}^2 \hat{\mathbb{K}} \left(\frac{W_i - w^*}{h_{\#I_b^c}} \right) dw^*,$$

where I_b denotes the set of observations in the bootstrap sample b, I_b^c is the complement of this set (i.e., the out-of-bag observations), and $\#I_b^c$ denotes the cardinality of the set I_b^c. Finally, $S(a_n)$ is estimated by taking an average of $\hat{S}_b(a_n)$ over the bootstrap samples

$$\hat{S}(a_n) = \frac{1}{B} \sum_{b=1}^{B} \hat{S}_b(a_n),$$

and a_n^* is chosen as the minimiser of $\hat{S}(a_n)$.

To choose the auxiliary bandwidth parameter $h_{\#I_b^c}$, we suggest using the approach of Delaigle and Gijbels (2004). From initial simulation results (not presented), the bandwidth selection procedure is relatively insensitive to the choice of $h_{\#I_b^c}$ providing that it is chosen in a sensible manner.

Figure 1 shows a representative plot of $\hat{S}(a_n)$ against a_n for DGP3 with 250 observations and Laplace errors which are assumed unknown but symmetric. We found that the performance of the procedure converged with a relatively small number of bootstrap resamples; thus, only 50 bootstrap resamples were used in this example (this number was also used by Efron and Tibshirani, 1997, in a similar bootstrap cross-validation procedure). There is a unique minimum at 0.36 (indicated by the dashed line) which gives the chosen bandwidth in this sample.

While Dong, Otsu and Taylor (2020) provide a theoretical justification for this method for nonparametric regression, this does not ensure the theoretical validity of the approach in our context. This is beyond the scope of this paper; however, below, we provide finite sample evidence of the suitability of this method.

Figures 2, 3, and 4 plot the mean integrated squared error (MISE) over bootstrap replications for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$, respectively, (between the 10-th and 90-th percentile of W^*) as a function of the difference between the bandwidth used and the 'optimal' bandwidth chosen by the selection procedure. Again, the results are based on DGP3 with 250 observations and Laplace errors which are assumed unknown but symmetric. The three figures show that the selected bandwidth is slightly larger than the optimum for estimating β_0, but smaller than the optimum for estimating β_1 and β_2. Thus, the selection procedure balances the respective costs across the three functions to be estimated. It is not surprising that these estimated functions behave differently for various bandwidth choices since the degree of nonlinearity is different for each.

This raises the question of whether a separate bandwidth should be chosen that is optimal for each respective β coefficient. There are two issues with this idea. First, as can be seen in (2.3), the estimation procedure does not lend itself easily to using several different bandwidths. Secondly, the bandwidth selection procedure - as well as other potential procedures - is based on predicting the outcome, Y, rather than the individual coefficient functions because these functions are inherently unobservable. Thus, the procedure cannot be artificially directed towards any one coefficient function in the model.

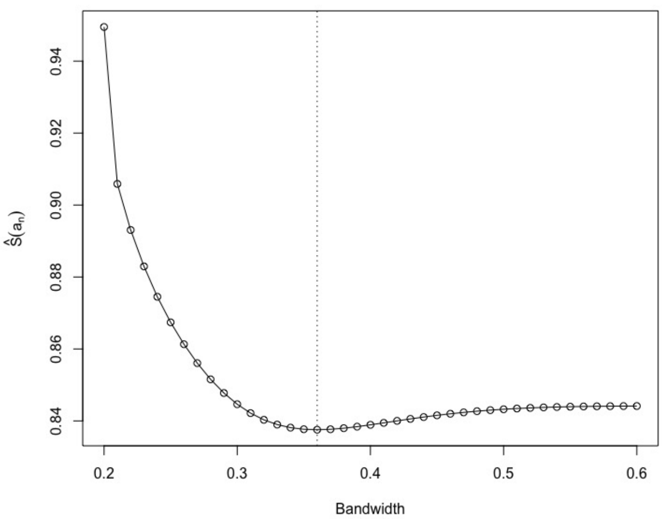

FIGURE 1. Representative plot of $\hat{S}(a_n)$ against a_n for DGP3 with 250 observations and Laplace errors where the measurement error is assumed unknown but symmetric. 50 bootstrap resamples were used. The dashed line (at 0.36) indicates the selected bandwidth in this sample.

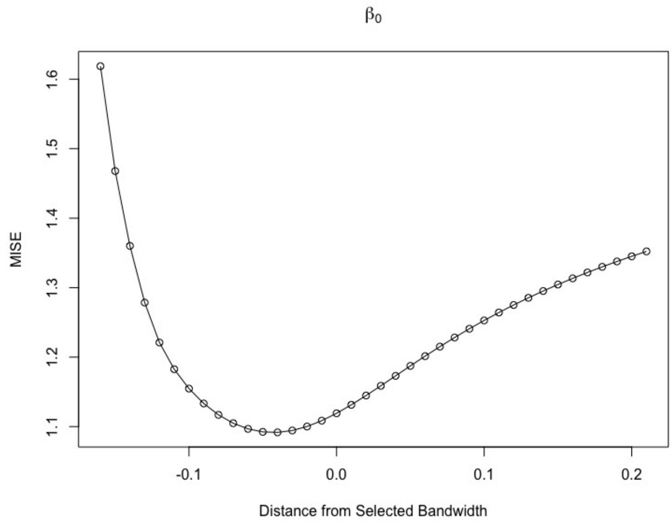

FIGURE 2. Plot of the MISE (over 500 bootstrap replications) for $\hat{\beta}_0(\cdot)$ (between the 10-th and 90-th percentile of W^*) as a function of the difference between the bandwidth used and the 'optimal' bandwidth chosen by the selection procedure. As in Figure 1, this plot is based on DGP3 with 250 observations and Laplace errors where the measurement error is assumed unknown but symmetric and 50 bootstrap resamples were used in the selection procedure.

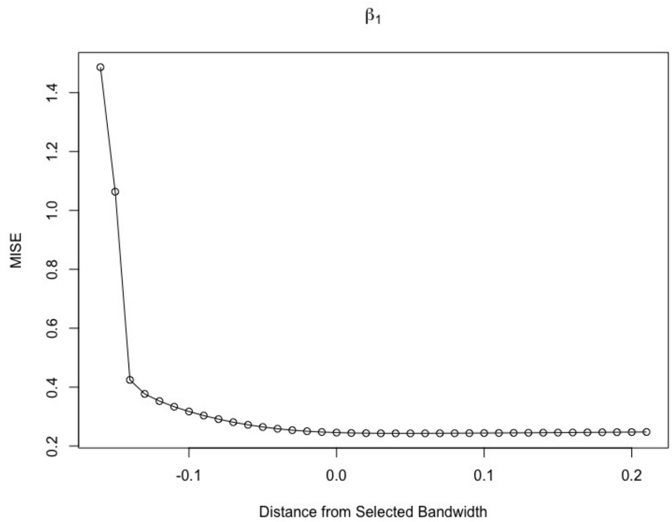

FIGURE 3. Plot of the mean (over 500 bootstrap replications) integrated squared error (MISE) for $\hat{\beta}_1(\cdot)$ (between the 10-th and 90-th percentile of W^*) as a function of the difference between the bandwidth used and the 'optimal' bandwidth chosen by the selection procedure. As in Figure 1, this plot is based on DGP3 with 250 observations and Laplace errors where the measurement error is assumed unknown but symmetric and 50 bootstrap resamples were used in the selection procedure.

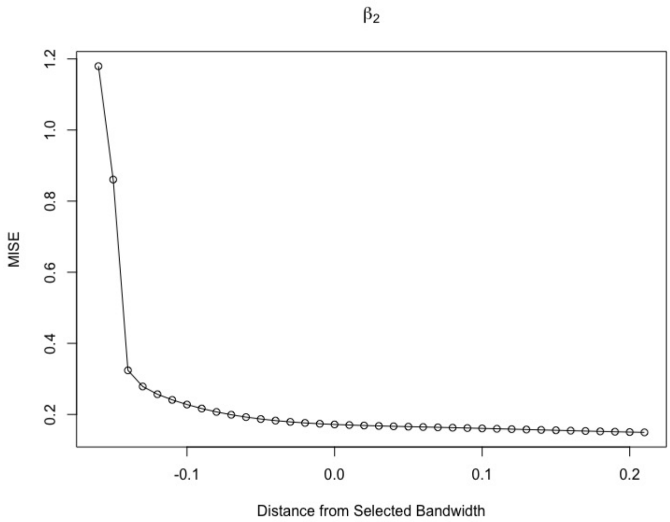

FIGURE 4. Plot of the mean (over 500 bootstrap replications) integrated squared error (MISE) for $\hat{\beta}_2(\cdot)$ (between the 10-th and 90-th percentile of W^*) as a function of the difference between the bandwidth used and the 'optimal' bandwidth chosen by the selection procedure. As in Figure 1, this plot is based on DGP3 with 250 observations and Laplace errors where the measurement error is assumed unknown but symmetric and 50 bootstrap resamples were used in the selection procedure.

4.2. **Results.** In Tables 1-3, results are reported for several varying coefficient estimators. The row labelled 'Known' denotes the estimator of this paper for a known error characteristic function. 'Symm' refers to the same estimator but using the repeated measurements of the noisy regressor and assuming a symmetric error density ('Symm'). 'Asymm' presents results for our varying coefficient estimator when the measurement error is assumed unknown and possibly asymmetric. 'LHLF' denotes the estimator of Li *et al.* (2002) using the mismeasured regressor, W, as if it were the truth. Finally, 'LHLF*' is the estimator of Li *et al.* (2002) using the correctly measured regressor, W^*. Note that this final estimator would be infeasible in practice, but acts as a useful benchmark with which to compare the other estimators. The bandwidths for LHLF and LHLF* are chosen by leave-one-out cross validation. Results are given for the MISE of each β coefficient between the 10-th and 90-th percentile of W^*.

Table 1: DGP 1

Estimator	β_0				β_1				β_2			
Error Type	**OS**		**SS**		**OS**		**SS**		**OS**		**SS**	
Sample Size	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**
Known	1.47	0.82	1.57	0.93	0.36	0.19	0.39	0.23	0.37	0.22	0.37	0.21
Symm	1.40	0.78	1.56	0.93	0.33	0.17	0.38	0.23	0.35	0.20	0.37	0.21
Asymm	1.39	0.77	1.58	0.94	0.32	0.17	0.40	0.24	0.35	0.20	0.38	0.22
LHLF	1.47	0.78	1.64	0.93	0.37	0.20	0.44	0.26	0.39	0.23	0.43	0.25
LHLF*	1.25	0.65	1.42	0.75	0.32	0.16	0.38	0.20	0.34	0.18	0.38	0.19

Table 2: DGP 2

Estimator	β_0				β_1				β_2			
Error Type	**OS**		**SS**		**OS**		**SS**		**OS**		**SS**	
Sample Size	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**
Known	1.19	0.67	1.30	0.77	0.26	0.15	0.30	0.17	0.26	0.15	0.28	0.16
Symm	1.16	0.65	1.29	0.78	0.25	0.13	0.30	0.17	0.25	0.14	0.28	0.16
Asymm	1.16	0.66	1.32	0.78	0.24	0.13	0.32	0.17	0.25	0.14	0.29	0.17
LHLF	1.22	0.66	1.38	0.78	0.30	0.17	0.37	0.22	0.31	0.18	0.36	0.20
LHLF*	1.11	0.58	1.28	0.72	0.28	0.14	0.29	0.19	0.29	0.15	0.33	0.18

Table 3: DGP 3

Estimator	β_0				β_1				β_2			
Error Type	**OS**		**SS**		**OS**		**SS**		**OS**		**SS**	
Sample Size	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**	**250**	**500**
Known	1.15	0.64	1.10	0.66	0.25	0.14	0.25	0.15	0.25	0.14	0.24	0.14
Symm	1.12	0.62	1.10	0.66	0.24	0.13	0.24	0.15	0.24	0.12	0.24	0.14
Asymm	1.14	0.62	1.12	0.67	0.23	0.13	0.26	0.16	0.23	0.12	0.26	0.16
LHLF	1.22	0.66	1.24	0.74	0.30	0.17	0.34	0.18	0.30	0.17	0.34	0.17
LHLF*	1.13	0.58	1.09	0.70	0.28	0.14	0.32	0.18	0.30	0.14	0.31	0.16

The results appear encouraging and display several interesting features. First, as is expected, the performance of all five estimators is poorer when faced with supersmooth error than with ordinary smooth. Somewhat surprisingly, the relative performance of the three estimators studied in this paper depends on the smoothness of the measurement error. For Gaussian error, Known dominates Symm and Symm dominates Asymm, in accordance with the theoretical results. However, for Laplace error, the Known estimator is dominated by the other two (which appear to be almost identical); Delaigle, Hall and Meister (2008) find a similar result with the performance of the deconvolution kernel regression estimator improving in some cases when the measurement error is estimated.

It is unsurprising to see that the performance of the estimators of this paper, relative to the correctly measured benchmark, LHLF*, deteriorate with the sample size. This reflects the slower convergence rates of our estimators in comparison to the estimator of LHLF*. It is interesting to see that as the degree of nonlinearity in the model increases, moving from DGP1 through to DGP3, the performance of our estimators increase relative to LHLF* and, for some coefficients in some settings, is actually superior to LHLF*.

In relation to LHLF, the estimators of this paper improve with the sample size, which reflects the inherent bias in LHLF from not accounting for the measurement error. While this relative performance does not appear to depend on the smoothness of the error. Note that the MISE for LHLF does fall as the sample size increases; this reflects a decrease in the variance of the estimator and not a decrease in the bias. Finally, as to be expected, the dominance of the estimators of this paper in comparison to LHLF increases with the nonlinearity in the model, i.e. moving from DGP1 through to DGP3.

5. Empirical Application

In this section, we apply our estimator to real data. In particular, we consider the role risk preferences play in explaining earnings, and how this relationship is affected by cognitive ability.

It has been consistently shown that cognitive ability has a large effect on labour market outcomes (see, for example, Herrnstein and Murray, 1994; Cawley, Heckman and Vytlacil, 2001). Although comparatively less attention has been given to the role of risk preferences in determining earnings, it has been shown that risk averse individuals earn less than their risk-loving counterparts (Bonin *et al.*, 2007). In addition, there is growing evidence that cognitive ability and attitudes towards risk are intimately related (Dohmen, Falk, Huffman, Sunde, 2010, 2018).

Taken together, these results suggest that in an analysis of the determinants of wages, both cognitive ability and risk preferences should be accounted for. Moreover, their effects are likely to interact, i.e. the effect of one of these variables on earnings is likely to depend on the level of the other. Furthermore, it is well-known that measures of cognitive ability are not perfect and can be subject to substantial measurement error. Hence, the varying-coefficient model of this paper, with cognitive ability modifying the coefficients, is perfectly suited to this setting.

This is by no means the first investigation which allows the coefficients of a wage equation to depend on cognitive ability. Heckman and Vytlacil (2001) use a nonparametric approach to investigate how the returns to education vary across individuals' ability. While Heckman, Stixrud and Urzua (2006) and Cunha, Heckman and Schennach (2010) also explore this link but go beyond the previous literature by taking seriously the role of measurement error in cognitive ability.

The data used for this empirical study come from the National Longitudinal Survey of Youth 1979 (NLSY79). This dataset follows a sample of Americans born between 1957-1964 who were aged between 14-22 when first interviewed in 1979 and were periodically surveyed until the present day (the most recent survey was conducted in 2016). As is typical in the literature using the NLSY79 dataset, we restrict the sample to white males who work in the formal labour market, giving a sample size of 1473.

The dataset contains a plethora of socio-economic information on the respondents; however, we focus on the following variables. The outcome of interest is the natural logarithm of annual earnings (recorded in 1996). The regressor of interest is the risk preference of the individual taken from a survey question in 2010 (the only year in which this question is asked) regarding the willingness to take risks while driving. The measure ranges from 0 - 10, with 0 being the most risk averse.[3] Dohmen *et al.* (2011) show that a direct question on risk taking preferences - as opposed to results from lottery experiments - provides the best predictor of risk-taking behaviour. Furthermore, we opt to use driving risk preferences - instead of, for example, financial risk preferences - in an attempt to avoid possible simultaneity bias. We also choose to control for age and age squared, years of education, years of work experience, and number of children (all measured in 1996). Each regressor is standardised to have zero mean and unit variance.

Out of a battery of ASVAB test scores (taken in 1981), the three tests which are most highly correlated with earnings are used as three repeated measures of cognitive ability, denoted (W_1, W_2, W_3).[4] A factor model structure is used, as in Cunha, Heckman and Schennach (2010), which assumes that each measure is constructed as

$$W_j = \alpha_j C + \epsilon_j \qquad \text{for } j = 1, 2, 3. \tag{5.1}$$

We are free to make one normalising restriction in order to identify the α parameters. Without loss of generality take $\alpha_1 = 1$. Then, under the maintained assumption of classical measurement error, α_2 can be identified via $\alpha_2 = Cov(W_2, W_3)/Cov(W_1, W_3)$ and α_3 can be obtained analogously.

Unfortunately, there exist no tests for the classical measurement error assumption. However, we hope to provide some evidence of this assumption being satisfied in our context. Table 4 gives results of a regression of log of earnings on the three principal components of (W_1, W_2, W_3), showing that only one component affects income. This suggests that the measurement error is unrelated to earnings and, thus, is plausibly unrelated to cognitive ability. Note that only two of the three principal components can be interpreted as measurement error. Thus, since $(\epsilon_1, \epsilon_2, \epsilon_3)$ are constructed from two uncorrelated factors, this approach implicitly violates the assumption that the measurement errors are independent across the repeated measures. So,

[3]The survey question is phrased as, "How would you rate your willingness to take risks (on a scale of 0-10) while driving?".

[4]These are: ‘paragraph comprehension’, ‘numerical operations’, and ‘coding speed’.

while this method does not prove the classical error assumption holds, it does give weight to its validity.

TABLE 4. Principal Component Regression

	Dependent variable:
	log Earnings
Principal Component 1	-0.272^{***}
	(0.016)
Principal Component 2	0.026
	(0.034)
Principal Component 3	-0.038
	(0.044)
Constant	-0.000
	(0.024)
Observations	1,473
R^2	0.165

Notes: Results from a regression of log earnings (in 1996) on the three principal components of the ASVAB test scores from the NLSY79 dataset. * indicates significance at 10%, ** indicates significance at 5%, and *** indicates significance at 1%.

The full model is written as

$$\begin{aligned}\log Y &= \beta_0(C) + \beta_1(C)Risk + \beta_2(C)Age + \beta_3(C)Age^2 \\ &\quad + \beta_4(C)Educ + \beta_5(C)Exp + \beta_6(C)Child + U, \end{aligned} \tag{5.2}$$

where Y is earnings, C is (unobserved) cognitive ability, $Risk$ is our risk measure, Age is the age of the respondent, $Educ$ is the number of years of educations, Exp is the number of years of work experience, and $Child$ is the number of children the respondent has. We assume the measurement error comes from a symmetric distribution and use the bandwidth selection mechanism introduced in Section 4.1 with 50 bootstrap resamples.

We compare our results to the estimator of Li *et al.* (2002). Recall, that this estimator is not designed for contaminated data. As such, the average of the three ASVAB test scores is used as cognitive ability as if it is correctly measured. As in Section 4, the bandwidth is chosen by leave-one-out cross-validation.

In the interest of space, and since our focus is on the interplay between cognitive ability and risk preference, we report results only for β_0 and β_1. Figure 5 plots $\hat{\beta}_0$ between the 10^{th} and 90^{th} percentile of the ASVAB test score. Note that for ease of interpretation, the y-axis has been transformed such that earnings are displayed as the annual dollar value rather than the natural logarithm. The solid line gives the estimate using the estimator of this paper and the dashed line denotes the estimate of Li *et al.* (2002). Unsurprisingly, cognitive ability has a relatively large positive impact on earnings. Moving from one standard deviation below the mean to one standard deviation above the mean leads to an increase of approximately \$8400 per annum (note that median earnings in the data are \$32000). In contrast, the estimator of Li *et al.* (2002) predicts a \$5500 increase. Moreover, while the general shape of the two estimators is similar, the estimator of Li *et al.* (2002) appears to struggle to find the nonlinearity in the lower tail of cognitive ability that the estimator of this paper shows.

