

# 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 component 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*, nor do we assume that the error is symmetrically distributed. Instead, we suppose we have access to a repeated measurement of the noisy regressor and use 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. Finally, we study the small sample behaviour of our estimator in a simulation study.

## 1. INTRODUCTION

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

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

where  $Y$  is a scalar dependent variable,  $X = (X_1, \dots, X_k)'$  and  $Z$  are covariates,  $(\beta_0(\cdot), \dots, \beta_k(\cdot))$  are unknown functions of  $Z$ , and  $U$  is an error term.  $X$  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 way. 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 propose a deconvolution based estimator for the varying coefficient model when the covariates are mismeasured. In particular, we suppose that a subset of the covariates,  $Z$ , in the nonparametric component is contaminated with classical measurement error. We show that the estimator is consistent and asymptotically normally distributed under both ordinary smooth and supersmooth errors. We do not suppose that we know the distribution of the measurement error *a priori* but assume that we have access to an independent repeated measurement of the contaminated regressors. Furthermore, we do not assume that the measurement error is symmetrically distributed.

---

The authors 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).

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, 2019). 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 for 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 properties of varying coefficient models highlights their growing popularity (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 are put to great use in many applied settings (for example, Mamuneas, Savvides and Stengos, 2006, Heshmati, Kumbhakar and Sun, 2014, and Feng, Gao, Peng and Zhang, 2017, among many others). For empirical work, one of the biggest appeals is their similarity to conventional linear regression models, which also facilitates a straightforward interpretation of the estimation results.

From a more theoretical perspective, varying coefficient models mitigate the effects of the ‘curse-of-dimensionality’. Typically, a single covariate is used in the nonparametric component; for example, one may be interested in how an effect changes over the lifecycle of an individual. In this case, estimators of these effects converge at the rate  $\sqrt{nb_n}$ , where  $b_n$  refers to the bandwidth parameter. This is in contrast to a fully nonparametric model where the convergence rate is  $\sqrt{nb_n^d}$  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.

There is a rich literature on estimating varying coefficient models starting with the seminal papers of Cleveland, Grosse and Shyu (1991) and Hastie and Tibshirani (1993). These models have been extended in many different directions, including allowing for measurement error in the covariates. In many real-world applications, data are often error-prone. This may be the result of poor data collection, imprecise measurement instruments, or imperfection in survey responses. 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 of these cases,  $\sqrt{n}$  convergence is obtained for this finite-dimensional parameter using profile least squares estimation. 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 papers. 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 our estimator. Furthermore, the rate of convergence depends sensitively on the degree of smoothness of the measurement error density.

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 under this more general setting. Section 4 presents the small sample properties of our estimator in a simulation study. Finally, Section 5 concludes.

## 2. CASE OF KNOWN MEASUREMENT ERROR DISTRIBUTION

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

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

where  $X = (X_1, \dots, X_k)'\in \mathbb{R}^k$  is a vector of observable covariates,  $W^* \in \mathbb{R}$  is an error-free covariate, and  $\beta(\cdot) = (\beta_1(\cdot), \dots, \beta_k(\cdot))'$  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

$$W = W^* + \epsilon,$$

and  $\epsilon$  is a measurement error. In this paper, we assume the measurement error is classical; that is,  $W^*$  is independent of  $\epsilon$ . Although this is somewhat restrictive, it also facilitates a multiplicative form  $W = W^*\epsilon$  - allowing for the variance of  $\epsilon$  to depend on  $W^*$  - which can then be transformed to an additive structure by taking the natural logarithm. Furthermore, we do not require full independence, 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  $f_A^{\text{ft}}$  denotes the characteristic 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. By premultiplying (2.1) by  $X$  and taking the conditional expectation, the object of interest  $\beta(w^*)$  can be written as

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

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

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

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

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

and  $K: \mathbb{R} \rightarrow \mathbb{R}$  is an ordinary kernel function.