Figure 6 displays the results of $\hat{\beta}_1$. Note that in this case, the y-axis has been transformed to give the effect of risk preference on earnings as a percent, i.e. a value of 2 indicates that a one standard deviation increase in risk preference leads to a 2 percent increase in earnings. Again, the two estimates follow a similar shape, with our estimator detecting greater nonlinearity in the lower tail. According to the results of our estimator, an individual at the mean level of cognitive ability faces the following effect: a one standard deviation increase in risk preferences leads to a 1.5 percent increase in earnings. However, the results suggest that the effect of risk preferences on earnings depends critically on the level of cognitive ability. While those with high cognitive ability are expected to earn more from risk-loving preferences, the opposite is true for those with low cognitive ability. An individual with cognitive ability at one standard deviation below the mean faces a risk-preference-effect-on-earnings that is 3.6 percentage points below that of a respondent with cognitive ability one standard deviation above the mean. Again, the estimator of Li *et al.* (2002) predicts a smaller effect: a 1.9 percentage point drop.

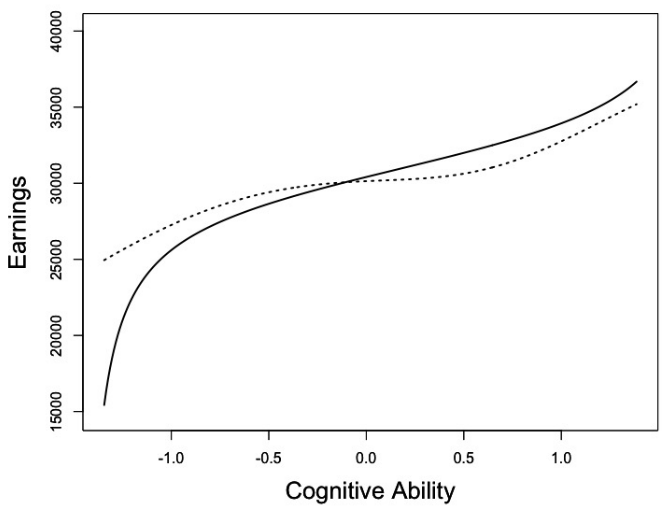

FIGURE 5. Plot of $\hat{\beta}_0$ between the 10^{th} and 90^{th} percentile of the ASVAB test score from the regression in (5.2) using the NLSY79 dataset. Note that the y-axis has been transformed such that earnings are displayed as the annual dollar value rather than the natural logarithm. The solid line gives the estimate using the estimator of this paper when the measurement error is assumed unknown but symmetric. The dashed line denotes the estimate of Li *et al.* (2002) using the mean of the ASVAB test scores as if it is true cognitive ability.

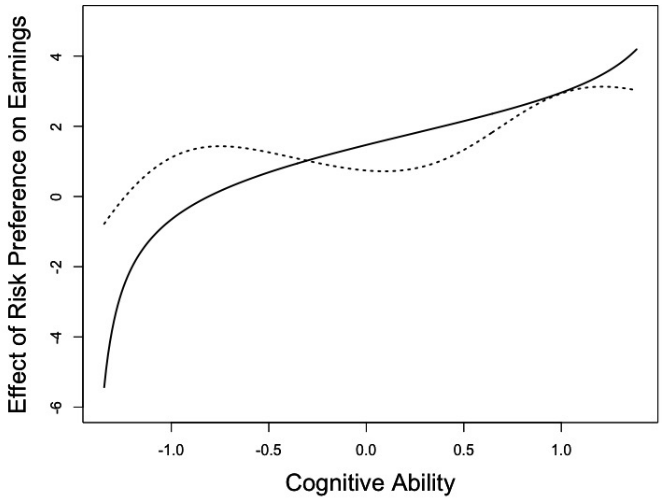

FIGURE 6. Plot of $\hat{\beta}_1$ between the 10^{th} and 90^{th} percentile of the ASVAB test score from the regression in (5.2) using the NLSY79 dataset. Note that the y-axis has been transformed to give the effect of risk preference on earnings as a percent, i.e. a value of 2 indicates that a one standard deviation increase in risk preference leads to a 2% increase in earnings. The solid line gives the estimate using the estimator of this paper when the measurement error is assumed unknown but symmetric. The dashed line denotes the estimate of Li *et al.* (2002) using the mean of the ASVAB test scores as if it is true cognitive ability.

Our analysis is agnostic with regards to the causal channels through which risk preferences affect wages. However, the previous literature sheds light on some potential mechanisms. Skriabikova, Dohmen and Kriechel (2014) investigate the relationship between risk aversion and self-employment and show that those with a higher preference for risk are more likely to be self-employed which may result in higher earnings (Levine and Rubinstein, 2017). Jaeger *et al.* (2010) find that those with higher risk preferences are also more likely to migrate; this greater flexibility imposes less constraints on a worker when attempting to maximise wages. Finally, Bonin *et al.* (2007) show that those with a higher preference for risk sort themselves into occupations with higher wage risk. Moreover, these jobs are associated with higher average wages.

6. Concluding remarks

In this paper, we develop an estimator for the varying coefficient model when the covariate in the coefficient functions is contaminated with classical measurement error. Using deconvolution kernel methods, the estimator is constructed following a similar two-step procedure as used in Li *et al.* (2002) for the error-free case. We study three separate cases based on knowledge of the error density. In the first, the error density is assumed known, this is then relaxed to allow for an unknown but symmetric error distribution where a repeated measurement of the noisy covariate is available. Finally, we allow for this density to potentially be asymmetric and again assume that a repeated measure is available. Furthermore, in each case we consider both ordinary smooth and supersmooth measurement error.

In all cases, the proposed estimator is shown to be consistent and asymptotically normally distributed. In particular, when the error density is unknown and symmetric, we use the approach of Delaigle, Hall and Meister (2008) to estimate the error characteristic function. While for the unknown and asymmetric case we use the approach of Li and Vuong (1998) based on Kotlarski's (1967) identity. Although this approach significantly increases the applicability of the estimator, we show that it does reduce the convergence rate.

A novel bandwidth selection procedure is proposed based on the out-of-bag prediction error and the finite sample performance of this method and the estimator in general is investigated by Monte Carlo simulation. Finally, we apply our estimator to study how the effect of risk preferences on earnings is affected by cognitive ability. In this setting, we show that accounting for measurement error is indeed important.

Throughout this paper, we focus on the case where a single mismeasured covariate W^* enters the coefficient functions β to keep the notation simple. The proposed method, however, can easily be adapted to the multivariate case. In particular, when β is a function of a set of covariates (W^*, Z) with $W^* \in \mathbb{R}^{d_w}$ and $Z \in \mathbb{R}^{d_z}$, i.e.,

$$Y = X'\beta(W^*, Z) + U, \qquad E[U|X, W^*, Z] = 0, \tag{6.1}$$

we wish to estimate $\beta(w^*, z)$ at a given point $(w^{*\prime}, z')' \in \mathbb{R}^{d_w+d_z}$ using an i.i.d. sample of (Y, X, W, Z), where W is a noisy measurement of W^* generated by $W = W^* + \tilde{\epsilon}$ and $\tilde{\epsilon} = (\epsilon_1, \ldots, \epsilon_{d_w})$ is mutually independent and independent of W^*. Similar to the case when β is a univariate function of a scalar W^*, by premultiplying (6.1) by X and taking the conditional expectation, the object of interest $\beta(w^*, z)$ can be written as

$$\beta(w^*, z) = M_{XX}(w^*, z)^{-1} M_{XY}(w^*, z),$$

where $M_{XX}(w^*, z) = E[XX'|W^* = w^*, Z = z]$ and $M_{XY}(w^*, z) = E[XY|W^* = w^*, Z = z]$. The conditional moments on the right hand side can be estimated by

$$\begin{aligned}
\hat{M}_{XX}(w^*, z) &= \frac{\sum_{j=1}^{n} X_j X_j' \mathbb{K}\left(\frac{w^*-W_j}{a_n}\right) L\left(\frac{z-Z_j}{b_n}\right)}{\sum_{j=1}^{n} \mathbb{K}\left(\frac{w^*-W_j}{a_n}\right) L\left(\frac{z-Z_j}{b_n}\right)}, \\
\hat{M}_{XY}(w^*, z) &= \frac{\sum_{j=1}^{n} X_j Y_j \mathbb{K}\left(\frac{w^*-W_j}{a_n}\right) L\left(\frac{z-Z_j}{b_n}\right)}{\sum_{j=1}^{n} \mathbb{K}\left(\frac{w^*-W_j}{a_n}\right) L\left(\frac{z-Z_j}{b_n}\right)},
\end{aligned}$$

respectively, where a_n and b_n are bandwidths, L is an ordinary kernel function, and $\mathbb{K}$ is a deconvolution kernel function defined by $\mathbb{K}(x) = \prod_{l=1}^{d_w} \mathbb{K}_l(x_l)$ with

$$\mathbb{K}_l(x_l) = \frac{1}{2\pi} \int e^{-\mathrm{i}tx_l} \frac{K^{\mathrm{ft}}(t)}{f_{\epsilon_l}^{\mathrm{ft}}(t/a_n)} dt,$$

and $K : \mathbb{R} \to \mathbb{R}$ is an ordinary univariate kernel. Based on these deconvolution estimators for the conditional moments, $\beta(w^*, z)$ can be estimated by

$$\hat{\beta}(w^*, z) = \left[\sum_{j=1}^{n} X_j X_j' \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right) L\left(\frac{z - Z_j}{b_n}\right)\right]^{-1} \sum_{j=1}^{n} X_j Y_j \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right) L\left(\frac{z - Z_j}{b_n}\right).$$

We expect that analogous results to our main theorems can be established for this estimator as well.

Finally, it is interesting to study optimal convergence rates for our estimation problems on $\beta(\cdot)$ and to develop adaptive bandwidth selection procedures. However, this is a substantial challenge even for the conventional nonparametric deconvolution regression (see, e.g., Chichignoud *et al.*, 2017, for the case where the distributions of the regression error term and measurement error are completely known), and we leave such extensions for future research.

Appendix A. Proofs for Section 2

Notation. For the results in Section 2, define

$$S_n = \frac{1}{na_n}\sum_{j=1}^n X_j X_j' \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right), \quad T_n = \frac{1}{n}\sum_{j=1}^n (\xi_j - E[\xi_j]), \quad B_n = E[\xi_j], \tag{A.1}$$

$$\xi_j = \frac{1}{a_n} V_j \mathbb{K}\left(\frac{w^* - W_j}{a_n}\right), \quad V_j = X_j[U_j + X_j'\{\beta(W_j^*) - \beta(w^*)\}]. \tag{A.2}$$

Then the estimator $\hat{\beta}(w^*)$ can be written as

$$\hat{\beta}(w^*) = \beta(w^*) + S_n^{-1}(T_n + B_n). \tag{A.3}$$

To simplify the notation, hereafter we suppress the dependence of S_n, T_n, B_n, ξ_j, and V_j on w^*. Also let $\xi_j^{k_1}$ and $V_j^{k_1}$ denote the k_1-th element of ξ_j and V_j, respectively, and $S_n^{k_1,k_2}$ and S^{k_1,k_2} denote the (k_1, k_2)-th element of S_n and S, respectively, for $k_1, k_2 = 1, \ldots, k$, where S is defined in Theorem 1.

A.1. **Proof of Theorem 1 (i) .** By (A.3), we have $|\hat{\beta}(w^*) - \beta(w^*)|^2 \leq 2\{\lambda_{\min}(S_n)\}^{-2}(|T_n|^2 + |B_n|^2)$. Note that $\lambda_{\min}(S_n) \geq \inf_{|\omega|=1} \omega'(S_n - S)\omega + \lambda_{\min}(S)$ and $\lambda_{\min}(S) > 0$ (Assumption M (2)). Thus, we obtain

$$|\hat{\beta}(w^*) - \beta(w^*)|^2 = O_p(|T_n|^2 + |B_n|^2), \tag{A.4}$$

if we can show

$$S_n \xrightarrow{p} S. \tag{A.5}$$

To show (A.5), note that for each $k_1, k_2 = 1, \ldots, k$,

$$\begin{aligned} E[S_n^{k_1,k_2}] &= a_n^{-1} E\left[X_{k_1} X_{k_2} \mathbb{K}\left(\frac{w^* - W}{a_n}\right)\right] = a_n^{-1} E\left[X_{k_1} X_{k_2} K\left(\frac{w^* - W^*}{a_n}\right)\right] \\ &= \int \{E[X_{k_1} X_{k_2} | W^* = \cdot] f_{W^*}(\cdot)\}(w^* - a_n u) K(u) du = S^{k_1,k_2} + O(a_n^p), \end{aligned} \tag{A.6}$$

where the second step follows by Lemma 1 as in Appendix D with $g(W^*) = E[X_{k_1} X_{k_2} | W^*]$ and last step follows by the smoothness of $E[X_{k_1} X_{k_2} | W^* = \cdot]$ and $f_{W^*}(\cdot)$ (Assumption M (2)) and properties of the p-th order kernel function K (Assumption M (3)). Also note that for any $k_1, k_2 = 1, \ldots, k$,

$$\begin{aligned} Var(S_n^{k_1,k_2}) &\leq \frac{1}{na_n^2} E\left[X_{k_1}^2 X_{k_2}^2 \mathbb{K}^2\left(\frac{w^* - W}{a_n}\right)\right] \\ &= \frac{1}{na_n^2} \iint \mathbb{K}^2\left(\frac{w^* - u - v}{a_n}\right) \{E[X_{k_1}^2 X_{k_2}^2 | W^* = \cdot] f_{W^*}(\cdot)\}(u) f_\epsilon(v) du dv \\ &= \frac{1}{na_n} \int \mathbb{K}^2(\tilde{u}) \left\{\int \{E[X_{k_1}^2 X_{k_2}^2 | W^*] f_{W^*}\}(w^* - v - a_n \tilde{u}) f_\epsilon(v) dv\right\} d\tilde{u} \\ &= O\left(\frac{1}{na_n} \int \mathbb{K}^2(\tilde{u}) d\tilde{u}\right) = O\left(\frac{1}{na_n}\left(\inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}\right), \end{aligned} \tag{A.7}$$

where the third step follows by the change of variables $\tilde{u} = \frac{w^* - u - v}{a_n}$, the fourth step follows by the boundedness of $E[X_{k_1}^2 X_{k_2}^2 | W^*]$ and f_{W^*} (Assumption M (2)), and the last step follows by

Lemma 2 as in Appendix D. Under Assumption O (1), (A.7) implies

$$Var(S_n^{k_1,k_2}) = O(n^{-1}a_n^{-(1+2\alpha)}). \tag{A.8}$$

Thus, (A.5) follows by (A.6) and (A.8) combined with $a_n \to 0$ and $na_n^{1+2\alpha} \to \infty$ (Assumption O (2)).

It remains to characterize the stochastic orders of $|T_n|^2$ and $|B_n|^2$. For $|T_n|^2$, note that for $k_1 = 1, \ldots, k$,

$$\begin{aligned} E|\xi_j^{k_1}|^2 &= a_n^{-2} \iint \mathbb{K}^2\left(\frac{w^* - u - v}{a_n}\right)\{E[|V_j^{k_1}|^2|W^* = \cdot]f_{W^*}(\cdot)\}(u)f_\epsilon(v)dudv \\ &= \frac{1}{a_n}\int \mathbb{K}^2(\tilde{u})\left\{\int \{E[|V_j^{k_1}|^2|W^* = \cdot]f_{W^*}(\cdot)\}(w^* - v - a_n\tilde{u})f_\epsilon(v)dv\right\}d\tilde{u} \\ &= O\left(\frac{1}{na_n}\int \mathbb{K}^2(\tilde{u})d\tilde{u}\right) = O\left(\frac{1}{na_n}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\text{ft}}(t)|\right)^{-2}\right), \end{aligned} \tag{A.9}$$

where the second step follows by the change of variables $\tilde{u} = \frac{w^*-u-v}{a_n}$, the third step follows by the definition of V_j and the boundedness of $E[X_{k_1}^2X_{k_2}^2|W^*]$, $E[U^2|X,W^*]$, β, and f_{W^*} (Assumption M (2)), and the last step follows by Lemma 2. Then, by $|T_n|^2 = O_p\left(n^{-1}\max_{1\le k_1\le k} E|\xi_j^{k_1}|^2\right)$, (A.9) implies

$$|T_n|^2 = O\left(\frac{1}{na_n}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\text{ft}}(t)|\right)^{-2}\right), \tag{A.10}$$

which under Assumption O (1) gives

$$|T_n|^2 = O_p(n^{-1}a_n^{-(1+2\alpha)}). \tag{A.11}$$

For $|B_n|^2$, note that $|B_n|^2 \le k\max_{1\le k_1\le k}|E[\xi_j^{k_1}]|^2$, and for $k_1 = 1, \ldots, k$,

$$\begin{aligned} E[\xi_j^{k_1}] &= a_n^{-1}E\left[V^{k_1}\mathbb{K}\left(\frac{w^* - W}{a_n}\right)\right] = a_n^{-1}E\left[V^{k_1}K\left(\frac{w^* - W^*}{a_n}\right)\right] \\ &= \int\{E[V^{k_1}|W^* = \cdot]f_{W^*}(\cdot)\}(w^* - a_nu)K(u)du = O(a_n^p), \end{aligned} \tag{A.12}$$

where the second step follows by Lemma 1 with $g(W^*) = E[V^{k_1}|W^*]$ and last step follows by the definition of V, the smoothness of $\beta(\cdot)$, $f_{W^*}(\cdot)$, and $E[X_{k_1}X_{k_2}|W^* = \cdot]$ (Assumption M (2)) and properties of the p-th order kernel function K (Assumption M (3)). Then, it follows

$$|B_n|^2 = O(a_n^{2p}). \tag{A.13}$$

Combining (A.4), (A.11), and (A.13), the conclusion follows.

A.2. **Proof of Theorem 1 (ii) .** By (A.3) and (A.5), it is sufficient to establish the asymptotic normality of T_n and the asymptotic negligibility of B_n. The asymptotic negligibility of B_n (i.e., $\sqrt{na_n^{1+2\alpha}}B_n \to 0$) is an immediate result of (A.12) and Assumption O (5). So, we focus on the asymptotic normality of T_n, i.e.,

$$\sqrt{na_n^{1+2\alpha}}T_n \xrightarrow{d} N(0, \Sigma). \tag{A.14}$$

By the Cramér–Wold device, (A.14) is equivalent to

$$\sqrt{na_n^{1+2\alpha}}\omega' T_n \overset{d}{\to} N(0, \omega'\Sigma\omega) \tag{A.15}$$

for each $\omega \in \mathbb{R}^k$ with $|\omega| = 1$. We first establish the standardised version of (A.15), i.e.

$$\frac{\omega' T_n}{\sqrt{Var(\omega' T_n)}} \overset{d}{\to} N(0, 1). \tag{A.16}$$

For (A.16), it suffices to check the Lyapunov condition for some $\eta > 0$, i.e.,

$$\frac{E|\omega'\xi_j - E[\omega'\xi_j]|^{2+\eta}}{n^{\eta/2}\{Var(\omega'\xi_j)\}^{1+\eta/2}} \to 0. \tag{A.17}$$

For the numerator of (A.17), we have

$$\begin{aligned}
& E|\omega'\xi_j - E[\omega'\xi_j]|^{2+\eta} \leq 2^{1+\eta} E|\omega'\xi_j|^{2+\eta} \\
= \;& O\left(a_n^{-(2+\eta)} \iint \left|\mathbb{K}\left(\frac{w^* - u - v}{a_n}\right)\right|^{2+\eta} \{E[|\omega' V|^{2+\eta}|W^* = \cdot] f_{W^*}(\cdot)\}(u) f_\epsilon(v) du dv\right) \\
= \;& O\left(a_n^{-(1+\eta)} \int |\mathbb{K}(\tilde{u})|^{2+\eta} \left\{\int \{E[|\omega' V|^{2+\eta}|W^* = \cdot] f_{W^*}(\cdot)\}(w^* - v - a_n\tilde{u}) f_\epsilon(v) dv\right\} d\tilde{u}\right) \\
= \;& O\left(a_n^{-(1+\eta)} \int |\mathbb{K}(\tilde{u})|^{2+\eta} d\tilde{u}\right) = O\left(a_n^{-(1+\eta)} \left(\inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|\right)^{-(2+\eta)}\right),
\end{aligned} \tag{A.18}$$

where the first step follows by Jensen's inequality, the second step follows by the definition of ξ_j, the third step follows by the change of variables $\tilde{u} = \frac{w^* - u - v}{a_n}$, the fourth step follows by the definition of V and the boundedness of $E[|X_{k_1}|^{2+\eta}|W^*]$, $E[|U|^{2+\eta}|W^*]$, β, and f_{W^*} (Assumption M (4)), and the last equality follows by Lemma 2. Under Assumption O (1), (A.18) implies

$$E|\omega'\xi_j - E[\omega'\xi_j]|^{2+\eta} = O(a_n^{-1-2\alpha-\eta(1+\alpha)}). \tag{A.19}$$

For the denominator of (A.17), for any $\omega_1, \omega_2 \in \mathbb{R}^k$ with $|\omega_1| = 1$ and $|\omega_2| = 1$, we have

$$\begin{aligned}
& E[\omega_1'\xi_j\xi_j'\omega_2] = a_n^{-2} \iint \left|\mathbb{K}\left(\frac{w^* - u - v}{a_n}\right)\right|^2 \{E[|\omega_1' V V' \omega_2|\,|W^* = \cdot] f_{W^*}(\cdot)\}(u) f_\epsilon(v) du dv \\
= \;& a_n^{-2} \int \left|\mathbb{K}\left(\frac{w^* - \tilde{u}}{a_n}\right)\right|^2 \left\{\int \{E[|\omega_1' V V' \omega_2|\,|W^* = \cdot] f_{W^*}(\cdot)\}(\tilde{u} - v) f_\epsilon(v) dv\right\} d\tilde{u} \\
= \;& a_n^{-(1+2\alpha)} C \int \{E[|\omega_1' V V' \omega_2|\,|W^* = \cdot] f_{W^*}(\cdot)\}(w^* - v) f_\epsilon(v) dv (1 + o(1)),
\end{aligned} \tag{A.20}$$

where the first equality follows by the definition of ξ_j, the second equality follows by the change of variables $\tilde{u} = u + v$, and the last equality follows by the continuity of $\int \{E[|\omega_1' V V' \omega_2|\,|W^* = \cdot] f_{W^*}(\cdot)\}(\tilde{u} - v) f_\epsilon(v) dv$ as a function of $\tilde{u}$ for any $\omega_1, \omega_2 \in \mathbb{R}^k$ with $|\omega_1| = 1$ and $|\omega_2| = 1$ (Assumption M (2)), Assumption O (4), Fan (1991, Lemma 2.1), Parseval's identity, and the definition of C.