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

$$\hat{\beta}(w^*) = \left[ \sum_{j=1}^n X_j X_j' \mathbb{K} \left( \frac{w^* - W_j}{a_n} \right) \right]^{-1} \sum_{j=1}^n X_j Y_j \mathbb{K} \left( \frac{w^* - W_j}{a_n} \right). \quad (2.2)$$

Li *et al.* (2002) considered a similar estimator when there is no measurement error and  $W^*$  is directly observed. The main difference is that we employ the deconvolution kernel  $\mathbb{K}$  to deal with contamination in  $W^*$ .

Throughout this paper, we focus on the case where a single mismeasured covariate  $W^*$  enters the coefficient functions  $\beta$  to keep the notation simple. The proposed method, however, can easily be adapted to the multivariate case. In particular, when  $\beta$  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.,

$$Y = X' \beta(W^*, Z) + U, \quad E[U|X, W^*, Z] = 0, \quad (2.3)$$

we wish to estimate  $\beta(w^*, z)$  at a given point  $(w^*, z)' \in \mathbb{R}^{d_w + d_z}$  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  $\tilde{\epsilon} = (\epsilon_1, \dots, \epsilon_{d_w})$  is mutually independent and independent of  $W^*$ . Similar to the case when  $\beta$  is a univariate function of a scalar  $W^*$ , by premultiplying (2.3) 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

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

and  $K: \mathbb{R} \rightarrow \mathbb{R}$  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. Hereafter we focus on the estimator  $\hat{\beta}(w^*)$  in (2.2) and study its asymptotic properties.

**2.2. 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 behaviour of the characteristic function of the measurement error. The first is known as ordinary smooth error and is characterised by the characteristic function of the measurement error decaying to zero at some polynomial rate. The second is known as supersmooth error and is defined by an exponentially decaying characteristic function.

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

**Assumption M.**  $\{Y_j, X_j, W_j\}_{j=1}^n$  is an i.i.d. sample of  $(Y, X, W)$  satisfying (2.1), and  $\epsilon$  is independent of  $(Y, X, W^*)$ . The function  $\beta$ , the density  $f_{W^*}$  of  $W^*$ , and the conditional expectation  $E[X_{k_1}X_{k_2}|W^* = w^*]$  for  $k_1, k_2 = 1, \dots, k$  have  $p$  continuous, bounded, and integrable derivatives, where  $p$  is a positive integer.  $E[U^2|X, W^* = w^*]$  and  $E[X_{k_1}^2 X_{k_2}^2|W^* = w^*]$  for  $k_1, k_2 = 1, \dots, k$  are bounded. Also,  $\lambda_{\min}(E[XX'|W^* = w^*]) > 0$  for almost every  $w^*$ .

These constitute mild assumptions on the data, and smoothness and boundedness of the densities and conditional moments. The last condition guarantees the identification of  $\beta(w^*)$ .

We begin with the ordinary smooth error case and impose the following assumptions.

**Assumption OS.**

(1): There exist positive constants  $\alpha_\epsilon$  and  $c_{\epsilon,0}^{\text{OS}} \leq c_{\epsilon,1}^{\text{OS}}$  such that

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

(2):  $K : \mathbb{R} \rightarrow \mathbb{R}$  satisfies  $\int K(x)dx = 1$ ,  $\int x^p K(x)dx \neq 0$  for some positive integer  $p$ , and  $\int x^q K(x)dx = 0$  for all positive integer  $q < p$ . Also  $K^{\text{ft}}$  is compactly supported on  $[-1, 1]$ , symmetric around zero, and bounded.

(3):  $a_n \rightarrow 0$  and  $na_n^{1+2\alpha_\epsilon} \rightarrow \infty$  as  $n \rightarrow \infty$ .

(4):  $f_\epsilon^{\text{ft}}(t)|t|^{\alpha_\epsilon} \rightarrow c_\epsilon^{\text{OS}}$  as  $|t| \rightarrow \infty$  for some positive constant  $c_{\epsilon,0}^{\text{OS}} \leq c_\epsilon^{\text{OS}} \leq c_{\epsilon,1}^{\text{OS}}$ .