Then, noting that $Var(\omega'\xi_j)$ is dominated by $E|\omega'\xi_j|^2$, (A.20) implies

$$\{Var(\omega'\xi_j)\}^{-(1+\eta/2)} = O(a_n^{(1+2\alpha)(1+\eta/2)}), \tag{A.21}$$

Therefore, (A.19) and (A.21) imply that (A.17) is satisfied if $na_n \to \infty$ as $n \to \infty$, which holds under Assumption O (2).

To show (A.14), besides (A.16), we also need

$$na_n^{1+2\alpha} Var(T_n) \to \Sigma. \tag{A.22}$$

For (A.22), decompose $na_n^{1+2\alpha} Var(T_n) = \Sigma_{n,1} + \Sigma_{n,2}$, where

$$\Sigma_{n,1} = a_n^{1+2\alpha} E[\xi_j \xi_j'], \qquad \Sigma_{n,2} = -a_n^{1+2\alpha} E[\xi_j] E[\xi_j]'.$$

By (A.12) and $a_n \to 0$ (Assumption O (2)), it holds $\Sigma_{n,2} = o(1)$. By (A.20), the (k_1, k_2)-element of $\Sigma_{n,1}$ converges to $C \int \{E[|V^{k_1} V^{k_2}| \,|W^* = \cdot] f_{W^*}(\cdot)\}(w^* - v) f_\epsilon(v) dv$, which implies $\Sigma_{n,1} \to \Sigma$, and (A.22) follows.

A.3. **Proof of Theorem 2 (i) .** First, note that under Assumption S (1), (A.7) implies

$$Var(S_n^{k_1,k_2}) = O(n^{-1} a_n^{-1} e^{2\mu a_n^{-\gamma}}). \tag{A.23}$$

Then, (A.4) follows by (A.5), which is implied by (A.6), (A.23), $a_n \to 0$, and $na_n e^{-2\mu a_n^{-\gamma}} \to \infty$ (Assumption S (2)).

Also note that under Assumption S (1), (A.10) implies

$$|T_n|^2 = O(n^{-1} a_n^{-1} e^{2\mu a_n^{-\gamma}}). \tag{A.24}$$

Combining (A.4), (A.13), and (A.24), the conclusion follows.

A.4. **Proof of Theorem 2 (ii).** For the numerator of (A.16), under Assumption S (1), (A.18) implies

$$E|\omega'\xi_j - E[\omega'\xi_j]|^{2+\eta} = O(a_n^{-(1+\eta)} e^{(2+\eta)\mu a_n^{-\gamma}}). \tag{A.25}$$

To establish the lower bound of the denominator of (A.16), we leverage the approach of van Es and Uh (2004) and Uh (2003). In particular, define $c(a_n) = \int_0^1 e^{\mu (s/a_n)^\gamma} ds$, $D_{n,j} = \frac{W_j - w^*}{a_n}$ mod 2π, and $\Psi_\theta(t) = (1 - it)^{-(1+\theta)}$. By Uh (2003, Theorem 4.7), under Assumptions S (3) and (4), we have

$$\mathbb{K}\left(\frac{w^* - W_j}{a_n}\right) = \frac{A\Gamma(\theta+1) a_n^{\gamma\theta} c(a_n)}{\pi(\mu\gamma)^\theta c^{\mathrm{ss}}} \left\{ \begin{array}{c} \cos(D_{n,j}) \Re \Psi_\theta\left(\frac{(W_j - w^*) a_n^{\gamma-1}}{\mu\gamma}\right) \\ -\sin(D_{n,j}) \Im \Psi_\theta\left(\frac{(W_j - w^*) a_n^{\gamma-1}}{\mu\gamma}\right) \end{array} \right\} + O_p(a_n^{\gamma(1+\theta)} c(a_n)),$$

where $\Re f$ and $\Im f$ denote the real and imaginary parts of a complex-valued function f, respectively. So, $Var(\omega'\xi_j)$ is dominated by $a_n^{2(\gamma\theta-1)} c(a_n)^2 Var(\omega'\psi_j)$ with

$$\psi_j = V_j \left\{ \cos(D_{n,j}) \Re \Psi_\theta\left(\frac{(W_j - w^*) a_n^{\gamma-1}}{\mu\gamma}\right) - \sin(D_{n,j}) \Im \Psi_\theta\left(\frac{(W_j - w^*) a_n^{\gamma-1}}{\mu\gamma}\right) \right\},$$

which, together with van Es and Uh (2004, Lemma 2.1), gives

$$\{Var(\omega'\xi_j)\}^{-(1+\eta/2)} = O\left(a_n^{-(\gamma\theta+\gamma-1)(2+\eta)} e^{-(2+\eta)\mu a_n^{-\gamma}} \{Var(\omega'\psi_j)\}^{-(1+\eta/2)}\right). \tag{A.26}$$

By (A.26), for the lower bound of the denominator of (A.16), we can focus on the lower bound of $Var(\omega'\psi_j)$, for which it is sufficient to derive the lower bound of $E|\omega'\psi_j|^2$. Using the law of iterated expectation, we have

$$E|\omega'\psi_j|^2 = E\left[E[|\omega' V_j|^2|W_j] \left\{ \begin{array}{c} \cos(D_{n,j})\Re\Psi_\theta\left(\frac{(W_j-w^*)a_n^{\gamma-1}}{\mu\gamma}\right) \\ -\sin(D_{n,j})\Im\Psi_\theta\left(\frac{(W_j-w^*)a_n^{\gamma-1}}{\mu\gamma}\right) \end{array} \right\}^2 \right]. \tag{A.27}$$

Let D be a random variable that is uniformly distributed on $[0, 2\pi]$ and is independent of W. Following van Es and Uh (2004), we consider three separate cases based on the value of γ. In particular, using van Es and Uh (2004, Lemma 3.1), (A.27) implies

$$E|\omega'\psi_j|^2 = \begin{cases} a_n^{2(1-\gamma)(1+\theta)} E\left|\omega' V_j \left\{ \begin{array}{c} \cos(D_j)\Re\left(\frac{W_j-w^*}{\mathrm{i}\mu\gamma}\right)^{-(1+\theta)} \\ -\sin(D_j)\Im\left(\frac{W_j-w^*}{\mathrm{i}\mu\gamma}\right)^{-(1+\theta)} \end{array} \right\}\right|^2 (1+o(1)) & \text{if } 1/3 < \gamma < 1 \\ E\left|\omega' V_j \left\{ \begin{array}{c} \cos(D_j)\Re\Psi_\theta\left(\frac{W_j-w^*}{\mu\gamma}\right) \\ -\sin(D_j)\Im\Psi_\theta\left(\frac{W_j-w^*}{\mu\gamma}\right) \end{array} \right\}\right|^2 (1+o(1)) & \text{if } \gamma = 1 \\ E[|\omega' V_j|^2 \cos^2(D_j)](1+o(1)) & \text{if } 1 < \gamma \le 2 \end{cases},$$

which together with (A.26) gives

$$\{Var(\omega'\xi_j)\}^{-(1+\eta/2)} = \begin{cases} O\left(a_n^{-(2+\theta)(2+\eta)} e^{-(2+\eta)\mu a_n^{-\gamma}}\right) & \text{if } 1/3 < \gamma < 1 \\ O\left(a_n^{-(\gamma\theta+\gamma-1)(2+\eta)} e^{-(2+\eta)\mu a_n^{-\gamma}}\right) & \text{if } 1 \le \gamma \le 2 \end{cases}. \tag{A.28}$$

Therefore, (A.25) and (A.28) imply that (A.17) is satisfied if $na_n^{2\gamma(\theta+1)(1+2/\eta)-2} \to \infty$ for some $\eta > 0$ in the case when $1 \le \gamma \le 2$ or if $na_n^{2(3+\theta)(1+2/\eta)-2} \to \infty$ for some $\eta > 0$ in the case when $1/3 < \gamma < 1$, which holds true under Assumption S (2).

The conclusion then follows by

$$\frac{n^{1/2}\omega' B_n}{\sqrt{Var(\omega'\xi_j)}} = o_p(1),$$

for which we combine (A.13) and (A.28) with $na_n^{2p-2(2+\theta)}e^{-2\mu a_n^{-\gamma}} \to 0$ if $1/3 < \gamma < 1$ and $na_n^{2p-2\gamma(\theta+1)+2}e^{-2\mu a_n^{-\gamma}} \to 0$ if $1 \le \gamma \le 2$ (Assumption S (5)).

Appendix B. Proofs for Section 3

Notation. For the results in Section 3, we introduce $\hat{\mu}_s(t) = \frac{1}{n}\sum_{l=1}^n \mu_{s,l}(t)$ and $\mu_s(t) = E[\mu_{s,1}(t)]$ with $\mu_{s,l}(t) = \cos\{t(W_l - W_l^r)\}$, which give $\hat{f}^{\text{ft}}_{\epsilon,s}(t) = |\hat{\mu}_s(t)|^{1/2}$ and $f^{\text{ft}}_\epsilon(t) = |\mu_s(t)|^{1/2}$, i.e., $\hat{\mathbb{K}}_s$ and $\mathbb{K}$ are functionals of $\hat{\mu}_s$ and μ_s, respectively. Define

$$\hat{\Pi}_s(t) = \frac{1}{n}\sum_{l=1}^n \Pi_{s,l}(t), \quad \Pi_{s,l}(t) = \frac{\mu_s(t) - \mu_{s,l}(t)}{2\mu_s(t)},$$

$$\hat{\Pi}_s^{\text{res}}(t) = \frac{(2|\hat{\mu}_s(t)|^{1/2} + |\mu_s(t)|^{1/2})(|\hat{\mu}_s(t)|^{1/2} - |\mu_s(t)|^{1/2})^2}{\mu_s(t)|\hat{\mu}_s(t)|^{1/2}}.$$

By expanding $\hat{\mu}_s$ around μ_s, we obtain

$$\hat{\mathbb{K}}_s(u) = \mathbb{K}(u) + \mathbb{A}_s(u) + \mathbb{R}_s(u), \tag{B.1}$$

where

$$\mathbb{A}_s(u) = \frac{1}{2\pi}\int e^{-itu}\frac{K^{\rm ft}(t)}{f_\epsilon^{\rm ft}(t/a_n)}\hat{\Pi}_s(t/a_n)dt,$$
$$\mathbb{R}_s(u) = \frac{1}{2\pi}\int e^{-itu}\frac{K^{\rm ft}(t)}{f_\epsilon^{\rm ft}(t/a_n)}\hat{\Pi}_s^{\rm res}(t/a_n)dt.$$

Here $\mathbb{A}_s(u)$ is the Fréchet derivative of $\hat{\mathbb{K}}_s(u)$ as a functional of $\hat{\mu}_s$ at μ_s in the direction of $\hat{\mu}_s - \mu_s$. Also note that $\hat{\Pi}_s^{\rm res}(t)$ is of a higher order than $\hat{\Pi}_s(t)$. So, the remainder $\mathbb{R}_s(u)$ should be dominated by $\mathbb{A}_s(u)$ asymptotically.

Define

$$\hat{S}_{n,s} = \frac{1}{na_n}\sum_{j=1}^n X_jX_j'\left\{\mathbb{A}_s\left(\frac{w^*-W_j}{a_n}\right)+\mathbb{R}_s\left(\frac{w^*-W_j}{a_n}\right)\right\},$$
$$\hat{A}_{n,s} = \frac{1}{n}\sum_{j=1}^n\{\xi_{s,a,j}-E[\xi_{s,a,j}]\},\quad \hat{B}_{n,s}=E[\xi_{s,a,j}],\quad \hat{R}_{n,s}=\frac{1}{n}\sum_{j=1}^n\xi_{s,r,j},$$

where $\xi_{s,a,j} = \frac{1}{a_n}V_j\mathbb{A}_s\left(\frac{w^*-W_j}{a_n}\right)$, $\xi_{s,r,j} = \frac{1}{a_n}V_j\mathbb{R}_s\left(\frac{w^*-W_j}{a_n}\right)$, and V_j is defined in (A.2). Then, using the approximation result for $\hat{\mathbb{K}}_s$ as in (B.1), we have

$$\tilde{\beta}_s(w^*) = \beta(w^*) + (S_n+\hat{S}_{n,s})^{-1}(T_n+\hat{A}_{n,s}+B_n+\hat{B}_{n,s}+\hat{R}_{n,s}), \tag{B.2}$$

where S_n, T_n, and B_n are defined in (A.1). Also, let $\hat{S}_{n,s}^{k_1,k_2}$ denote the (k_1,k_2)-th element of $\hat{S}_{n,s}$, $\xi_{s,a,j}^{k_1}$ and $\xi_{s,r,j}^{k_1}$ separately denote the k_1-element of $\xi_{s,a,j}$ and $\xi_{s,r,j}$ for $k_1,k_2=1,\ldots,k$, and $E_j[\cdot]=E[\cdot|W_j^*,X_j,Y_j,\epsilon_j,\epsilon_j^r]$ for $j=1,\ldots,n$. The proofs of Theorems 3 and 4 follow along similar lines as Theorems 1 and 2 respectively, as such, we only explain the parts that differ.

B.1. **Proof of Theorem 3 (i).** By (B.2), (A.5), and $\lambda_{\min}(S)>0$ (Assumption M (2)), we have

$$|\tilde{\beta}_s(w^*)-\beta(w^*)|^2 = O_p(|T_n|^2+|B_n|^2+|\hat{A}_{n,s}+\hat{B}_{n,s}|^2+|\hat{R}_{n,s}|^2), \tag{B.3}$$

if we can show

$$\hat{S}_{n,s} \xrightarrow{p} 0. \tag{B.4}$$

To show (B.4), note that for any $k_1,k_2=1,\ldots,k$,

$$|\hat{S}_{n,s}^{k_1,k_2}| \leq \frac{1}{na_n}\sum_{j=1}^n\left|X_{k_1,j}X_{k_2,j}\left\{\mathbb{A}_s\left(\frac{w^*-W_j}{a_n}\right)+\mathbb{R}_s\left(\frac{w^*-W_j}{a_n}\right)\right\}\right|$$
$$\leq \frac{1}{na_n}\sum_{j=1}^n|X_{k_1,j}X_{k_2,j}|\int|K^{\rm ft}(t)|\frac{|\hat{\Pi}_s(t/a_n)|+|\hat{\Pi}_s^{\rm res}(t/a_n)|}{|f_\epsilon^{\rm ft}(t/a_n)|}dt$$
$$= O_p\left(\frac{\sup_{|t|\leq a_n^{-1}}|\hat{\Pi}_s(t)|+\sup_{|t|\leq a_n^{-1}}|\hat{\Pi}_s^{\rm res}(t)|}{a_n\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\rm ft}(t)|}\right), \tag{B.5}$$

where the second step follows by definitions of $\mathbb{A}_s$ and $\mathbb{R}_s$, and the last step uses the fact that $K^{\rm ft}$ is supported on $[-1,1]$ (Assumption M (3)). By $n^{-1/2}a_n^{-2\alpha}\log(1/a_n)^{1/2}\to 0$ (Assumption OS (1)) and Lemma 4 as in Appendix D, under Assumption O (1), (B.5) implies

$$|\hat{S}_{n,s}| = O_p\left(n^{-1/2}a_n^{-(1+3\alpha)}\log(1/a_n)+n^{-1}a_n^{-(1+5\alpha)}\log(1/a_n)^2\right), \tag{B.6}$$

and (B.4) follows by (B.6) and $n^{-1/2}a_n^{-(1+3\alpha)}\log(1/a_n)\to 0$ (Assumption OS (1)).

Since stochastic orders of $|T_n|^2$ and $|B_n|^2$ are obtained in Appendix A.1, it remains to characterize the orders of $|\hat{R}_{n,s}|^2$ and $|\hat{A}_{n,s}+\hat{B}_{n,s}|^2$. For $|\hat{R}_{n,s}|^2$, note that

$$\begin{aligned}|\hat{R}_{n,s}| &\leq \sqrt{k}\max_{1\leq k_1\leq k}\left\{\frac{1}{n}\sum_{j=1}^{n}|\xi_{s,r,j}^{k_1}|\right\}\leq\sqrt{k}\max_{1\leq k_1\leq k}\left\{\frac{1}{na_n}\sum_{j=1}^{n}|V_j^{k_1}|\int|K^{\mathrm{ft}}(t)|\frac{|\hat{\Pi}_s^{\mathrm{res}}(t/a_n)|}{|f_\epsilon^{\mathrm{ft}}(t/a_n)|}dt\right\}\\ &= O_p\left(\frac{\sup_{|t|\leq a_n^{-1}}|\hat{\Pi}_s^{\mathrm{res}}(t)|}{a_n\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|}\right), \qquad (B.7)\end{aligned}$$

where the second step follows by definitions of $\xi_{s,r,j}$ and $\mathbb{R}_s$, and the last step uses the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)). By $n^{-1/2}a_n^{-2\alpha}\log(1/a_n)\to 0$ (Assumption OS (1)) and Lemma 4, under Assumption O (1), (B.7) implies

$$|\hat{R}_{n,s}|^2 = O_p\left(n^{-2}a_n^{-2(1+5\alpha)}\log(1/a_n)^4\right). \qquad (B.8)$$

For $|\hat{A}_{n,s}+\hat{B}_{n,s}|^2$, note that

$$\begin{aligned}|\hat{A}_{n,s}+\hat{B}_{n,s}| &\leq \sqrt{k}\max_{1\leq k_1\leq k}\left\{\frac{1}{n}\sum_{j=1}^{n}|\xi_{s,a,j}^{k_1}|\right\}\leq\sqrt{k}\max_{1\leq k_1\leq k}\left\{\frac{1}{na_n}\sum_{j=1}^{n}|V_j^{k_1}|\int|K^{\mathrm{ft}}(t)|\frac{|\hat{\Pi}_s(t/a_n)|}{|f_\epsilon^{\mathrm{ft}}(t/a_n)|}dt\right\}\\ &= O_p\left(\frac{\sup_{|t|\leq a_n^{-1}}|\hat{\Pi}_s(t)|}{a_n\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|}\right), \qquad (B.9)\end{aligned}$$

where the second step follows by definitions of $\xi_{s,a,j}$ and $\mathbb{A}_s$, and the last step uses the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)). By Lemma 4, under Assumption O (1), (B.9) implies

$$|\hat{A}_{n,s}+\hat{B}_{n,s}|^2 = O_p\left(n^{-1}a_n^{-2(1+3\alpha)}\log(1/a_n)^2\right). \qquad (B.10)$$

Combining (A.11), (A.13), (B.8), and (B.10) with (B.3), the conclusion follows.

B.2. **Proof of Theorem 3 (ii) .** Define

$$T_{n,s}^* = \frac{2}{n}\sum_{j=1}^{n}\{\xi_j - E[\xi_j]+\hat{\xi}_{s,j}\},\quad \hat{T}_{n,s} = \binom{n}{2}^{-1}\sum_{j=1}^{n-1}\sum_{l=j+1}^{n}\{p_{s,j,l}+p_{s,l,j}\}\quad T_{n,s}^r = \frac{1}{n^2}\sum_{j=1}^{n}p_{s,j,j},$$

where $\hat{\xi}_{s,j} = E_j[\xi_{s,a,l,j}]$ for $l\neq j$ and $p_{s,j,l} = \xi_j + \xi_{s,a,j,l} - E[\xi_j + \xi_{s,a,j,l}]$ with

$$\xi_{s,a,j,l} = \frac{1}{2\pi a_n}\int V_j e^{-\mathrm{i}t\left(\frac{w^*-W_j}{a_n}\right)}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}\Pi_{s,l}(t/a_n)dt.$$

Decompose

$$T_n + \hat{A}_{n,s} = \frac{n-1}{2n}T_{n,s}^* + \frac{n-1}{2n}\{\hat{T}_{n,s}-T_{n,s}^*\} + T_{n,s}^r. \qquad (B.11)$$

To understand (B.11), note that $T_n + \hat{A}_{n,s} = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{l=1}^{n}p_{s,j,l}$ since $\xi_{a,j} = \frac{1}{n}\sum_{l=1}^{n}\xi_{a,j,l}$, $T_{n,s}^r$ characterizes the diagonal elements of $T_n+\hat{A}_{n,s}$, $\hat{T}_{n,s}$ characterizes the off-diagonal elements of $T_n+\hat{A}_{n,s}$ and is a second-order U-statistic with symmetric kernel $p_{s,j,l}+p_{s,l,j}$ and the Hájek projection $T_{n,s}^*$. Also note that $\hat{\xi}_{s,j}=0$ (as $\Pi_{s,j}(t/a_n)=0$) when f_ϵ is known, which implies

$\frac{n-1}{2n}T^*_{n,s} = T_n(1-1/n)$, i.e., $\frac{n-1}{2n}T^*_{n,s}$ is a natural extension of T_n to the case when $f^{\rm ft}_\epsilon$ is estimated by $\hat{f}^{\rm ft}_{\epsilon,s}$ since $\Pi_{s,j}(t/a_n)$ in $\xi_{s,j}$ characterizes the additional estimation error brought by $\hat{f}^{\rm ft}_{\epsilon,s}$.

So, the conclusion follows if we can establish the asymptotic normality of $T^*_{n,s}$ and show the asymptotic negligibility of $\hat{T}_{n,s} - T^*_{n,s}$, $T^r_{n,s}$, B_n, $\hat{B}_{n,s}$, and $\hat{R}_{n,s}$. Using the Cramér–Wold device, this is equivalent to establishing the asymptotic normality of $\omega' T^*_{n,s}$ and showing the asymptotic negligibility of $\omega'(\hat{T}_{n,s} - T^*_{n,s})$, $\omega' T^r_{n,s}$, $\omega' B_n$, $\omega' \hat{B}_{n,s}$, and $\omega' \hat{R}_{n,s}$ for each $\omega \in \mathbb{R}^k$ with $|\omega| = 1$.

For the asymptotic normality of $\omega' T^*_{n,s}$, we show

$$\frac{\omega' T^*_{n,s}}{\sqrt{Var(\omega' T^*_{n,s})}} \xrightarrow{d} N(0,1). \tag{B.12}$$

To this end, it suffices to check Lyapunov's condition for some $\eta > 0$, i.e.,

$$\frac{E|\omega'(\xi_j + \hat{\xi}_{s,j}) - E[\omega'\xi_j]|^{2+\eta}}{n^{\eta/2}\{Var(\omega'(\xi_j + \hat{\xi}_{s,j}))\}^{1+\eta/2}} \to 0. \tag{B.13}$$

For the numerator of (B.13), note that

$$\begin{aligned}
& E|\omega'(\xi_j + \hat{\xi}_{s,j}) - E[\omega'\xi_j]|^{2+\eta} \leq 2^{1+\eta}\{E|\omega'\hat{\xi}_{s,j}|^{2+\eta} + E|\omega'\xi_j - E[\omega'\xi_j]|^{2+\eta}\} \\
= \;& O\left(E\left| \frac{1}{2\pi a_n} \int E\left[\omega' V e^{-\mathrm{i}t\left(\frac{w^* - W^*}{a_n}\right)}\right] \Pi_{s,j}(t/a_n) K^{\rm ft}(t) dt \right|^{2+\eta} + E|\omega'\xi_j - E[\omega'\xi_j]|^{2+\eta} \right) \\
= \;& O\left(a_n^{-(2+\eta)} (E|\omega' V|)^{2+\eta} E\left[\left(\sup_{|t| \leq a_n^{-1}} |\Pi_{s,j}(t)| \right)^{2+\eta} \right] + E|\omega'\xi_j - E[\omega'\xi_j]|^{2+\eta} \right) \\
= \;& O\left(a_n^{-(2+\eta)} \left(\inf_{|t| \leq a_n^{-1}} |f^{\rm ft}_\epsilon(t)| \right)^{-2(2+\eta)} \right),
\end{aligned} \tag{B.14}$$

where the first step follows by Jensen's inequality, the third step follows from the fact that $K^{\rm ft}$ is supported on $[-1,1]$ (Assumption M (3)), and the last step follows by Lemma 5 as in Appendix D and (A.18). Under Assumption O (1), (B.14) implies

$$E|\omega'(\xi_j + \hat{\xi}_{s,j}) - E[\omega'\xi_j]|^{2+\eta} = O(a_n^{-(2+\eta)(1+2\alpha)}). \tag{B.15}$$

For the denominator of (B.13), note that $Var(\omega'(\xi_j + \hat{\xi}_{s,j})) = Var(\omega'(\xi_j + \xi^*_{s,j}))$ with

$$\xi^*_{s,j} = -\frac{1}{4\pi} \int \frac{E[V e^{\mathrm{i}tW^*}] K^{\rm ft}(ta_n)}{\{f^{\rm ft}_\epsilon(t)\}^2} \cos\{t(\epsilon_j - \epsilon^r_j)\} e^{-\mathrm{i}tw^*} dt. \tag{B.16}$$

By Assumption MS (4), for the lower bound of $Var(\omega'(\xi_j + \xi^*_{s,j}))$, it suffices to focus on $Var(\omega'\xi_j)$ and $Var(\omega'\xi^*_{s,j})$, which are dominated by $E|\omega'\xi_j|^2$ and $E|\omega'\xi^*_{s,j}|^2$, respectively. For $E|\omega'\xi^*_{s,j}|^2$, note that for some finite interval $I_s \subset \mathbb{R}$ containing $\pm w^*$, we have

$$\begin{aligned}
& E|\omega'\xi^*_{s,j}|^2 = \frac{1}{16\pi^2} \int \left| \int \frac{\{\omega' b(\cdot)\}^{\rm ft}(t) K^{\rm ft}(ta_n)}{\{f^{\rm ft}_\epsilon(t)\}^2} \cos(tu) e^{-\mathrm{i}tw^*} dt \right|^2 f_{\epsilon - \epsilon^r}(u) du \\
= \;& \frac{1}{64\pi^2} \int \left| \int \frac{\{\omega' b(\cdot)\}^{\rm ft}(t)}{\{f^{\rm ft}_\epsilon(t)\}^2} \{e^{-\mathrm{i}t(w^* - u)} + e^{-\mathrm{i}t(w^* + u)}\} K^{\rm ft}(ta_n) dt \right|^2 f_{\epsilon - \epsilon^r}(u) du \\
\geq \;& c_s \int_{u \in I_s} \left| \int \frac{\{\omega' b(\cdot)\}^{\rm ft}(t)}{\{f^{\rm ft}_\epsilon(t)\}^2} \{e^{-\mathrm{i}t(w^* - u)} + e^{-\mathrm{i}t(w^* + u)}\} K^{\rm ft}(ta_n) dt \right|^2 du,
\end{aligned} \tag{B.17}$$

where the first step follows from the definition of $\xi^*_{s,j}$, the second step uses $\cos(x) = \{e^{\mathrm{i}x}+e^{-\mathrm{i}x}\}/2$, and the last step follows by choosing some constant $c_s > 0$ such that $\inf_{u\in I} f_{\epsilon-\epsilon^r}(u) > 64\pi^2 c_s$, where such a c_s exists due to the compactness of I_s and the fact that f_ϵ is continuous and non-vanishing everywhere (Assumption MS (2)).

As $\frac{d}{dt}\left\{\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2}\right\} \in \mathbb{W}$ (Assumption MS (3)), Schennach (2004, Lemma 10) implies

$$\lim_{|u|\to\infty} (w^* \pm u) \int \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2} e^{-\mathrm{i}t(w^*\pm u)} dt = 0,$$

and it follows

$$\int_{u\in I^c} \left| \int \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2} \{e^{-\mathrm{i}t(w^*-u)} + e^{-\mathrm{i}t(w^*+u)}\} dt \right|^2 du = O\left(\int_{u\in I^c} \frac{1}{(w^{*2}-u^2)^2} du \right) = O(1). \tag{B.18}$$

Thus, for all n large enough and some constant $C_s > 0$, we have

$$\begin{aligned} E|\omega'\xi^*_{s,j}|^2 &\geq C_s \int \left| \int \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2} \left\{ e^{-\mathrm{i}t(w^*-u)} + e^{-\mathrm{i}t(w^*+u)} \right\} K^{\mathrm{ft}}(ta_n) dt \right|^2 du \\ &= 2\pi C_s \iint \left\{ \begin{array}{c} \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)\{\omega' b(\cdot)\}^{\mathrm{ft}}(-s) K^{\mathrm{ft}}(ta_n) K^{\mathrm{ft}}(-sa_n)}{\{f_\epsilon^{\mathrm{ft}}(t) f_\epsilon^{\mathrm{ft}}(-s)\}^2} e^{\mathrm{i}(s-t)w^*} \\ \times \frac{1}{2\pi} \int \left\{ e^{\mathrm{i}(t-s)u} + e^{\mathrm{i}(t+s)u} + e^{\mathrm{i}(-t-s)u} + e^{\mathrm{i}(-t+s)u} \right\} du \end{array} \right\} ds dt \\ &= 4\pi C_s \left\{ \int \left| \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2} K^{\mathrm{ft}}(ta_n) \right|^2 dt + \int \left(\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2} K^{\mathrm{ft}}(ta_n) \right)^2 e^{\mathrm{i}2tw^*} dt \right\} \\ &= 8\pi C_s \int \left| \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t) \cos(tw^*) K^{\mathrm{ft}}(ta_n)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2} \right|^2 dt, \end{aligned} \tag{B.19}$$

where the first step follows by (B.17) and (B.18), the third step follows by $\int \delta(x-b)f(x)dx = f(b)$ with the Dirac function $\delta(x) = \frac{1}{2\pi}\int e^{-\mathrm{i}tx} dx$, and the last step follows by the symmetry of $E[XX'|W^* = \cdot] f_{W^*}(\cdot)$, $\beta(\cdot)$, and $K(\cdot)$ (Assumption MS (2)) and the identity $1 + \cos(2x) = 2\cos^2(x)$.

Thus, (B.19) together with (A.21) implies

$$\begin{aligned} &\{Var(\omega'(\xi_j + \hat{\xi}_{s,j}))\}^{-(1+\eta/2)} \\ = \;& O\left(\left(\int \left| \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t) \cos(tw^*) K^{\mathrm{ft}}(ta_n)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2} \right|^2 dt + a_n^{-(1+2\alpha)} \right)^{-(1+\eta/2)} \right). \end{aligned} \tag{B.20}$$

Combining (B.15) and (B.20), (B.13) holds under Assumption OS (2), and (B.12) follows.

For $\omega'(\hat{T}_{n,s} - T^*_{n,s})$, by Ahn and Powell (1993, Lemma A.3), we have

$$\omega'(\hat{T}_{n,s} - T^*_{n,s}) = o_p(n^{-1/2}), \tag{B.21}$$

if

$$E|\omega'\{p_{s,j,l} + p_{s,l,j}\}|^2 = O(n). \tag{B.22}$$

To show (B.22), note that

$$E|\omega'\xi_{s,a,j,l}|^2 = a_n^{-2}E\left[\left(\int |\Pi_{s,l}(t/a_n)|\frac{|K^{\mathrm{ft}}(t)|}{|f_\epsilon^{\mathrm{ft}}(t/a_n)|}dt\right)^2\right]E|\omega' V_j|^2$$

$$= O\left(a_n^{-2}E\left[\left(\sup_{|t|\le a_n^{-1}}|\Pi_{s,l}(t)|\right)^2\right]\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}\right)$$

$$= O\left(a_n^{-2}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-6}\right), \tag{B.23}$$

where the first step follows by random sampling (Assumption M (1)), the second step follows from the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)), and the last step follows by Lemma 5. Then, we have

$$E|\omega'\{p_{s,j,l}+p_{s,l,j}\}|^2 \le 8\{E|\omega'\xi_{s,a,j,l}|^2 + E|\omega'\xi_j|^2\} = O\left(a_n^{-2}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-6}\right), \tag{B.24}$$

where the first step follows by Jensen's inequality and the second step follows by (B.23) and (A.9). Under Assumption O (1), (B.24) implies $E|\omega'\{p_{j,l}+p_{l,j}\}|^2 = O(a_n^{-2(1+3\alpha)})$ and (B.22) follows by Assumption OS (1).

For $\omega' T_{n,s}^r$, note that

$$|\omega' T_{n,s}^r|^2 = O\left(n^{-3}\{E|\omega'\xi_{s,a,1,1}|^2 + E|\omega'\xi_1|^2\}\right) = O\left(n^{-3}a_n^{-2}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-6}\right), \tag{B.25}$$

where the last step follows by (B.23) and (A.9). Under Assumption O (1), (B.25) implies

$$|\omega' T_{n,s}^r| = O_p(n^{-3/2}a_n^{-(1+3\alpha)}). \tag{B.26}$$

For $\omega'\hat{B}_{n,s}$, note that

$$|\omega'\hat{B}_{n,s}| = \frac{1}{2\pi n a_n}\int E\left[\omega' V_j e^{-\mathrm{i}t\left(\frac{w^*-W_j}{a_n}\right)}\Pi_{s,j}(t/a_n)\right]\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}dt$$

$$= \frac{1}{4\pi n a_n}\int E\left[\omega' V_j e^{-\mathrm{i}t\left(\frac{w^*-W_j^*}{a_n}\right)}\right]E\left[e^{\mathrm{i}t\epsilon_j/a_n}\left\{\frac{1}{f_\epsilon^{\mathrm{ft}}(t/a_n)} - \frac{\cos\{t(\epsilon_j-\epsilon_j^r)/a_n\}}{\{f_\epsilon^{\mathrm{ft}}(t/a_n)\}^3}\right\}\right]K^{\mathrm{ft}}(t)dt$$

$$= \frac{1}{4\pi n a_n}\int E\left[\omega' V_j e^{-\mathrm{i}t\left(\frac{w^*-W^*}{a_n}\right)}\right]\left\{1 - \frac{f_\epsilon^{\mathrm{ft}}(2t/a_n)+1}{2\{f_\epsilon^{\mathrm{ft}}(t/a_n)\}^2}\right\}K^{\mathrm{ft}}(t)dt$$

$$= O\left(\frac{1}{n a_n}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}\right), \tag{B.27}$$

where the first step follows from definitions of $\hat{B}_{n,s}$ and $\xi_{s,a,j,j}$, the second step follows by independence between ϵ and (Y,X,W^*) (Assumption M (1)) and the definition of $\Pi_{s,j}$, the third step follows from $\cos(x) = \{e^x + e^{-x}\}/2$ and the fact that ϵ^r is independent of ϵ and has the same distribution as ϵ (Assumption RS), and the penultimate step follows by the fact that K^{ft}

is supported on $[-1,1]$ (Assumption M (3)). Under Assumption O (1), (B.27) implies

$$|\hat{B}_{n,s}| = O(n^{-1}a_n^{-(1+2\alpha)}). \tag{B.28}$$

Therefore, the conclusion follows by

$$\frac{n^{1/2}\omega'\{\hat{T}_{n,s} - T_{n,s}^* + T_{n,s}^r + B_n + \hat{B}_{n,s} + \hat{R}_{n,s}\}}{2\sqrt{Var(\omega'(\xi_j + \xi_{s,j}^*))}} = o_p(1), \tag{B.29}$$

for which we combine (A.13), (B.8), (B.20), (B.21), (B.26), and (B.28) with Assumption OS (2).

B.3. **Proof of Theorem 4 (i).** First, note that by $n^{-1/2}e^{2\mu a_n^{-\gamma}}\log(1/a_n) \to 0$ (Assumption SS (1)) and Lemma 4, under Assumption S (1), (B.5) implies

$$|\hat{S}_{n,s}| = O_p\left(n^{-1/2}e^{3\mu a_n^{-\gamma}}a_n^{-1}\log(1/a_n) + n^{-1}e^{5\mu a_n^{-\gamma}}a_n^{-1}\log(1/a_n)^2\right), \tag{B.30}$$

which together with $n^{-1/2}e^{3\mu a_n^{-\gamma}}a_n^{-1}\log(1/a_n) \to 0$ (Assumption SS (1)) gives (B.4), and thus (B.3) follows.

Also note that by $n^{-1/2}e^{2\mu a_n^{-\gamma}}\log(1/a_n) \to 0$ (Assumption SS (1)) and Lemma 4, under Assumption S (1), (B.7) implies

$$|\hat{R}_{n,s}|^2 = O_p\left(n^{-2}e^{10\mu a_n^{-\gamma}}a_n^{-2}\log(1/a_n)^4\right), \tag{B.31}$$

and (B.9) implies

$$|\hat{A}_{n,s} + \hat{B}_{n,s}|^2 = O_p\left(n^{-1}e^{6\mu a_n^{-\gamma}}a_n^{-2}\log(1/a_n)^2\right). \tag{B.32}$$

Combining (A.13), (A.24), (B.31), and (B.32) with (B.3), the conclusion follows.

B.4. **Proof of Theorem 4 (ii).** Under Assumption S (1), (B.14) implies

$$E|\omega'(\xi_j + \hat{\xi}_{s,j}) - E[\omega'\xi_j]|^{2+\eta} = O\left(e^{2(2+\eta)\mu a_n^{-\gamma}}a_n^{-(2+\eta)}\right), \tag{B.33}$$

and (B.19) together with (A.28) implies

$$\begin{aligned}
&\{Var(\omega'(\xi_j + \hat{\xi}_{s,j}))\}^{-(1+\eta/2)} \\
&= \begin{cases} O\left(\left(\int\left|\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}\cos(tw^*)K^{\mathrm{ft}}(ta_n)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2}\right|^2 dt + e^{2\mu a_n^{-\gamma}}a_n^{2(2+\theta)}\right)^{-(1+\eta/2)}\right) & \text{if } 1/3 < \gamma < 1 \\ O\left(\left(\int\left|\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}\cos(tw^*)K^{\mathrm{ft}}(ta_n)}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2}\right|^2 dt + e^{2\mu a_n^{-\gamma}}a_n^{2(\gamma\theta+\gamma-1)}\right)^{-(1+\eta/2)}\right) & \text{if } 1 \le \gamma \le 2 \end{cases}
\end{aligned} \tag{B.34}$$

Combining (B.33) and (B.34), (B.13) holds under Assumption SS (2), and thus (B.12) follows.

Under Assumption S (1), (B.24) implies $E|\omega'\{p_{s,j,l} + p_{s,l,j}\}|^2 = O\left(a_n^{-2}e^{6\mu a_n^{-\gamma}}\right)$ and (B.22) follows by $na_n^2e^{-6\mu a_n^{-\gamma}} \to \infty$ (Assumption SS (1)), (B.25) implies

$$|\omega' T_{n,s}^r| = O_p\left(n^{-3/2}e^{3\mu a_n^{-\gamma}}a_n^{-1}\right), \tag{B.35}$$

and (B.27) implies

$$|\hat{B}_{n,s}| = O\left(n^{-1}e^{2\mu a_n^{-\gamma}}a_n^{-1}\right). \tag{B.36}$$

Combining (A.13), (B.21), (B.31), (B.34), (B.35), and (B.36), (B.29) holds under Assumption SS (2). The conclusion then follows by (B.12) and (B.29) .