(5):  $E[|U|^{2+\varsigma}|X, W^* = w^*]$  and  $E[|X_{k_1}|^{2+\varsigma}|W^* = w^*]$  for  $k_1 = 1, \dots, k$  are bounded for some  $\varsigma > 0$ .  $E[X_{k_1}X_{k_2}U^2|W^* = w^*]$  and  $E[X_{k_1}^2 X_{k_2}^2|W^* = w^*]$  for  $k_1, k_2 = 1, \dots, k$  are continuous.

(6):  $na_n^{1+2\alpha_\epsilon+2p} \rightarrow 0$  as  $n \rightarrow \infty$ .

Assumption OS (1) says that  $f_\epsilon$  is ordinary smooth of order  $\alpha_\epsilon$ . Popular examples of ordinary smooth densities include the Laplace and gamma density. Assumption OS (2) concerns the kernel function  $K$ . We require  $K$  to be a higher-order kernel to control the estimation bias. Assumption OS (3) gives conditions on the bandwidth  $a_n$ . The first restriction is required for the bias term, and the second is to control the variance term. Assumption OS (4)-(6) are additional conditions to derive the asymptotic distribution of  $\hat{\beta}(w^*)$ . Assumption OS (4) characterises the exact tail behaviour of  $f_\epsilon^{\text{ft}}$ , which is typically needed to derive the distributional result for the deconvolution based estimators. Assumption OS (5) contains Lyapunov's conditions to apply the central limit theorem and some smoothness conditions for conditional moments. Assumption OS (6) 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 our deconvolution estimator  $\hat{\beta}(w^*)$  for  $\beta(w^*)$  are obtained as follows.

**Theorem 1.**

(i): Under Assumptions M and OS (1)-(3), it holds

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

(ii): Under Assumptions M and OS (1)-(6), it holds

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

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

$$S(w^*) = E[XX'|W^* = w^*]f_{W^*}(w^*),$$

$$\Sigma(w^*) = C \int E[XX'\{U + X'\{\beta(W^*) - \beta(w^*)\}\}^2|W^* = w^* - \eta]f_{W^*}(w^* - \eta)f_\epsilon(\eta)d\eta,$$

and  $C = (2\pi)^{-1}(c_\epsilon^{\text{os}})^{-2} \int |K^{\text{ft}}(t)|^2|t|^{2\alpha_\epsilon}dt$  is a constant that depends on both  $K$  and  $f_\epsilon$ .

Theorem 1 (i) characterises the  $L_2$ -risk property of our deconvolution estimator  $\hat{\beta}(w^*)$ . The second term  $a_n^{2p}$  in the convergence rate characterises the magnitude of the estimation bias, which is identical to that of the error-free case. The first term  $n^{-1}a_n^{-(1+2\alpha_\epsilon)}$  characterises the magnitude of the estimation variance. Compared to that of the error-free case, the estimation variance of  $\hat{\beta}(w^*)$  decays slower due to the term  $a_n^{-2\alpha_\epsilon}$ : a smoother error distribution, which is characterised by a larger  $\alpha_\epsilon$ , 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 (2019) 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  $S(w^*)^{-1}\Sigma(w^*)S(w^*)^{-1}$ . Note that  $f_\epsilon$  can be set as the Dirac delta function when there is no measurement error. So in the error-free context,  $\Sigma(w^*) = \int K^2(x)dx E[XX'U^2|W^* = w^*]f_{W^*}(w^*)$ , which implies that 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 assumptions.

**Assumption SS.** (1): For some positive constants  $c_\epsilon^{\text{ss}}$ ,  $\mu_\epsilon$ , and  $1/3 < \gamma_\epsilon \leq 2$ , it holds

$$f_\epsilon^{\text{ft}}(t) \rightarrow c_\epsilon^{\text{ss}} e^{-\mu_\epsilon|t|^{\gamma_\epsilon}} \quad \text{as } |t| \rightarrow \infty.$$

(2):  $K : \mathbb{R} \rightarrow \mathbb{R}$  satisfies  $\int K(x)dx = 1$ ,  $\int x^p K(x)dx \neq 0$  for some positive integer  $p$ , and  $\int x^q K(x)dx = 0$  for all positive integer  $q < p$ . Also  $K^{\text{ft}}$  is compactly supported on  $[-1, 1]$ , symmetric around zero, and satisfies  $K^{\text{ft}}(1-t) = At^\theta + o(t^\theta)$  as  $t \rightarrow 0$ , for some constants  $A$  and  $\theta \geq 0$ .

- (3):  $a_n \rightarrow 0$  and  $na_n^{2-2\gamma_\epsilon(1+\theta)} e^{-2\mu_\epsilon a_n^{-\gamma_\epsilon}} \rightarrow \infty$  as  $n \rightarrow \infty$ .
- (4):  $E[|U|^{2+\varsigma}|X, W^*]$ , and  $E[|X_{k_1}|^{2+\varsigma}|W^*]$  for  $k_1 = 1, \dots, k$  are bounded for some  $\varsigma > 0$ .
- (5): If  $1 \leq \gamma_\epsilon \leq 2$ , then  $na_n^{2-2\gamma_\epsilon(1+\theta)+2p} e^{-2\mu_\epsilon a_n^{-\gamma_\epsilon}} \rightarrow 0$  as  $n \rightarrow \infty$ . If  $1/3 < \gamma_\epsilon < 1$ , then  $na_n^{1-\gamma_\epsilon(1+2\theta)+2p} e^{-2\mu_\epsilon a_n^{-\gamma_\epsilon}} \rightarrow 0$  as  $n \rightarrow \infty$ . Also,  $na_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

Assumption SS (1) says that  $f_\epsilon$  is supersmooth. Popular examples of supersmooth densities include the Gaussian and Cauchy density. We restrict  $\gamma_\epsilon$  to be less than or equal to 2 to ensure that  $f_\epsilon$  is a density (Chung, 1975, 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 SS (2) is the same as in the ordinary smooth case with 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  $\theta = 3$ . Assumption SS (3) concerns the bandwidth  $a_n$ ; similar comments to the ordinary smooth case apply here. Equally, Assumption SS (4) and (5) are analogous to Assumption OS (5) and (6) in the ordinary smooth case.

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

**Theorem 2.**

(i): Suppose that Assumptions M and SS (1)-(3) hold true. Then

$$|\hat{\beta}(w^*) - \beta(w^*)|^2 = \begin{cases} O_p\left(n^{-1}a_n^{-2+2\gamma_\epsilon(1+\theta)}e^{2\mu_\epsilon a_n^{-\gamma_\epsilon}} + a_n^{2p}\right) & \text{for } 1 \leq \gamma_\epsilon \leq 2, \\ O_p\left(n^{-1}a_n^{-1+\gamma_\epsilon(1+2\theta)}e^{2\mu_\epsilon a_n^{-\gamma_\epsilon}} + a_n^{2p}\right) & \text{for } 1/3 < \gamma_\epsilon < 1. \end{cases}$$

(ii): Suppose that Assumptions M and SS (1)-(5) hold true. Then

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

where  $\Omega_n(w^*) = \text{Var}(\xi(w^*))$  and

$$\xi(w^*) = \frac{1}{a_n}X[U + X'\{\beta(W^*) - \beta(w^*)\}]\mathbb{K}\left(\frac{w^* - W}{a_n}\right).$$

$$\text{Also, } \Omega_n(w^*)^{-1/2} = \begin{cases} O\left(n^{1/2}a_n^{1-\gamma_\epsilon(1+\theta)}e^{-\mu_\epsilon a_n^{-\gamma_\epsilon}}\right) & \text{for } 1 \leq \gamma_\epsilon \leq 2 \\ O\left(n^{1/2}a_n^{\{1-\gamma_\epsilon(1+2\theta)\}/2}e^{-\mu_\epsilon a_n^{-\gamma_\epsilon}}\right) & \text{for } 1/3 < \gamma_\epsilon < 1 \end{cases}.$$

Similar comments to Theorem 1 apply here. However, in the supersmooth setting, we distinguish between two cases depending on the smoothness parameter  $\gamma_\epsilon$ . In particular, as also shown in van Es and Uh (2004), the limiting behaviour of our estimator changes at the 'Cauchy boundary', where  $\gamma_\epsilon = 1$ . Compared to the ordinary smooth case, the rate of convergence is much slower, reflecting the more difficult task of deconvolution in the presence of supersmooth contamination. Comparably slow rates have been observed in other supersmooth settings, such as Fan (1992) and Schennach (2004).