Appendix C. Lemmas

Lemma 1. *Under Assumption M, for any function g such that $gf_{W^*} \in L^2(\mathbb{R})$, it holds*

$$E\left[g(W^*)\mathbb{K}\left(\frac{w^*-W}{a_n}\right)\right] = E\left[g(W^*)K\left(\frac{w^*-W^*}{a_n}\right)\right].$$

Proof.

$$\begin{aligned} E\left[g(W^*)\mathbb{K}\left(\frac{w^*-W}{a_n}\right)\right] &= \frac{1}{2\pi}\int e^{-\mathrm{i}tw^*/a_n} E\left[g(W^*)e^{\mathrm{i}tW/a_n}\right]\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}dt \\ &= \frac{1}{2\pi}\int e^{-\mathrm{i}tw^*/a_n}\{gf_{W^*}\}^{\mathrm{ft}}(t/a_n)K^{\mathrm{ft}}(t)dt \\ &= E\left[g(W^*)K\left(\frac{w^*-W^*}{a_n}\right)\right], \end{aligned}$$

where the first step follows by the definition of $\mathbb{K}$, the second step follows by the independence between ϵ and W^* (Assumption M (1)), and the last step follows by the Plancherel's isometry. □

Lemma 2. *Under Assumption M, for $\eta > 0$, it holds*

$$\int |\mathbb{K}(x)|^{2+\eta}dx = O\left(\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-(2+\eta)}\right).$$

Proof. The conclusion follows by $\int|\mathbb{K}(x)|^{2+\eta}dx \le \sup_x|\mathbb{K}(x)|^\eta\int\mathbb{K}^2(x)dx$ with

$$\begin{aligned} \sup_x|\mathbb{K}(x)|^\eta &= O\left(\left(\int\left|\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}\right|dt\right)^\eta\right) = O\left(\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-\eta}\right), \\ \int\mathbb{K}^2(x)dx &= O\left(\int\left|\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}\right|^2dt\right) = O\left(\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}\right), \end{aligned}$$

where we use the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)) and Parseval's identity. □

Lemma 3. *Under Assumptions RS and MS (1), it holds*

$$\sup_{|t|\le a_n^{-1}}|\hat{\mu}_s(t)-\mu_s(t)| = O_p\left(n^{-1/2}\log(1/a_n)\right).$$

Proof. Note that $|\hat{\mu}_s(t)-\mu_s(t)| \le |\frac{1}{n}\sum_{j=1}^n e^{\mathrm{i}t(\epsilon_j-\epsilon_j^r)}-\mu_s(t)|$. The conclusion then follows by $E|\epsilon|^{2+\varsigma}<\infty$ for some $\varsigma>0$ (Assumption MS (1)) and Kurisu and Otsu (2020, Lemma 1). □

Lemma 4. *Under Assumptions RS and MS (1), it holds*

$$\sup_{|t|\le a_n^{-1}}|\hat{\Pi}_s(t)| = O_p\left(\frac{\log(1/a_n)}{n^{1/2}\{\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\}^2}\right).$$

Moreover, if $n^{-1/2}\log(1/a_n)\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}\to 0$, it holds

$$\sup_{|t|\le a_n^{-1}}|\hat{\Pi}_s^{\mathrm{res}}(t)| = O_p\left(\frac{\log(1/a_n)^2}{n\{\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\}^4}\right).$$

Proof. The first statement follows by Lemma 3 and $\sup_{|t|\le a_n^{-1}} |\hat{\Pi}_s(t)| = \frac{\sup_{|t|\le a_n^{-1}} |\hat{\mu}_s(t)-\mu_s(t)|}{2\{\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\}^2}$. For the second statement, note that

$$\begin{aligned}
&\sup_{|t|\le a_n^{-1}} |\hat{\Pi}_s^{\mathrm{res}}(t)| \le \sup_{|t|\le a_n^{-1}} \left| \frac{2(\mu_s(t)-\hat{\mu}_s(t))(|\mu_s(t)|^{1/2}-|\hat{\mu}_s(t)|^{1/2})}{\mu_s(t)|\hat{\mu}_s(t)|^{1/2}} \right| \\
= &\sup_{|t|\le a_n^{-1}} \left| \frac{2(\hat{\mu}_s(t)-\mu_s(t))^2}{\mu_s(t)|\hat{\mu}_s(t)|^{1/2}(|\mu_s(t)|^{1/2}+|\hat{\mu}_s(t)|^{1/2})} \right| \le \sup_{|t|\le a_n^{-1}} \left| \frac{2(\hat{\mu}_s(t)-\mu_s(t))^2}{\mu_s(t)\{(\hat{\mu}_s(t)-\mu_s(t))+\mu_s(t)\}} \right| \\
\le &\frac{2\left(\sup_{|t|\le a_n^{-1}} |\hat{\mu}_s(t)-\mu_s(t)|\right)^2}{\inf_{|t|\le a_n^{-1}} \mu_s(t) \inf_{|t|\le a_n^{-1}} \{(\hat{\mu}_s(t)-\mu_s(t))+\mu_s(t)\}}.
\end{aligned}$$

The conclusion then follows by Lemma 3 and $n^{-1/2}\log(1/a_n)\left(\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2} \to 0$. ☐

Lemma 5. *Under Assumption RS, for $\eta > 0$, it holds*

$$E\left[\left(\sup_{|t|\le a_n^{-1}} |\Pi_{s,1}(t)|\right)^{2+\eta}\right] = O\left(\left(\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)\right)^{-2(2+\eta)}\right).$$

Proof. The conclusion follows by

$$\sup_{|t|\le a_n^{-1}} 2|\Pi_{s,1}(t)| = \sup_{|t|\le a_n^{-1}} \left\{1 - \frac{\cos\{t(W_1-W_1^r)\}}{\{f_\epsilon^{\mathrm{ft}}(t)\}^2}\right\} \le 1 + \left(\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}.$$

☐

References

[1] Adusumilli, K. and T. Otsu (2018) Nonparametric instrumental regression with errors in variables, *Econometric Theory*, 34, 1256-1280.

[2] Ahn, H. and J.L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism, *Journal of Econometrics*, 58, 3-29.

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

[4] Bonin, H., Dohmen, T., Falk, A., Huffman, D. and U. Sunde (2007) Cross-sectional earnings risk and occupational sorting: The role of risk attitudes, *Labour Economics,* 14, 926-937.

[5] Breiman, L. (2001) Random Forests, *Machine Learning*, 45, 5-32.

[6] Butucea, C. and A.B. Tsybakov (2008) Sharp optimality in density deconvolution with dominating bias. I, *Theory of Probability & Its Applications*, 52, 24-39.

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

[8] Cawley, J., Heckman, J. and E. Vytlacil (2001) Three observations on wages and measured cognitive ability, *Labour Economics*, 8, 419-442.

[9] Chichignoud, M., Hoang, V. H., Ngoc, T. M. P. and V. Rivoirard (2017) Adaptive wavelet multivariate regression with errors in variables, *Electronic Journal of Statistics*, 11, 682-724.

[10] Chung, K. L. (1974) A Course in Probability Theory, Academic Press, London.

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

[12] Comte, F. and C. Lacour (2013) Anisotropic adaptive kernel deconvolution, *Annales de l'IHP Probabilités et statistiques,* 49, 569-609.

[13] Comte, F. and M.L. Taupin (2007) Adaptive estimation in a nonparametric regression model with errors-in-variables, *Statistica Sinica*, 1065-1090.

[14] Cunha, F., Heckman, J. J. and S. M. Schennach (2010) Estimating the technology of cognitive and noncognitive skill formation, *Econometrica*, 78, 883-931.

[15] Delaigle, A. and I. Gijbels (2004) Bootstrap bandwidth selection in kernel density estimation from a contaminated sample, *Annals of the Institute of Statistical Mathematics*, 56, 19-47.

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

[17] Delaigle, A. and P. Hall (2016) Methodology for non-parametric deconvolution when the error distribution is unknown, *Journal of the Royal Statistical Society: B*, 78, 231-252.

[18] Dohmen, T., Falk, A., Huffman, D. and U. Sunde (2010) Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100, 1238-60.

[19] Dohmen, T., Falk, A., Huffman, D. and U. Sunde (2018) On the relationship between cognitive ability and risk preference, *Journal of Economic Perspectives*, 32, 115-34.

[20] Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and G. G. Wagner (2011) Individual risk attitudes: Measurement, determinants, and behavioral consequences, *Journal of the European Economic Association*, 9, 522-550.

[21] Dong, H. and T. Otsu (2019) Nonparametric estimation of additive model with errors-in-variables, Working paper.

[22] Dong, H., Otsu, T. and L. Taylor (2020) Bandwidth selection for nonparametric regression with errors-in-variables, Working paper.

[23] Efron, B. and R. Tibshirani (1997) Improvements on cross-validation: the 632+ bootstrap method, *Journal of the American Statistical Association*, 92, 548-560.

[24] Fan, J. (1991) Asymptotic normality for deconvolution kernel density estimators, *Sankhyā: The Indian Journal of Statistics, Series A*, 97-110.

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

[26] Fan, J. and J.Y. Koo (2002) Wavelet deconvolution, *IEEE Transactions on Information Theory*, 48, 734-747.

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

[28] Feng, G., Gao, J., Peng, B. and X. Zhang (2017) A varying-coefficient panel data model with fixed effects: Theory and an application to US commercial banks, *Journal of Econometrics*, 196, 68-82.

[29] Hastie, T. and R. Tibshirani (1993) Varying-coefficient models, *Journal of the Royal Statistical Society: B*, 55, 757-779.

[30] Herrnstein, R. J. and C. Murray (1994) *The Bell Curve: Intelligence and Class Structure in American Life*, New York: Free Press.

[31] He, K., Lian, H., Ma, S. and J. Z. Huang (2018) Dimensionality reduction and variable selection in multivariate varying-coefficient models with a large number of covariates, *Journal of the American Statistical Association*, 113, 746-754.

[32] Heckman, J. J., Stixrud, J. and S. Urzua (2006) The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior, *Journal of Labor Economics*, 24, 411-482.

[33] Heckman, J. and E. Vytlacil (2001) Identifying the role of cognitive ability in explaining the level of and change in the return to schooling, *Review of Economics and Statistics*, 83, 1-12.

[34] Heshmati, A., Kumbhakar, S. C. and K. Sun (2014) Estimation of productivity in Korean electric power plants: A semiparametric smooth coefficient model, *Energy Economics*, 45, 491-500.

[35] Horowitz, J. L. and M. Markatou (1996) Semiparametric estimation of regression models for panel data, *Review of Economic Studies*, 63, 145-168.

[36] Jaeger, D. A., Dohmen, T., Falk, A., Huffman, D., Sunde, U. and H. Bonin (2010) Direct evidence on risk attitudes and migration, *Review of Economics and Statistics*, 92, 684-689.

[37] Kappus, J. and G. Mabon (2014) Adaptive density estimation in deconvolution problems with unknown error distribution, *Electronic Journal of Statistics*, 8, 2879-2904.

[38] Kotlarski, I. (1967) On characterizing the gamma and the normal distribution, *Pacific Journal of Mathematics*, 20, 69-76.

[39] Kurisu, D. and T. Otsu (2020) On uniform convergence of deconvolution estimator from repeated measurements, Working paper.

[40] Lepski, O.V. (2018) A new approach to estimator selection, *Bernoulli*, 24, 2776-2810.

[41] Lepski, O.V. and T. Willer (2019) Oracle inequalities and adaptive estimation in the convolution structure density model, *Annals of Statistics*, 47, 233-287.

[42] Levine, R. and Y. Rubinstein (2017) Smart and illicit: who becomes an entrepreneur and do they earn more? *Quarterly Journal of Economics,* 132, 963-1018.

[43] Li, L. and T. Greene (2008) Varying coefficients model with measurement error, *Biometrics*, 64, 519-526.

[44] Li, Q., Huang, C. J., Li, D. and T.-T. Fu (2002) Semiparametric smooth coefficient models, *Journal of Business & Economic Statistics*, 20, 412-422.

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

[46] Ma, S. and P. X. K. Song (2015) Varying index coefficient models, *Journal of the American Statistical Association*, 110, 341-356.

[47] Mamuneas, T. P., Savvides, A. and T. Stengos (2006) Economic development and the return to human capital: a smooth coefficient semiparametric approach, *Journal of Applied Econometrics*, 21, 111-132.

[48] McMurry, T. L. and D. N. Politis (2004) Nonparametric regression with infinite order flat-top kernels, *Journal of Nonparametric Statistics*, 16, 549-562.

[49] Otsu, T. and L. Taylor (2020) Specification testing for errors-in-variables models, *Econometric Theory*, Forthcoming.

[50] Pensky, M. and B. Vidakovic (1999) Adaptive wavelet estimator for nonparametric density deconvolution, *Annals of Statistics*, 27, 2033-2053.

[51] Robinson, P. M. (1988) Root-N-consistent semiparametric regression, *Econometrica*, 931-954.

[52] Schennach, S. M. (2004) Nonparametric regression in the presence of measurement error, *Econometric Theory*, 20, 1046-1093.

[53] Schennach, S. M. (2016) Recent advances in the measurement error literature, *Annual Review of Economics*, 8, 341-377.

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

[55] Skriabikova, O. J., Dohmen, T. and B. Kriechel (2014) New evidence on the relationship between risk attitudes and self-employment, *Labour Economics*, 30, 176-184.

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

[57] Uh, H. W. (2003) Kernel Deconvolution, PhD thesis, University of Amsterdam.

[58] Van Es, A. J. and H. W. Uh (2004) Asymptotic normality of nonparametric kernel type deconvolution density estimators: crossing the Cauchy boundary, *Nonparametric Statistics*, 16, 261-277.

[59] Yao, F., Zhang, F. and S. C. Kumbhakar (2019) Semiparametric smooth coefficient stochastic frontier model with panel data, *Journal of Business & Economic Statistics*, 37, 556-572.

[60] You, J. and G. Chen (2006) Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model, *Journal of Multivariate Analysis*, 97, 324-341.

[61] Zhou, Y. and H. Liang (2009) Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates, *Annals of Statistics*, 37, 427-458.

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

Email address: haod@smu.edu

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

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

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

Email address: lntaylor@econ.au.dk

ESTIMATION OF VARYING COEFFICIENT MODELS WITH MEASUREMENT ERROR - SUPPLEMENTARY MATERIAL

HAO DONG, TAISUKE OTSU, AND LUKE TAYLOR

In this supplementary appendix, we discuss our varying coefficient regression estimator when the measurement error distribution is unknown and possibly asymmetric.

1. Setup and estimator

If we wish to avoid assuming f_ϵ to be symmetric around zero, to identify the distribution of ϵ, we require the following assumption.

Assumption RK. *ϵ^r is independent of (Y, X, W^*, ϵ) and has the same distribution as ϵ, $f^{\text{ft}}_{W^*}$ and f^{ft}_ϵ vanish nowhere, and $E[\epsilon] = 0$.*

These assumptions are common in the literature (e.g., Li and Vuong, 1998, and Kurisu and Otsu, 2020). Under Assumption RK, Kotlarski's (1967) identity gives $f^{\text{ft}}_{W^*}(t) = \exp\left(\int_0^t \frac{\mathrm{i}E[W^r e^{\mathrm{i}sW}]}{E[e^{\mathrm{i}sW}]}ds\right)$ and the error distribution is identified by $f^{\text{ft}}_\epsilon(t) = \frac{f^{\text{ft}}_W(t)}{f^{\text{ft}}_{W^*}(t)}$. Based on an i.i.d. sample $\{W_j, W_j^r\}_{j=1}^n$ of (W, W^r), $f^{\text{ft}}_\epsilon(t)$ can be estimated by (Li and Vuong, 1998)

$$\hat{f}^{\text{ft}}_\epsilon(t) = \frac{\hat{f}^{\text{ft}}_W(t)}{\hat{f}^{\text{ft}}_{W^*}(t)},$$

where $\hat{f}^{\text{ft}}_W(t) = n^{-1}\sum_{j=1}^n e^{\mathrm{i}tW_j^r}$ and $\hat{f}^{\text{ft}}_{W^*}(t) = \exp\left(\int_0^t \frac{\mathrm{i}\sum_{j=1}^n W_j^r e^{\mathrm{i}sW_j}}{\sum_{j=1}^n e^{\mathrm{i}sW_j}}ds\right)$.

Based on this estimator, we propose to estimate $\beta(w^*)$ by

$$\tilde{\beta}(w^*) = \left[\sum_{j=1}^n X_j X_j' \hat{\mathbb{K}}\left(\frac{w^* - W_j}{a_n}\right)\right]^{-1} \sum_{j=1}^n X_j Y_j \hat{\mathbb{K}}\left(\frac{w^* - W_j}{a_n}\right), \tag{1.1}$$

where $\hat{\mathbb{K}}$ is the deconvolution kernel function obtained by replacing $f^{\text{ft}}_\epsilon(t)$ in (2.4) with $\hat{f}^{\text{ft}}_\epsilon(t)$. A similar plug-in estimator was used in Schennach (2004) for a regression function when the regressor is mismeasured and repeated noisy measurements are available.

1.1. Asymptotic properties.

To analyse the asymptotic properties of $\tilde{\beta}(w^*)$, we first impose the following assumptions for both the ordinary smooth and supersmooth error cases.

Assumption MK.

(1): *$E|W^*|^{2+\eta} < \infty$ and $E|\epsilon|^{2+\eta} < \infty$ for some $\eta > 0$.*

(2): *$f_{W^*}(\cdot)$ and $E[W^{r^2}|W^* = \cdot]$ are continuous and non-vanishing almost everywhere.*

(3): *$\frac{d}{dt}\left\{\{\omega' b(\cdot)\}^{\text{ft}}(t)\{f^{\text{ft}}_\epsilon(t)\}^{-1}\{f^{\text{ft}}_{W^*}(t)\}^{-1}\right\} \in \mathbb{W}$,*
$\frac{d}{dt}\left\{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\text{ft}}(s)e^{-\mathrm{i}sw^*}ds\{f^{\text{ft}}_\epsilon(t)\}^{-1}\{f^{\text{ft}}_{W^*}(t)\}^{-1}\right\} \in \mathbb{W}$, and
$\frac{d}{dt}\left\{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\text{ft}}(s)e^{-\mathrm{i}sw^*}ds\{f^{\text{ft}}_\epsilon(t)\}^{-1}\frac{d}{dt}\{f^{\text{ft}}_{W^*}(t)\}\{f^{\text{ft}}_{W^*}(t)\}^{-2}\right\} \in \mathbb{W}$ for any $\omega \in \mathbb{R}^k$ with $|\omega| = 1$.

(4): *For any* $\omega \in \mathbb{R}^k$ *with* $|\omega| = 1$, *there exists some constant* $\bar{c} > 0$ *such that*

$$Var(\omega'(\xi_1 + \xi_{1,1}^* + \xi_{2,1}^* + \xi_{3,1}^*)) \geq \bar{c} \max\{Var(\omega'\xi_1), Var(\omega'\xi_{1,1}^*), Var(\omega'\xi_{2,1}^*), Var(\omega'\xi_{3,1}^*)\},$$

where $\xi_{1,1}^*$, $\xi_{2,1}^*$, *and* $\xi_{3,1}^*$ *are defined in (D.16), (D.17), and (D.18), respectively, in Appendix.*

Assumption MK (1) is a regularity condition required by Lemma 1, which is used to characterize the uniform convergence rate of the empirical characteristic function of (W, W^r) and its first-order derivative over an expanding region. Assumptions MK (2)-(4) contain further conditions required to derive the distribution of our estimator when the error distribution is unknown and possibly asymmetric. We emphasise that the convergence rate result below does not require Assumptions MK (2)-(4). Assumptions MK (2)-(3) are regularity conditions used to derive the lower bound for the variance of $\tilde{\beta}(w^*)$; similar assumptions are used in Schennach (2004, Assumptions 12 and 13). In particular, Assumption MK (3) characterizes the tail behavior of the dominant component of $\tilde{\beta}(w^*)$'s asymptotic representation brought by the estimation error of $\hat{f}_\epsilon^{\text{ft}}$, and is used together with Assumption O (3) in the ordinary smooth case or Assumption S (3) in the supersmooth case to derive the lower bound for the variance of $\tilde{\beta}(w^*)$. Assumption MK (4) states that the variance of $\tilde{\beta}(w^*)$ is of an order no less than any term in its asymptotic representation; a similar assumption is used in Schennach (2004, Assumption 14).

We again start with the ordinary smooth measurement error case and introduce the following additional assumptions.

Assumption OK.