### 3. CASE OF UNKNOWN MEASUREMENT ERROR DISTRIBUTION

**3.1. Setup and estimator.** Assuming the measurement error distribution to be known is unrealistic in many applications. In this section, we consider the situation where  $f_\epsilon$  is unknown but repeated measurements on  $W^*$  are available. Suppose we have two independent noisy measurements of the error-free covariate  $W^*$ , i.e.,

$$W_j = W_j^* + \epsilon_j \quad \text{and} \quad W_j^r = W_j^* + \epsilon_j^r, \quad (3.1)$$

for  $j = 1, \dots, n$ . To identify the distribution of  $\epsilon$ , we impose the following assumption.

**Assumption R.**  $\epsilon^r$  has the same distribution as  $\epsilon$  and is independent of  $(Y, X, W^*)$ ,  $f_{W^*}$  and  $f_\epsilon$  are square-integrable,  $f_{W^*}^{\text{ft}}$  and  $f_\epsilon^{\text{ft}}$  vanish nowhere, and  $E[\epsilon] = 0$ .

These assumptions are common in the literature (e.g., Li and Vuong, 1998, and Kurisu and Otsu, 2019). Under Assumption R, Kotlarski's (1967) identity gives

$$f_{W^*}^{\text{ft}}(t) = \exp\left(\int_0^t \frac{iE[W^r e^{isW^*}]}{E[e^{isW^*}]} ds\right), \quad (3.2)$$

and the error distribution is identified by  $f_\epsilon^{\text{ft}}(t) = \frac{f_{W^*}^{\text{ft}}(t)}{f_{W^*}^{\text{ft}}(t)}$ . Based on an i.i.d. sample  $\{W_j, W_j^r\}_{j=1}^n$  of  $(W, W^r)$ ,  $f_\epsilon^{\text{ft}}(t)$  can be estimated by its sample analog (Li and Vuong, 1998)

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

where  $\hat{f}_W^{\text{ft}}(t) = n^{-1} \sum_{j=1}^n e^{itW_j}$  and  $\hat{f}_{W^*}^{\text{ft}}(t) = \exp\left(\int_0^t \frac{i \frac{1}{n} \sum_{j=1}^n W_j^r e^{isW_j}}{\frac{1}{n} \sum_{j=1}^n e^{isW_j}} ds\right)$ . By replacing  $f_\epsilon^{\text{ft}}(t)$  with  $\hat{f}_\epsilon^{\text{ft}}(t)$  in (2.2), we propose the following estimator

$$\tilde{\beta}(w^*) = \left[ \sum_{j=1}^n X_j X_j' \hat{\mathbb{K}}\left(\frac{w^* - W_j}{a_n}\right) \right]^{-1} \sum_{j=1}^n X_j Y_j \hat{\mathbb{K}}\left(\frac{w^* - W_j}{a_n}\right), \quad (3.3)$$

where  $\hat{\mathbb{K}}(x) = \frac{1}{2\pi} \int e^{-itx} \frac{K^{\text{ft}}(t)}{\hat{f}_\epsilon^{\text{ft}}(t/a_n)} dt$ . A similar plug-in estimator was used in Schennach (2004) for the regression function when the regressor is mismeasured and repeated noisy measurements are available.

**3.2. Asymptotic properties.** To analyse the asymptotic properties of  $\tilde{\beta}(w^*)$ , we use the following class of functions  $\mathbb{W}$ , which is introduced by Schennach (2004)

**Definition.** Let  $\mathbb{W}$  be the set of all functions  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  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 form  $|t|^c$ ,  $\text{sgn}(t)|t|^c$ ,  $\log|t|$ ,  $\sin(c_1 t)$ ,  $\cos(c_1 t)$ ,  $\exp(c_1 t^a)$  with  $c > 0$ ,  $c_1 \in \mathbb{R}$ , and  $a \in \mathbb{N}$ .