(1): *There exists positive constants* α_w *and* $c_{w,0}^{\text{os}}$ *such that*

$$|f_{W^*}^{\text{ft}}(t)|(1 + |t|)^{\alpha_w} \geq c_{w,0}^{\text{os}} \quad \textit{for all } t \in \mathbb{R}.$$

(2): $n^{-1/2} a_n^{-(2+3\alpha+2\alpha_w)} \log(1/a_n) \to 0$ *as* $n \to \infty$.

(3): *For any* $\omega \in \mathbb{R}^k$ *with* $|\omega| = 1$ *and for some* $\eta > 0$, *as* $n \to \infty$,

$$\left\{ \begin{array}{c} \min\{n^{-1} a_n^{-2p}, n a_n^{2(3+5\alpha+4\alpha_w)} \log(1/a_n)^{-4}, n^{\eta/(\eta+2)} a_n^{2(1+2\alpha)}\} \\ \times \max \left\{ \begin{array}{c} a_n^{-(1+2\alpha)}, \int \left| \frac{\{\omega' b(\cdot)\}^{\text{ft}}(t) K^{\text{ft}}(t a_n)}{f_\epsilon^{\text{ft}}(t) f_{W^*}^{\text{ft}}(t)} \right|^2 dt, \\ \int \left| \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\text{ft}}(s) K^{\text{ft}}(s a_n) e^{-isw^*} ds \frac{d}{dt}\{f_{W^*}^{\text{ft}}(t)\}}{f_\epsilon^{\text{ft}}(t)\{f_{W^*}^{\text{ft}}(t)\}^2} \right|^2 dt, \\ \int \left| \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\text{ft}}(s) K^{\text{ft}}(s a_n) e^{-isw^*} ds}{f_\epsilon^{\text{ft}}(t) f_{W^*}^{\text{ft}}(t)} \right|^2 dt \end{array} \right\} \end{array} \right\} \to \infty.$$

Assumption OK (1) assumes that f_{W^*} is ordinary smooth of order α_w, which is introduced to guarantee the consistency of the estimated error characteristic function $\hat{f}_\epsilon^{\text{ft}}$. Similar to Assumption O (1) for f_ϵ, in contrast to the traditional ordinary smooth assumption, we only need to impose a lower bound on $|f_{W^*}^{\text{ft}}(t)|$ to study the upper bound of the risk as in Theorem 1 (i). Assumption OK (2) is imposed to control the magnitude of the estimation error from using $\hat{f}_\epsilon^{\text{ft}}$ when the measurement error is ordinary smooth. Assumption OK (3) imposes additional bandwidth conditions to derive the asymptotic normality of $\tilde{\beta}(w^*)$. The component $a_n^{-(1+2\alpha)}$ characterizes the estimation variance if f_ϵ is known, and depends on the smoothness of f_ϵ as in the standard deconvolution literature. The components $\int \left| \frac{\{\omega' b(\cdot)\}^{\text{ft}}(t) K^{\text{ft}}(ta_n)}{f_\epsilon^{\text{ft}}(t) f_{W^*}^{\text{ft}}(t)} \right|^2 dt$, $\int \left| \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\text{ft}}(s) K^{\text{ft}}(sa_n) e^{-isw^*} ds \frac{d}{dt}\{f_{W^*}^{\text{ft}}(t)\}}{f_\epsilon^{\text{ft}}(t)\{f_{W^*}^{\text{ft}}(t)\}^2} \right|^2 dt$, and $\int \left| \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\text{ft}}(s) K^{\text{ft}}(sa_n) e^{-isw^*} ds}{f_\epsilon^{\text{ft}}(t) f_{W^*}^{\text{ft}}(t)} \right|^2 dt$ separately characterize the estimation errors brought by using $\frac{1}{n}\sum_{j=1}^n e^{\text{i}tW_j^r}$, $\frac{1}{n}\sum_{j=1}^n e^{\text{i}tW_j}$, and $\frac{1}{n}\sum_{j=1}^n W_j^r e^{\text{i}tW_j}$ in place of $E[e^{\text{i}tW^r}]$, $E[e^{\text{i}tW}]$, and $E[W^r e^{\text{i}tW}]$, respectively, and depend on the relative smoothness of f_ϵ and f_{W^*} to the conditional moment $E[XX'|W^*]$ and the coefficient function β. These four terms together characterizes (up to $1/n$) the magnitude of the lower bound of the variance of $\tilde{\beta}(w^*)$. For $\min\{n^{-1}a_n^{-2p}, na_n^{2(3+5\alpha+4\alpha_w)}\log(1/a_n)^{-4}, n^{\eta/(\eta+2)}a_n^{2(1+2\alpha)}\}$, the first two terms reflect the estimation bias when f_ϵ is known and the estimation error from using $\hat{f}_\epsilon^{\text{ft}}$, respectively, and the last term reflects Lyapunov's condition to apply the central limit theorem. So, Assumption OK (3) contains further restrictions on the bandwidth to use Lyapunov's central limit theorem and to ensure the asymptotic negligibility of the estimation bias and higher order terms in the estimation error from using $\hat{f}_\epsilon^{\text{ft}}$.

Theorem 1.

(i): *Under Assumptions M (1)-(4), O (1)-(2), RK, MK (1), and OK (1)-(2), it holds*

$$|\tilde{\beta}(w^*) - \beta(w^*)|^2 = O_p\left(n^{-1}a_n^{-2(2+3\alpha+2\alpha_w)}\log(1/a_n)^2 + a_n^{2p}\right).$$

(ii): *Under Assumptions M (1)-(5), O (1)-(5), RK, MK (1)-(4), and OK (1)-(3), it holds*

$$\hat{\Omega}_n(w^*)^{-1/2}\{\tilde{\beta}(w^*) - \beta(w^*)\} \xrightarrow{d} N(0, I_k),$$

where $\hat{\Omega}_n(w^*) = n^{-1}S(w^*)^{-1}Var(\xi_1 + \xi_{1,1}^* + \xi_{2,1}^* + \xi_{3,1}^*)S(w^*)^{-1}$.

Theorem 1 (i) shows the L_2-risk of our estimator $\tilde{\beta}(w^*)$ when the measurement error is ordinary smooth and possibly asymmetric. The first term relates to the estimation error of $\hat{f}_\epsilon^{\text{ft}}$. The second term is the usual bias term from an error-free nonparametric estimator. Theorem 1 (ii) shows that the estimator retains its asymptotic normality when the measurement error characteristic function is estimated using the approach of Li and Vuong (1998). Compared to the rate result on $\hat{\beta}(w^*)$ established in Theorem 1 of the main body, where the error distribution is known, $\tilde{\beta}(w^*)$ converges more slowly due to the estimation error from $\hat{f}_\epsilon^{\text{ft}}$. Compared to the rate result on $\tilde{\beta}_s(w^*)$ established in Theorem 3 of the main body, where the error distribution is unknown but is symmetric around zero, the symmetry of the error distribution allows $\tilde{\beta}_s(w^*)$ to converge faster than $\tilde{\beta}(w^*)$.

We now turn to the supersmooth measurement error case which requires the following additional assumptions.

Assumption SK.

(1): *For some positive constants* μ_w, γ_w, *and* $c_{w,0}^{\text{ss}}$, *it holds*

$$|f_{W^*}^{\text{ft}}(t)|e^{\mu_w|t|^{\gamma_w}} \geq c_{w,0}^{\text{ss}} \quad \textit{for all } t \in \mathbb{R}.$$

(2): $n^{-1/2}e^{3\mu a_n^{-\gamma} + 2\mu_w a_n^{-\gamma_w}}a_n^{-2}\log(1/a_n) \to 0$ *as* $n \to \infty$.

(3): *For any* $\omega \in \mathbb{R}^k$ *with* $|\omega| = 1$ *and for some* $\eta > 0$*, as* $n \to \infty$*,*

$$\left\{ \begin{array}{c} \min\{n^{-1}a_n^{-2p}, ne^{-10\mu a_n^{-\gamma}-8\mu_w a_n^{-\gamma_w}} a_n^6 \log(1/a_n)^{-4}, n^{\eta/(\eta+2)} e^{-2\mu a_n^{-\gamma}-2\mu_w a_n^{-\gamma_w}} a_n^2\} \\ \times \max \left\{ \begin{array}{c} e^{2\mu a_n^{-\gamma}} a_n^{2(2+\theta)}, \int \left| \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t) K^{\mathrm{ft}}(ta_n)}{f_\epsilon^{\mathrm{ft}}(t) f_{W^*}^{\mathrm{ft}}(t)} \right|^2 dt, \\ \int \left| \frac{\{\int_t^{\infty} + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\mathrm{ft}}(s) K^{\mathrm{ft}}(sa_n) e^{-isw^*} ds \frac{d}{dt}\{f_{W^*}^{\mathrm{ft}}(t)\}}{f_\epsilon^{\mathrm{ft}}(t)\{f_{W^*}^{\mathrm{ft}}(t)\}^2} \right|^2 dt, \\ \int \left| \frac{\{\int_t^{\infty} + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\mathrm{ft}}(s) K^{\mathrm{ft}}(sa_n) e^{-isw^*} ds}{f_\epsilon^{\mathrm{ft}}(t) f_{W^*}^{\mathrm{ft}}(t)} \right|^2 dt \end{array} \right\} \end{array} \right\} \to \infty,$$

if $1/3 < \gamma < 1$ *and*

$$\left\{ \begin{array}{c} \min\{n^{-1}a_n^{-2p}, ne^{-10\mu a_n^{-\gamma}-8\mu_w a_n^{-\gamma_w}} a_n^6 \log(1/a_n)^{-4}, n^{\eta/(\eta+2)} e^{-2\mu a_n^{-\gamma}-2\mu_w a_n^{-\gamma_w}} a_n^2\} \\ \times \max \left\{ \begin{array}{c} e^{2\mu a_n^{-\gamma}} a_n^{2(\gamma\theta+\gamma-1)}, \int \left| \frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t) K^{\mathrm{ft}}(ta_n)}{f_{W^*}^{\mathrm{ft}}(t) f_\epsilon^{\mathrm{ft}}(t)} \right|^2 dt, \\ \int \left| \frac{\{\int_t^{\infty} + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\mathrm{ft}}(s) K^{\mathrm{ft}}(sa_n) e^{-isw^*} ds \frac{d}{dt}\{f_{W^*}^{\mathrm{ft}}(t)\}}{\{f_{W^*}^{\mathrm{ft}}(t)\}^2 f_\epsilon^{\mathrm{ft}}(t)} \right|^2 dt, \\ \int \left| \frac{\{\int_t^{\infty} + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\mathrm{ft}}(s) K^{\mathrm{ft}}(sa_n) e^{-isw^*} ds}{f_{W^*}^{\mathrm{ft}}(t) f_\epsilon^{\mathrm{ft}}(t)} \right|^2 dt \end{array} \right\} \end{array} \right\} \to \infty,$$

if $1 \le \gamma \le 2$.

Assumption SK (1) assumes that f_{W^*} is supersmooth of order γ_w. Similar to Assumption S (1) for f_ϵ, different from traditional supersmooth assumption, to study the upper bound of the risk as in Theorem 2 (i), we only impose a lower bound on $|f_{W^*}^{\mathrm{ft}}(t)|$. As in the ordinary smooth case, Assumption SK (2) is imposed to control the magnitude of the estimation error from using $\hat{f}_\epsilon^{\mathrm{ft}}$. Assumption SK (3) contains similar additional bandwidth conditions to derive the asymptotic normality of $\tilde{\beta}(w^*)$ as were used in the ordinary smooth setting, except that conditions are separately imposed for the cases when $1/3 < \gamma < 1$ and $1 \le \gamma \le 2$, which is compatible with Assumption S (5) as in the case when f_ϵ is known.

Theorem 2.

(i): *Under Assumptions M (1)-(4), S (1)-(3), RK, MK (1), and SK (1)-(2), it holds*

$$|\tilde{\beta}(w^*) - \beta(w^*)|^2 = O_p\left(n^{-1}e^{6\mu a_n^{-\gamma}+4\mu_w a_n^{-\gamma_w}} a_n^{-4}\log(1/a_n)^2 + a_n^{2p}\right).$$

(ii): *Under Assumptions M (1)-(5), S (1)-(3), RK, MK (1)-(4), and SK (1)-(3), it holds*

$$\hat{\Omega}_n(w^*)^{-1/2}\{\tilde{\beta}(w^*) - \beta(w^*)\} \xrightarrow{d} N(0, I_k),$$

where $\hat{\Omega}_n(w^*) = n^{-1}S(w^*)^{-1}Var(\xi_1 + \xi_{1,1}^* + \xi_{2,1}^* + \xi_{3,1}^*)S(w^*)^{-1}$.

Similar comments to Theorem 1 apply here. Theorem 2 (i) shows the L_2-risk of our estimator $\tilde{\beta}(w^*)$ when the measurement error is supersmooth and possibly asymmetric. Compared to the ordinary smooth case where $\tilde{\beta}(w^*)$ obtains a polynomial rate, by similar arguments to the case when the error distribution is known, using the rate result presented in Theorem 2 (i), we can show that $\tilde{\beta}(w^*)$ converges at a considerably slower logarithmic rate in the supersmooth case. Theorem 2 (ii) shows that the estimator retains its asymptotic normality when the measurement error characteristic function is estimated using the approach of Li and Vuong (1998). Again,

$\tilde{\beta}(w^*)$ converges more slowly than $\hat{\beta}(w^*)$ and $\tilde{\beta}_s(w^*)$ due to the estimation error from $\hat{f}_\epsilon^{\text{ft}}$ and the error distribution's lack of symmetry.

Appendix D. Proofs

Notation. We introduce $\hat{\mu}_\iota(t) = \frac{1}{n}\sum_{j=1}^n \mu_{\iota,j}(t)$ and $\mu_\iota(t) = E[\mu_{\iota,1}(t)]$ for $\iota = 1,2,3$ with $\mu_{1,j}(t) = e^{\mathrm{i}tW_j^r}$, $\mu_{2,j}(t) = e^{\mathrm{i}tW_j}$, and $\mu_{3,j}(t) = W_j^r e^{\mathrm{i}tW_j}$, which implies $\hat{f}_\epsilon^{\mathrm{ft}}(t) = \hat{\mu}_1(t)\exp\left(-\int_0^t \frac{\mathrm{i}\hat{\mu}_3(s)}{\hat{\mu}_2(s)}ds\right)$ and $f_\epsilon(t) = \mu_1(t)\exp\left(-\int_0^t \frac{\mathrm{i}\mu_3(s)}{\mu_2(s)}ds\right)$, i.e., $\hat{\mathbb{K}}$ and $\mathbb{K}$ are functionals of $(\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3)$ and (μ_1,μ_2,μ_3) respectively.

Define

$$\hat{\Pi}(t) = \frac{1}{n}\sum_{l=1}^n \Pi_l(t), \quad \Pi_l(t) = -\frac{\delta_{1,l}(t)}{\mu_1(t)} + \mathrm{i}\int_0^t\left\{-\frac{\mu_3(s)\delta_{2,l}(s)}{\mu_2^2(s)} + \frac{\delta_{3,l}(s)}{\mu_2(s)}\right\}ds,$$

$$\begin{aligned}\hat{\Pi}^{\mathrm{res}}(t) &= \frac{\hat{\delta}_1^2(t)}{\mu_1(t)+\hat{\delta}_1(t)} - \int_0^t \mathrm{i}\left\{-\frac{\mu_3(s)\hat{\delta}_2(s)}{\mu_2^2(s)} + \frac{\hat{\delta}_3(s)}{\mu_2(s)}\right\}\frac{\hat{\delta}_2(s)}{\mu_2(s)+\hat{\delta}_2(s)}ds \\ &+ \int_0^t \mathrm{i}\left\{-\frac{\mu_3(s)\hat{\delta}_2(s)}{\mu_2^2(s)} + \frac{\hat{\delta}_3(s)}{\mu_2(s)}\right\}\left\{1-\frac{\hat{\delta}_2(s)}{\mu_2(s)+\hat{\delta}_2(s)}\right\}ds\left\{-\frac{\hat{\delta}_1(t)}{\mu_1(t)} + \frac{\hat{\delta}_1^2(t)}{\mu_1(t)+\hat{\delta}_1(t)}\right\} \\ &-\frac{1}{2}e^{\bar{\phi}(t)}\left(\int_0^t\left\{-\frac{\mu_3(s)\hat{\delta}_2(s)}{\mu_2^2(s)} + \frac{\hat{\delta}_3(s)}{\mu_2(s)}\right\}\left\{1-\frac{\hat{\delta}_2(s)}{\mu_2(s)+\hat{\delta}_2(s)}\right\}ds\right)^2\left\{1-\frac{\hat{\delta}_1(t)}{\mu_1(t)} + \frac{\hat{\delta}_1^2(t)}{\mu_1(t)+\hat{\delta}_1(t)}\right\},\end{aligned}$$

for some $|\bar{\phi}(t)| \leq \left|\int_0^t\left\{-\frac{\mu_3(s)\hat{\delta}_2(s)}{\mu_2^2(s)} + \frac{\hat{\delta}_3(s)}{\mu_2(s)}\right\}\left\{1-\frac{\hat{\delta}_2(s)}{\mu_2(s)+\hat{\delta}_2(s)}\right\}ds\right|$, where $\hat{\delta}_\iota(t) = \frac{1}{n}\sum_{j=1}^n \delta_{\iota,j}(t)$ with $\delta_{\iota,j}(t) = \mu_{\iota,j}(t) - \mu_\iota(t)$ for $\iota = 1,2,3$.

By expanding $(\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3)$ around (μ_1,μ_2,μ_3), we obtain

$$\hat{\mathbb{K}}(u) = \mathbb{K}(u) + \mathbb{A}(u) + \mathbb{R}(u), \tag{D.1}$$

where

$$\begin{aligned}\mathbb{A}(u) &= \frac{1}{2\pi}\int e^{-\mathrm{i}tu}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}\hat{\Pi}(t/a_n)dt, \\ \mathbb{R}(u) &= \frac{1}{2\pi}\int e^{-\mathrm{i}tu}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}\hat{\Pi}^{\mathrm{res}}(t/a_n)dt.\end{aligned}$$

Here $\mathbb{A}(u)$ is the Fréchet derivative of $\hat{\mathbb{K}}(u)$ as a functional of $(\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3)$ at (μ_1,μ_2,μ_3) in the direction of $(\hat{\delta}_1,\hat{\delta}_2,\hat{\delta}_3)$, and $\mathbb{R}(u)$ represents the dominant components of the remainder. Also note that $\hat{\Pi}^{\mathrm{res}}(t)$ is of a higher order than $\hat{\Pi}(t)$. So, the remainder $\mathbb{R}(u)$ is should be dominated by $\mathbb{A}(u)$ asymptotically.

Define

$$\begin{aligned}\hat{S}_n &= \frac{1}{na_n}\sum_{j=1}^n X_jX_j'\left\{\mathbb{A}\left(\frac{w^*-W_j}{a_n}\right) + \mathbb{R}\left(\frac{w^*-W_j}{a_n}\right)\right\}, \\ \hat{A}_n &= \frac{1}{n}\sum_{j=1}^n\{\xi_{a,j} - E[\xi_{a,j}]\}, \quad \hat{B}_n = E[\xi_{a,j}], \quad \hat{R}_n = \frac{1}{n}\sum_{j=1}^n \xi_{r,j},\end{aligned}$$

where $\xi_{a,j} = a_n^{-1}V_j\mathbb{A}\left(\frac{w^*-W_j}{a_n}\right)$, $\xi_{r,j} = a_n^{-1}V_j\mathbb{R}\left(\frac{w^*-W_j}{a_n}\right)$, and V_j is defined in (A.2). Then, using the approximation result for $\hat{\mathbb{K}}$ as in (D.1), we have

$$\tilde{\beta}(w^*) = \beta(w^*) + (S_n+\hat{S}_n)^{-1}(T_n + \hat{A}_n + B_n + \hat{B}_n + \hat{R}_n), \tag{D.2}$$

where S_n, T_n, and B_n are defined in (A.1). Also, let $\hat{S}_n^{k_1,k_2}$ denote the (k_1,k_2)-th element of $\hat{S}_n$, $\xi_{a,j}^{k_1}$ and $\xi_{r,j}^{k_1}$ separately denote the k_1-element of $\xi_{a,j}$ and $\xi_{r,j}$ for $k_1,k_2=1,\ldots,k$. The proofs of Theorem 1 and 2 follow along similar lines as Theorem 1 and 2 of the main body respectively, as such, we only explain the parts that differ.

D.1. **Proof of Theorem 1 (i).** By (A.5), (D.2), and $\lambda_{\min}(S)>0$ (Assumption M (2)), we have

$$|\tilde{\beta}(w^*)-\beta(w^*)|^2 = O_p(|T_n|^2+|B_n|^2+|\hat{A}_n+\hat{B}_n|^2+|\hat{R}_n|^2), \tag{D.3}$$

if we can show

$$\hat{S}_n \xrightarrow{p} 0. \tag{D.4}$$

To show (D.4), note that for any $k_1,k_2=1,\ldots,k$,

$$\begin{aligned}
|\hat{S}_n^{k_1,k_2}| &\leq \frac{1}{na_n}\sum_{j=1}^n \left| X_{k_1,j}X_{k_2,j}\left\{\mathbb{A}\left(\frac{w^*-W_j}{a_n}\right)+\mathbb{R}\left(\frac{w^*-W_j}{a_n}\right)\right\}\right| \\
&\leq \frac{1}{na_n}\sum_{j=1}^n |X_{k_1,j}X_{k_2,j}|\int |K^{\mathrm{ft}}(t)|\frac{|\hat{\Pi}(t/a_n)|+|\hat{\Pi}^{\mathrm{res}}(t/a_n)|}{|f_\epsilon^{\mathrm{ft}}(t/a_n)|}dt \\
&= O_p\left(\frac{\sup_{|t|\leq a_n^{-1}}|\hat{\Pi}(t)|+\sup_{|t|\leq a_n^{-1}}|\hat{\Pi}^{\mathrm{res}}(t)|}{a_n\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|}\right),
\end{aligned} \tag{D.5}$$

where the second step follows by the definitions of $\mathbb{A}$ and $\mathbb{R}$, and the last step uses the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)). By Lemma 2 as in Appendix D and $n^{-1/2}a_n^{-(1+2\alpha+2\alpha_w)}\log(1/a_n)\to 0$ (Assumption OK (2)), under Assumptions O (1) and OK (1), (D.5) implies

$$|\hat{S}_n| = O_p\left(n^{-1/2}a_n^{-(2+3\alpha+2\alpha_w)}\log(1/a_n)\right), \tag{D.6}$$

and (D.4) follows by $n^{-1/2}a_n^{-(2+3\alpha+2\alpha_w)}\log(1/a_n)\to 0$ as $n\to\infty$ (Assumption OK (2)).

Since stochastic orders of $|T_n|^2$ and $|B_n|^2$ are obtained in Appendix A.1, it remains to characterize the orders of $|\hat{R}_n|^2$ and $|\hat{A}_n+\hat{B}_n|^2$. For $|\hat{R}_n|^2$, note that

$$\begin{aligned}
|\hat{R}_n| &\leq \sqrt{k}\max_{1\leq k_1\leq k}\left\{\frac{1}{n}\sum_{j=1}^n|\xi_{r,j}^{k_1}|\right\} \leq \sqrt{k}\max_{1\leq k_1\leq k}\left\{\frac{1}{na_n}\sum_{j=1}^n|V_j^{k_1}|\int |K^{\mathrm{ft}}(t)|\frac{|\hat{\Pi}^{\mathrm{res}}(t/a_n)|}{|f_\epsilon^{\mathrm{ft}}(t/a_n)|}dt\right\} \\
&= O_p\left(\frac{\sup_{|t|\leq a_n^{-1}}|\hat{\Pi}^{\mathrm{res}}(t)|}{a_n\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|}\right),
\end{aligned} \tag{D.7}$$

where the second step follows by the definitions of $\xi_{r,j}$ and $\mathbb{R}$, and the last step uses the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)). By Lemma 2 and $n^{-1/2}a_n^{-(1+2\alpha+2\alpha_w)}\log(1/a_n)\to 0$ as $n\to\infty$ (Assumption OK (2)), under Assumptions O (1) and OK (1), (D.7) implies

$$|\hat{R}_n|^2 = O_p\left(n^{-2}a_n^{-(6+10\alpha+8\alpha_w)}\log(1/a_n)^4\right). \tag{D.8}$$

For $|\hat{A}_n + \hat{B}_n|^2$, note that

$$\begin{aligned} |\hat{A}_n + \hat{B}_n| &\leq \sqrt{k} \max_{1\leq k_1 \leq k} \left\{ \frac{1}{n} \sum_{j=1}^{n} |\xi_{a,j}^{k_1}| \right\} \\ &\leq \sqrt{k} \max_{1\leq k_1 \leq k} \left\{ \frac{1}{na_n} \sum_{j=1}^{n} |V_j^{k_1}| \int |K^{\text{ft}}(t)| \frac{|\hat{\Pi}(t/a_n)|}{|f_\epsilon^{\text{ft}}(t/a_n)|} dt \right\} \\ &= O_p\left(\frac{\sup_{|t|\leq a_n^{-1}} |\hat{\Pi}(t)|}{a_n \inf_{|t|\leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|} \right), \end{aligned} \quad \text{(D.9)}$$

where the second step follows by definitions of $\xi_{a,j}$ and $\mathbb{A}$, and the last step uses the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)). By Lemma 2, under Assumptions O (1) and OK (1), (D.9) implies

$$|\hat{A}_n + \hat{B}_n|^2 = O_p\left(n^{-1} a_n^{-(4+6\alpha+4\alpha_w)} \log(1/a_n)^2 \right). \quad \text{(D.10)}$$

Combining (A.11), (A.13), (D.8), and (D.10) with (B.3), the conclusion follows.

D.2. **Proof of Theorem 1 (ii).** Define

$$T_n^* = \frac{2}{n} \sum_{j=1}^{n} \{\xi_j - E[\xi_j] + \hat{\xi}_j\}, \quad \hat{T}_n = \binom{n}{2}^{-1} \sum_{j=1}^{n-1} \sum_{l=j+1}^{n} \{p_{j,l} + p_{l,j}\}, \quad T_n^r = n^{-2} \sum_{j=1}^{n} p_{j,j},$$

where $\hat{\xi}_j = E_j[\xi_{a,l,j}]$ for $l \neq j$ and $p_{j,l} = \xi_j + \xi_{a,j,l} - E[\xi_j + \xi_{a,j,l}]$ with

$$\xi_{a,j,l} = \frac{1}{2\pi a_n} \int V_j e^{-\text{i}t\left(\frac{w^* - W_j}{a_n}\right)} \frac{K^{\text{ft}}(t)}{f_\epsilon^{\text{ft}}(t/a_n)} \Pi_l(t/a_n) dt.$$

Decompose

$$T_n + \hat{A}_n = \frac{n-1}{2n} T_n^* + \frac{n-1}{2n} \{\hat{T}_n - T_n^*\} + T_n^r. \quad \text{(D.11)}$$

The conclusion follows if we can establish the asymptotic normality of T_n^* and show the asymptotic negligibility of $\hat{T}_n - T_n^*$, T_n^r, B_n, $\hat{B}_n$, and $\hat{R}_n$. Using the Cramér–Wold device, this is equivalent to establishing the asymptotic normality of $\omega' T_n^*$ and showing the asymptotic negligibility of $\omega'(\hat{T}_n - T_n^*)$, $\omega' T_n^r$, $\omega' B_n$, $\omega' \hat{B}_n$, and $\omega' \hat{R}_n$ for each $\omega \in \mathbb{R}^k$ with $|\omega| = 1$.

For the asymptotic normality of $\omega' T_n^*$, we show

$$\frac{\omega' T_n^*}{\sqrt{Var(\omega' T_n^*)}} \xrightarrow{d} N(0,1). \quad \text{(D.12)}$$

To this end, it suffices to check Lyapunov's condition for some $\eta > 0$, i.e.,

$$\frac{E|\omega'(\xi_j + \hat{\xi}_j) - E[\omega' \xi_j]|^{2+\eta}}{n^{\eta/2} \{Var(\omega'(\xi_j + \hat{\xi}_j))\}^{1+\eta/2}} \to 0. \quad \text{(D.13)}$$

For the numerator of (D.13), note that

$$
\begin{aligned}
& E|\omega'(\xi_j+\hat{\xi}_j)-E[\omega'\xi_j]|^{2+\eta} \le 2^{1+\eta}\{E|\omega'\hat{\xi}_j|^{2+\eta}+E|\omega'\xi_j-E[\omega'\xi_j]|^{2+\eta}\} \\
= & O\left(E\left|\frac{1}{2\pi a_n}\int E\left[\omega' V e^{-\mathrm{i}t\left(\frac{w^*-W^*}{a_n}\right)}\right]\Pi_j(t/a_n)K^{\mathrm{ft}}(t)dt\right|^{2+\eta}+E|\omega'\xi_j-E[\omega'\xi_j]|^{2+\eta}\right) \\
= & O\left(a_n^{-(2+\eta)}(E|\omega' V|)^{2+\eta}E\left[\left(\sup_{|t|\le a_n^{-1}}|\Pi_j(t)|\right)^{2+\eta}\right]+E|\omega'\xi_j-E[\omega'\xi_j]|^{2+\eta}\right) \\
= & O\left(a_n^{-2(2+\eta)}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\inf_{|t|\le a_n^{-1}}|f_{W^*}^{\mathrm{ft}}(t)|\right)^{-2(2+\eta)}\right), \qquad (D.14)
\end{aligned}
$$

where the first step follows by Jensen's inequality, the third step follows from the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)), and the last step follows by Lemma 3 as in Appendix D and (A.18). Under Assumptions O (1) and OK (1), (D.14) implies

$$
E|\omega'(\xi_j+\hat{\xi}_j)-E[\omega'\xi_j]|^{2+\eta}=O(a_n^{-2(2+\eta)(1+\alpha+\alpha_w)}). \qquad (D.15)
$$

For the denominator of (D.13), note that $Var(\omega'(\xi_j+\hat{\xi}_j))=Var(\omega'(\xi_j+\sum_{\iota=1}^3\xi_{\iota,j}^*))$, where

$$
\begin{aligned}
\xi_{1,j}^* &= -\frac{1}{2\pi a_n}\int E\left[\omega' V e^{-\mathrm{i}t\left(\frac{w^*-W^*}{a_n}\right)}\right]\frac{K^{\mathrm{ft}}(t)}{\mu_1(t/a_n)}\mu_{1,j}(t/a_n)dt, \qquad (D.16)\\
\xi_{2,j}^* &= -\frac{\mathrm{i}}{2\pi a_n}\int E\left[\omega' V e^{-\mathrm{i}t\left(\frac{w^*-W^*}{a_n}\right)}\right]K^{\mathrm{ft}}(t)\left\{\int_0^{t/a_n}\frac{\mu_3(s)\mu_{2,j}(s)}{\mu_2^2(s)}ds\right\}dt, \qquad (D.17)\\
\xi_{3,j}^* &= \frac{\mathrm{i}}{2\pi a_n}\int E\left[\omega' V e^{-\mathrm{i}t\left(\frac{w^*-W^*}{a_n}\right)}\right]K^{\mathrm{ft}}(t)\left\{\int_0^{t/a_n}\frac{\mu_{3,j}(s)}{\mu_2(s)}ds\right\}dt. \qquad (D.18)
\end{aligned}
$$

By Assumption MK (4), for the lower bound of $Var(\omega'(\xi_j+\sum_{\iota=1}^3\xi_{\iota,j}^*))$, it suffices to focus on $Var(\omega'\xi_j)$ and $Var(\omega'\xi_{\iota,j}^*)$ for $\iota=1,2,3$, which are dominated by $E|\omega'\xi_j|^2$ and $E|\omega'\xi_{\iota,j}^*|^2$ for $\iota=1,2,3$, respectively.

For $E|\omega'\xi_{1,j}^*|^2$, note that for some finite interval $I_1\subset\mathbb{R}$ containing w^*, we have

$$
\begin{aligned}
E|\omega'\xi_{1,j}^*|^2 &= \frac{1}{4\pi^2}\int\left|\int\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)K^{\mathrm{ft}}(ta_n)}{f_\epsilon^{\mathrm{ft}}(t)f_{W^*}^{\mathrm{ft}}(t)}e^{\mathrm{i}t(u-w^*)}dt\right|^2 f_W(u)du \\
&\ge c_1\int_{u\in I_1}\left|\int\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)K^{\mathrm{ft}}(ta_n)}{f_\epsilon^{\mathrm{ft}}(t)f_{W^*}^{\mathrm{ft}}(t)}e^{\mathrm{i}t(u-w^*)}dt\right|^2 du, \qquad (D.19)
\end{aligned}
$$

where the first step follows from the definitions of $\xi_{1,j}^*$ and $b(\cdot)$ and the last step follows by choosing some constant $c_1>0$ such that $\inf_{u\in I_1}f_W(u)>4\pi^2c_1$, where such a c_1 exists due to the compactness of I_1 and the fact that f_W is continuous and non-vanishing everywhere (Assumption MK (2)).

As $\frac{d}{dt}\left\{\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{f_{W^*}^{\mathrm{ft}}(t)f_\epsilon^{\mathrm{ft}}(t)}\right\}\in\mathbb{W}$ (Assumption MK (3)), Schennach (2004, Lemma 10) implies

$$
\lim_{|u|\to\infty}(u-w^*)\int\frac{\{\omega' b(\cdot)\}^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t)f_{W^*}^{\mathrm{ft}}(t)}e^{\mathrm{i}t(u-w^*)}dt=0,
$$

and it follows

$$\int_{u\in I^c}\left|\int \frac{\{\omega' b(\cdot)\}^{\rm ft}(t)}{f_\epsilon^{\rm ft}(t) f_{W^*}^{\rm ft}(t)} e^{{\rm i}t(u-w^*)}dt\right|^2 du = O\left(\int_{u\in I^c}\frac{1}{(u-w^*)^2}du\right) = O(1). \tag{D.20}$$

Thus, for all n large enough and some constant $C_1>0$, we have

$$\begin{aligned} E|\omega'\xi^*_{1,j}|^2 &\geq C_1\int\left|\int \frac{\{\omega' b(\cdot)\}^{\rm ft}(t)K^{\rm ft}(ta_n)}{f_\epsilon^{\rm ft}(t) f_{W^*}^{\rm ft}(t)} e^{{\rm i}t(u-w^*)}dt\right|^2 du \\ &= 2\pi C_1\int\left|\frac{\{\omega' b(\cdot)\}^{\rm ft}(t)K^{\rm ft}(ta_n)}{f_\epsilon^{\rm ft}(t) f_{W^*}^{\rm ft}(t)}\right|^2 dt, \end{aligned} \tag{D.21}$$

where the first step follows by (D.19) and (D.20) and the last step follows by Parseval's identity.

For $E|\omega'\xi^*_{2,j}|^2$, note that for some finite interval $I_2\subset\mathbb{R}$, we have

$$\begin{aligned} E|\omega'\xi^*_{2,j}|^2 &= \frac{1}{4\pi^2}\int\left|\int\int_0^t \frac{\{\omega' b(\cdot)\}^{\rm ft}(t)\frac{d}{ds}\{f_{W^*}^{\rm ft}(s)\}e^{-{\rm i}tw^*}}{f_\epsilon^{\rm ft}(s)\{f_{W^*}^{\rm ft}(s)\}^2} e^{{\rm i}su}K^{\rm ft}(ta_n)dsdt\right|^2 f_W(u)du \\ &= \frac{1}{4\pi^2}\int\left|\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2} e^{{\rm i}tu}dt\right|^2 f_W(u)du \\ &\geq c_2\int_{u\in I_2}\left|\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2} e^{{\rm i}tu}dt\right|^2 du, \end{aligned} \tag{D.22}$$

where the first step follows from the definitions of $\xi^*_{2,j}$ and $b(\cdot)$, the second step uses $\int_{-\infty}^\infty\int_0^t f(t,s)dsdt = \int_0^\infty\int_t^\infty f(s,t)dsdt + \int_{-\infty}^0\int_t^{-\infty} f(s,t)dsdt$ for any absolutely integrable function f, and the last step follows by choosing some constant $c_2>0$ such that $\inf_{u\in I_2} f_W(u) > 4\pi^2 c_2$, where such a c_2 exists due to the compactness of I_2 and the fact that f_W is continuous and non-vanishing everywhere (Assumption MK (2)).

As $\frac{d}{dt}\left\{\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2}\right\}\in\mathbb{W}$ (Assumption MK (3)), Schennach (2004, Lemma 10) implies

$$\lim_{|u|\to\infty} u\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2} e^{{\rm i}tu}dt = 0,$$

and it follows

$$\int_{u\in I_2^c}\left|\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2} e^{{\rm i}tu}dt\right|^2 du = O\left(\int_{u\in I_2^c} u^{-2}du\right) = O(1). \tag{D.23}$$

Thus, for all n large enough and some constant $C_2>0$, we have

$$\begin{aligned} E|\omega'\xi^*_{2,j}|^2 &\geq C_2\int\left|\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2} e^{{\rm i}tu}dt\right|^2 du \\ &= 2\pi C_2\int\left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2}\right|^2 dt, \end{aligned} \tag{D.24}$$

where the first step follows by (D.22) and (D.23) and the last step follows by Parseval's identity.

For $E|\omega'\xi^*_{3,j}|^2$, note that for some finite interval $I_3 \subset \mathbb{R}$, we have

$$\begin{aligned} E|\omega'\xi^*_{3,j}|^2 &= \frac{1}{4\pi^2}\int\left|\int\int_0^t \frac{\{\omega' b(\cdot)\}^{\rm ft}(t)e^{-{\rm i}tw^*}}{f^{\rm ft}_\epsilon(s)f^{\rm ft}_{W^*}(s)} e^{{\rm i}su}K^{\rm ft}(ta_n)dsdt\right|^2 \{E[W^{r^2}|W=\cdot]f_W(\cdot)\}(u)du \\ &= \frac{1}{4\pi^2}\int\left|\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)} e^{{\rm i}tu}dt\right|^2 \{E[W^{r^2}|W=\cdot]f_W(\cdot)\}(u)du \\ &\geq c_3\int_{u\in I_3}\left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)} e^{{\rm i}tu}dt\right|^2 du, \end{aligned} \quad \text{(D.25)}$$

where the first step follows from the definitions of $\xi^*_{3,j}$ and $b(\cdot)$, the second step follows uses $\int_{-\infty}^\infty\int_0^t f(t,s)dsdt = \int_0^\infty\int_t^\infty f(s,t)dsdt + \int_{-\infty}^0\int_t^{-\infty} f(s,t)dsdt$ for any absolutely integrable function f, and the last step follows by choosing some constant $c_3 > 0$ such that $\inf_{u\in I_3}\{E[W^{r^2}|W = \cdot]f_W(\cdot)\}(u) > 4\pi^2 c_3$, and such a c_3 exists due to the compactness of I_3 and the fact that $E[W^{r^2}|W=\cdot]f_W(\cdot)$ is continuous and non-vanishing everywhere (Assumption MK (2)).

As $\frac{d}{dt}\left\{\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)}\right\} \in \mathbb{W}$ (Assumption MK (3)), Schennach (2004, Lemma 10) implies

$$\lim_{|u|\to\infty} u\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)} e^{{\rm i}tu}dt = 0,$$

and it follows

$$\int_{u\in I_3^c}\left|\int \frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)} e^{{\rm i}tu}dt\right|^2 du = O\left(\int_{u\in I_3^c} u^{-2}du\right) = O(1). \quad \text{(D.26)}$$

Thus, for all n large enough and some constant $C_3 > 0$, we have

$$\begin{aligned} E|\omega'\xi^*_{3,j}|^2 &\geq C_3\int\left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)} e^{{\rm i}tu}dt\right|^2 du \\ &= 2\pi C_3\int\left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)}\right|^2 dt, \end{aligned} \quad \text{(D.27)}$$

where the first step follows by (D.25) and (D.26) and the last step follows by Parseval's identity.

Thus, (D.21), (D.24), and (D.27) together with (A.21) implies

$$\begin{aligned} &\{Var(\omega'(\xi_j + \hat{\xi}_j))\}^{-(1+\eta/2)} \\ &= O\left(\left(\begin{array}{c} a_n^{-(1+2\alpha)} + \int\left|\frac{\{\omega' b(\cdot)\}^{\rm ft}(t)K^{\rm ft}(ta_n)}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)}\right|^2 dt \\ + \int\left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds\frac{d}{dt}\{f^{\rm ft}_{W^*}(t)\}}{f^{\rm ft}_\epsilon(t)\{f^{\rm ft}_{W^*}(t)\}^2}\right|^2 dt \\ + \int\left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-{\rm i}sw^*}ds}{f^{\rm ft}_\epsilon(t)f^{\rm ft}_{W^*}(t)}\right|^2 dt \end{array}\right)^{-(1+\eta/2)}\right). \end{aligned} \quad \text{(D.28)}$$

Combining (D.15) and (D.28), (D.13) holds under Assumption OK (3), and (D.12) follows.

For $\omega'(\hat{T}_n - T^*_n)$, by Ahn and Powell (1993, Lemma A.3), we have

$$\omega'(\hat{T}_n - T^*_n) = o_p(n^{-1/2}), \quad \text{(D.29)}$$

if

$$E|\omega'\{p_{j,l}+p_{l,j}\}|^2 = O(n). \tag{D.30}$$

To show (D.30), note that

$$\begin{aligned} E|\omega'\xi_{a,j,l}|^2 &= a_n^{-2}E\left[\left(\int |\Pi_l(t/a_n)|\frac{|K^{\mathrm{ft}}(t)|}{|f_\epsilon^{\mathrm{ft}}(t/a_n)|}dt\right)^2\right]E|\omega' V_j|^2 \\ &= O\left(a_n^{-2}E\left[\left(\sup_{|t|\leq a_n^{-1}}|\Pi_l(t)|\right)^2\right]\left(\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-2}\right) \\ &= O\left(a_n^{-2}\left(\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-6}\left(\inf_{|t|\leq a_n^{-1}}|f_{W^*}^{\mathrm{ft}}(t)|\right)^{-4}\right), \end{aligned} \tag{D.31}$$

where the first step follows by random sampling (Assumption M (1)), the second step follows from the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)), and the last step follows by Lemma 3. Then, we have

$$\begin{aligned} E|\omega'\{p_{j,l}+p_{l,j}\}|^2 &\leq 8\{E|\omega'\xi_{a,j,l}|^2+E|\omega'\xi_j|^2\} \\ &= O\left(a_n^{-2}\left(\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-6}\left(\inf_{|t|\leq a_n^{-1}}|f_{W^*}^{\mathrm{ft}}(t)|\right)^{-4}\right), \end{aligned} \tag{D.32}$$

where the first step follows by Jensen's inequality and the second step follows by (D.31) and (A.9). Under Assumptions O (1) and OK (1), (D.32) implies $E|\omega'\{p_{j,l}+p_{l,j}\}|^2 = O(a_n^{-2(1+3\alpha+2\alpha_w)})$ and (D.30) follows by $n^{-1/2}a_n^{-(2+3\alpha+2\alpha_w)} \to 0$ (Assumption OK (2)).

For $\omega' T_n^r$, note that

$$\begin{aligned} |\omega' T_n^r|^2 &= O\left(n^{-3}\{E|\omega'\xi_{a,1,1}|^2+E|\omega'\xi_1|^2\}\right) \\ &= O\left(n^{-3}a_n^{-2}\left(\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-6}\left(\inf_{|t|\leq a_n^{-1}}|f_{W^*}^{\mathrm{ft}}(t)|\right)^{-4}\right), \end{aligned} \tag{D.33}$$

where the last step follows by (D.31) and (A.9). Under Assumptions O (1) and OK (1), (D.33) implies

$$|\omega' T_n^r| = O_p(n^{-3/2}a_n^{-(1+3\alpha+2\alpha_w)}). \tag{D.34}$$

For $\omega'\hat{B}_n$, note that

$$
\begin{aligned}
|\omega'\hat{B}_n| &= \left|\frac{1}{2\pi n a_n}\int E\left[\omega' V_j e^{-\mathrm{i}t\left(\frac{w^*-W_j}{a_n}\right)}\Pi_j(t/a_n)\right]\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/a_n)}dt\right| \\
&\leq \frac{\{E|\omega' V_j|^2\}^{1/2}}{2\pi n a_n}\int \{E|\Pi_j(t/a_n)|^2\}^{1/2}\frac{|K^{\mathrm{ft}}(t)|}{|f_\epsilon^{\mathrm{ft}}(t/a_n)|}dt \\
&= O\left(n^{-1}a_n^{-1}\left(\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-1}\left\{E\left[\left(\sup_{|t|\leq a_n^{-1}}|\Pi_1(t/a_n)|\right)^2\right]\right\}^{1/2}\right) \\
&= O\left(n^{-1}a_n^{-2}\left(\inf_{|t|\leq a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\right)^{-3}\left(\inf_{|t|\leq a_n^{-1}}|f_{W^*}^{\mathrm{ft}}(t)|\right)^{-2}\right), \qquad \text{(D.35)}
\end{aligned}
$$

where the first step follows from the definitions of $\hat{B}_{n,s}$ and $\xi_{a,j,j}$ and random sampling (Assumption M (1)), the second step uses Cauchy-Schwarz inequality, the third step follows by the fact that K^{ft} is supported on $[-1,1]$ (Assumption M (3)), and the last step follows by Lemma 3. Under Assumptions O (1) and OK (1), (D.35) implies

$$|\hat{B}_n| = O(n^{-1}a_n^{-(2+3\alpha+2\alpha_w)}). \qquad \text{(D.36)}$$

The conclusion then follows by

$$\frac{n^{1/2}\omega'\{\hat{T}_n - T_n^* + T_n^r + B_n + \hat{B}_n + \hat{R}_n\}}{2\sqrt{Var(\omega'(\xi_j + \xi_{1,j}^* + \xi_{2,j}^* + \xi_{3,j}^*))}} = o_p(1), \qquad \text{(D.37)}$$

for which we combine (B.8), (A.13), (D.29), (D.34), (D.36), and (D.28) with Assumption OK (3).

D.3. **Proof of Theorem 2 (i).** First, note that by $n^{-1/2}e^{2\mu a_n^{-\gamma}+2\mu_w a_n^{-\gamma_w}}a_n^{-1}\log(1/a_n) \to 0$ as $n\to\infty$ (Assumption SK (2)) and Lemma 2, under Assumptions S (1) and SK (1), (D.5) implies

$$|\hat{S}_n| = O_p\left(n^{-1/2}e^{3\mu a_n^{-\gamma}+2\mu_w a_n^{-\gamma_w}}a_n^{-2}\log(1/a_n)\right), \qquad \text{(D.38)}$$

which together with $n^{-1/2}e^{3\mu a_n^{-\gamma}+2\mu_w a_n^{-\gamma_w}}a_n^{-2}\log(1/a_n) \to 0$ as $n\to\infty$ (Assumption SK (2)) gives (D.4), and (D.3) follows.

Also note that by $n^{-1/2}e^{2\mu a_n^{-\gamma}+2\mu_w a_n^{-\gamma_w}}a_n^{-1}\log(1/a_n) \to 0$ as $n\to\infty$ (Assumption SK (2)) and Lemma 2, under Assumptions S (1) and SK (1), (D.7) implies

$$|\hat{R}_n|^2 = O_p\left(n^{-2}e^{10\mu a_n^{-\gamma}+8\mu_w a_n^{-\gamma_w}}a_n^{-6}\log(1/a_n)^4\right), \qquad \text{(D.39)}$$

and (D.9) implies

$$|\hat{A}_n + \hat{B}_n|^2 = O_p\left(n^{-1}e^{6\mu a_n^{-\gamma}+4\mu_w a_n^{-\gamma_w}}a_n^{-4}\log(1/a_n)^2\right). \qquad \text{(D.40)}$$

Combining (A.13), (A.24), (D.39), and (D.40) with (B.3), the conclusion follows.

D.4. **Proof of Theorem 2 (ii).** Under Assumptions S (1) and SK (1), (B.14) implies

$$E|\omega'(\xi_j + \hat{\xi}_j) - E[\omega'\xi_j]|^{2+\eta} = O\left(e^{2(2+\eta)\mu a_n^{-\gamma}+2(2+\eta)\mu_w a_n^{-\gamma_w}}a_n^{-2(2+\eta)}\right), \qquad \text{(D.41)}$$

and (B.19) together with (A.28) implies

$$\{Var(\omega'(\xi_j+\hat{\xi}_j))\}^{-(1+\eta/2)}$$

$$= \begin{cases} O\left(\left(\begin{array}{c} e^{2\mu a_n^{-\gamma}} a_n^{2(2+\theta)} + \int \left|\frac{\{\omega' b(\cdot)\}^{\rm ft}(t)K^{\rm ft}(ta_n)}{f_\epsilon^{\rm ft}(t)f_{W^*}^{\rm ft}(t)}\right|^2 dt \\ + \int \left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-isw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2}\right|^2 dt \\ + \int \left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-isw^*}ds}{f_\epsilon^{\rm ft}(t)f_{W^*}^{\rm ft}(t)}\right|^2 dt \end{array}\right)^{-(1+\eta/2)}\right) & \text{if } 1/3<\gamma<1 \\ O\left(\left(\begin{array}{c} e^{2\mu a_n^{-\gamma}} a_n^{2(\gamma\theta+\gamma-1)} + \int \left|\frac{\{\omega' b(\cdot)\}^{\rm ft}(t)K^{\rm ft}(ta_n)}{f_\epsilon^{\rm ft}(t)f_{W^*}^{\rm ft}(t)}\right|^2 dt \\ + \int \left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-isw^*}ds\frac{d}{dt}\{f_{W^*}^{\rm ft}(t)\}}{f_\epsilon^{\rm ft}(t)\{f_{W^*}^{\rm ft}(t)\}^2}\right|^2 dt \\ + \int \left|\frac{\{\int_t^\infty + \int_t^{-\infty}\}\{\omega' b(\cdot)\}^{\rm ft}(s)K^{\rm ft}(sa_n)e^{-isw^*}ds}{f_\epsilon^{\rm ft}(t)f_{W^*}^{\rm ft}(t)}\right|^2 dt \end{array}\right)^{-(1+\eta/2)}\right) & \text{if } 1\le\gamma\le 2 \end{cases} \quad \text{(D.42)}$$

Combining (D.41) and (D.42), (B.13) holds under Assumption SK (3), and (B.12) follows.

Under Assumptions S (1) and SK (1), (D.32) implies $E|\omega'\{p_{j,l}+p_{l,j}\}|^2 = O\left(a_n^{-2}e^{6\mu a_n^{-\gamma}+4\mu_w a_n^{-\gamma_w}}\right)$ and (D.30) follows by $n^{-1/2}e^{3\mu a_n^{-\gamma}+2\mu_w a_n^{-\gamma_w}}a_n^{-2}\to 0$ (Assumption SK (2)), (D.33) implies

$$|\omega' T_n^r| = O_p\left(n^{-3/2}e^{3\mu a_n^{-\gamma}+2\mu_w a_n^{-\gamma_w}}a_n^{-1}\right), \quad \text{(D.43)}$$

and (D.35) implies

$$|\hat{B}_n| = O\left(n^{-1}e^{3\mu a_n^{-\gamma}+2\mu_w a_n^{-\gamma_w}}a_n^{-2}\right). \quad \text{(D.44)}$$

Combining (B.21), (D.43), (B.31), (A.13), (D.44), and (D.42), (B.29) holds under Assumption SK (3). The conclusion then follows by (B.12) and (B.29).

Appendix E. Lemmas

Lemma 1. *Under Assumptions RK and MK (1), for* $\iota = 1,2,3$, *it holds*

$$\sup_{|t|\le a_n^{-1}} |\hat{\delta}_\iota(t)| = O_p\left(n^{-1/2}\log(1/a_n)\right).$$

Proof. The conclusion follows by $E|W^*|^{2+\eta}<\infty$ and $E|\epsilon|^{2+\eta}<\infty$ for some $\eta>0$ (Assumption MK (1)) and Lemma 1 of Kurisu and Otsu (2020). □

Lemma 2. *Under Assumptions RK and MK (1), it holds*

$$\sup_{|t|\le b_n^{-1}} |\hat{\Pi}(t)| = O_p\left(\frac{\log(1/a_n)}{n^{1/2}a_n\{\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\rm ft}(t)|\}^2\{\inf_{|t|\le a_n^{-1}}|f_{W^*}^{\rm ft}(t)|\}^2}\right).$$

Moreover, if $\frac{n^{-1/4}a_n^{-1/2}\log(1/a_n)^{1/2}}{\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\rm ft}(t)|\inf_{|t|\le a_n^{-1}}|f_{W^*}^{\rm ft}(t)|}\to 0$, *it holds*

$$\sup_{|t|\le a_n^{-1}} |\hat{\Pi}^{\rm res}(t)| = O_p\left(\frac{\log(1/a_n)^2}{na_n^2\{\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\rm ft}(t)|\}^4\{\inf_{|t|\le a_n^{-1}}|f_{W^*}^{\rm ft}(t)|\}^4}\right).$$

Proof. The first statement then follows by

$$\sup_{|t|\le a_n^{-1}} |\hat{\Pi}(t)| = O_p\left(\frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)|} + a_n^{-1}\left\{\frac{\sup_{|t|\le a_n^{-1}} |\mu_3(t)| \sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\{\inf_{|t|\le a_n^{-1}} |\mu_2(t)|\}^2} + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_3(s)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(s)|}\right\}\right)$$
$$= O_p\left(\frac{\log(1/a_n)}{n^{1/2} a_n \{\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\}^2 \{\inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|\}^2}\right),$$

where the last step uses Lemma 1, $\inf_{|t|\le a_n^{-1}} |\mu_\iota(t)| \ge \inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)| \inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|$ for $\iota = 1, 2$ (Assumption RK), and $\sup_{|t|\le a_n^{-1}} |\mu_3(t)| = O(1)$ (Assumption MK (1)).

For the second statement, note that under $\frac{n^{-1/4} a_n^{-1/2} \log(1/a_n)^{1/2}}{\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)| \inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|} \to 0$ as $n \to \infty$,

$$\sup_{|t|\le a_n^{-1}} |\bar{\phi}(t)| = O_p\left(\begin{array}{c} a_n^{-1}\left\{\frac{\sup_{|t|\le a_n^{-1}} |\mu_3(t)| \sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\{\inf_{|t|\le a_n^{-1}} |\mu_2(t)|\}^2} + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_3(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(t)|}\right\} \\ \times\left\{1 + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(t)+\hat{\delta}_2(t)|}\right\} \end{array}\right),$$
$$= O_p\left(\frac{n^{-1/2} a_n^{-1} \log(1/a_n)}{\{\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\}^2 \{\inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|\}^2}\right) = o_p(1),$$

where the second step uses Lemma 1, $\inf_{|t|\le a_n^{-1}} |\mu_2(t)| \ge \inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)| \inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|$, and $\sup_{|t|\le a_n^{-1}} |\mu_3(t)| = O(1)$ (Assumption MK (1)), which implies $\sup_{|t|\le a_n^{-1}} e^{|\bar{\phi}(t)|} = O_p(1)$.

The conclusion then follows by $\frac{n^{-1/4} a_n^{-1/2} \log(1/a_n)^{1/2}}{\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)| \inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|} \to 0$ as $n \to \infty$ and

$$\sup_{|t|\le a_n^{-1}} |\hat{\Pi}^{\mathrm{res}}(t)| = O_p\left(\begin{array}{c} \frac{\{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|\}^2}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)+\hat{\delta}_1(t)|} + a_n^{-1}\left\{\frac{\sup_{|t|\le a_n^{-1}} |\mu_3(t)| \sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\{\inf_{|t|\le a_n^{-1}} |\mu_2(t)|\}^2} + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_3(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(t)|}\right\} \\ \times\left\{\begin{array}{c} \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)|} + \frac{\{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|\}^2}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)+\hat{\delta}_1(t)|} + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(t)+\hat{\delta}_2(t)|} \\ + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(t)+\hat{\delta}_2(t)|}\left\{\frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)|} + \frac{\{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|\}^2}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)+\hat{\delta}_1(t)|}\right\} \end{array}\right\} \\ + a_n^{-2}\left\{\frac{\sup_{|t|\le a_n^{-1}} |\mu_3(t)| \sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\{\inf_{|t|\le a_n^{-1}} |\mu_2(t)|\}^2} + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_3(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(t)|}\right\}^2 \\ \times\left\{1 + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_2(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_2(t)+\hat{\delta}_2(t)|}\right\}^2 \left\{1 + \frac{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)|} + \frac{\{\sup_{|t|\le a_n^{-1}} |\hat{\delta}_1(t)|\}^2}{\inf_{|t|\le a_n^{-1}} |\mu_1(t)+\hat{\delta}_1(t)|}\right\} \end{array}\right)$$
$$= O_p\left(\frac{\log(1/a_n)^2}{n a_n^2 \{\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|\}^4 \{\inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|\}^4}\right),$$

where the last step uses Lemma 1, $\inf_{|t|\le a_n^{-1}} |\mu_\iota(t)| \ge \inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)| \inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|$ for $\iota = 1, 2$ (Assumption RK), and $\sup_{|t|\le a_n^{-1}} |\mu_3(t)| = O(1)$ (Assumption MK (1)). □

Lemma 3. *Under Assumptions RK and MK (1), for $\eta > 0$, it holds*

$$E\left[\left(\sup_{|t|\le a_n^{-1}} |\Pi_1(t)|\right)^{2+\eta}\right] = O\left(a_n^{-(2+\eta)}\left(\inf_{|t|\le a_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)| \inf_{|t|\le a_n^{-1}} |f_{W^*}^{\mathrm{ft}}(t)|\right)^{-2(2+\eta)}\right).$$

Proof. The conclusion follows by

$$E\left[\left(\sup_{|t|\le a_n^{-1}}|\Pi_1(t)|\right)^{2+\eta}\right] = E\left[\left(\sup_{|t|\le a_n^{-1}}\left|-\frac{\delta_{1,1}(t)}{\mu_1(t)} + \mathrm{i}\int_0^t \left\{-\frac{\mu_3(s)\delta_{2,1}(s)}{\mu_2^2(s)} + \frac{\delta_{3,1}(s)}{\mu_2(s)}\right\} ds\right|\right)^{2+\eta}\right]$$

$$\le E\left[\left(\frac{\sup_{|t|\le a_n^{-1}}|\delta_{1,1}(t)|}{\inf_{|t|\le a_n^{-1}}|\mu_1(t)|} + a_n^{-1}\left\{\frac{\sup_{|t|\le a_n^{-1}}|\mu_3(t)|\sup_{|t|\le a_n^{-1}}|\delta_{2,1}(t)|}{\{\inf_{|t|\le a_n^{-1}}|\mu_2(t)|\}^2} + \frac{\sup_{|t|\le a_n^{-1}}|\delta_{3,1}(t)|}{\inf_{|t|\le a_n^{-1}}|\mu_2(t)|}\right\}\right)^{2+\eta}\right]$$

$$= O\left(\frac{1}{\{\inf_{|t|\le a_n^{-1}}|\mu_1(t)|\}^{2+\eta}} + \frac{\{\sup_{|t|\le a_n^{-1}}|\mu_3(t)|\}^{2+\eta}}{a_n^{2+\eta}\{\inf_{|t|\le a_n^{-1}}|\mu_2(t)|\}^{2(2+\eta)}} + \frac{1}{a_n^{2+\eta}\{\inf_{|t|\le a_n^{-1}}|\mu_2(t)|\}^{2+\eta}}\right)$$

$$= O\left(a_n^{-(2+\eta)}\left(\inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\inf_{|t|\le a_n^{-1}}|f_{W^*}^{\mathrm{ft}}(t)|\right)^{-2(2+\eta)}\right),$$

where the first step follows by the definition of $\Pi_1(t)$, the second step uses the triangular inequality, the third step follows by $\sup_{|t|\le a_n^{-1}}|\delta_{\iota,1}(t)| \le 2$ for $\iota = 1,2$ and $E[\{\sup_{|t|\le a_n^{-1}}|\delta_{3,1}(t)|\}^{2+\eta}] \le 2^{2+\eta}E|W^r|^{2+\eta} < \infty$ (Assumption MK (1)), and Jensen's inequality, and the last step follows by $\inf_{|t|\le a_n^{-1}}|\mu_\iota(t)| \ge \inf_{|t|\le a_n^{-1}}|f_\epsilon^{\mathrm{ft}}(t)|\inf_{|t|\le a_n^{-1}}|f_{W^*}^{\mathrm{ft}}(t)|$ for $\iota = 1,2$ (Assumption RK) and $\sup_{|t|\le a_n^{-1}}|\mu_3(t)| = O(1)$ (Assumption MK (1)). □

References

[1] Kurisu, D. and T. Otsu (2020) On uniform convergence of deconvolution estimator from repeated measurements, Working paper.

[2] Ahn, H. and J.L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism, *Journal of Econometrics*, 58, 3-29.

[3] Kotlarski, I. (1967) On characterizing the gamma and the normal distribution, *Pacific Journal of Mathematics*, 20, 69-76.

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

[5] Schennach, S. M. (2004) Nonparametric regression in the presence of measurement error, *Econometric Theory*, 20, 1046-1093.

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

Email address: haod@smu.edu

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

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

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

Email address: lntaylor@econ.au.dk