As Schennach (2004, p. 1062) argued, this class  $\mathbb{W}$  characterises a class of functions that are well behaved at infinity, and is useful to derive a lower bound for the asymptotic variance of the estimator. For the asymptotic distribution of  $\tilde{\beta}(w^*)$  (but not for the convergence rate), we impose the following assumptions for both the ordinary smooth and supersmooth error cases.



**Assumption M'.**

(1):  $f_W(w) > 0$ ,  $E[(W^r)^2|W = w] > 0$ , and  $E[(\lambda'X)^2\{U + X'\{\beta(W^*) - \beta(w^*)\}\}^2|W = w] > 0$  for all  $w \in \mathbb{R}$  and  $\lambda \in \mathbb{R}^k$  with  $\|\lambda\| = 1$ .

(2):  $\frac{dm_l^\lambda(t)}{dt} \in \mathbb{W}$  for each  $l = 1, 2, 3$  and  $\lambda \in \mathbb{R}^k$  with  $\|\lambda\| = 1$ , where

$$m_1^\lambda(t) = \frac{\lambda' A(t, w^*) df_{W^*}^{\text{ft}}(t)/dt}{f_{W^*}^{\text{ft}}(t)^2 f_\epsilon^{\text{ft}}(t)} + \frac{\lambda' \{E[XX'|W^* = \cdot]\{\beta(\cdot) - \beta(w^*)\} f_{W^*}(\cdot)\}^{\text{ft}}(t) e^{-itw^*}}{f_{W^*}^{\text{ft}}(t) f_\epsilon^{\text{ft}}(t)},$$

$$m_2^\lambda(t) = \frac{\lambda' A(t, w^*)}{f_{W^*}^{\text{ft}}(t) f_\epsilon^{\text{ft}}(t)}, \quad m_3^\lambda(t) = \frac{1}{f_\epsilon^{\text{ft}}(t)},$$

and

$$A(t, w^*) = \begin{cases} \int_t^\infty e^{-isw^*} \{E[XX'|W^* = \cdot]\{\beta(\cdot) - \beta(w^*)\} f_{W^*}(\cdot)\}^{\text{ft}}(s) ds & t \geq 0, \\ \int_t^{-\infty} e^{-isw^*} \{E[XX'|W^* = \cdot]\{\beta(\cdot) - \beta(w^*)\} f_{W^*}(\cdot)\}^{\text{ft}}(s) ds & t < 0. \end{cases}$$

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

$$\text{Var} \left( \sum_{l=1}^3 \lambda' \zeta_{l,1} \right) \geq \underline{c} \max_{l=1,2,3} \text{Var}(\lambda' \zeta_{l,1}),$$

where  $\zeta_{l,1}$  for  $l = 1, 2, 3$  are defined in (B.1) in Appendix.

Assumption M' contains further conditions to derive the distributional results of our estimator when the error distribution is unknown for both ordinary and supersmooth cases. We emphasise that the convergence rate results below do not require this assumption. Assumption M' (1)-(2) are regularity conditions to derive the lower bound for the estimation variance; these are also used in Schennach (2004, Assumptions 12 and 13). In particular, Assumption M' (2) is a smoothness condition, and it is essentially compatible with our existing conditions on the smoothness of  $f_\epsilon^{\text{ft}}$ , like Assumption OS (4) and SS (1), in the case of known measurement error distribution. To understand this, note that  $\frac{dm_3^\lambda(t)}{dt} = \frac{-df_\epsilon^{\text{ft}}(t)/dt}{f_\epsilon^{\text{ft}}(t)^2}$ . If  $f_\epsilon^{\text{ft}}(t) \propto (1 + |t|)^{-\alpha_\epsilon}$  for all  $t \in \mathbb{R}$ , then  $\frac{dm_3^\lambda(t)}{dt} \propto (1 + |t|)^{\alpha_\epsilon - 1}$ , which belongs to  $\mathbb{W}$  for any positive integer  $\alpha_\epsilon > 1$ . If  $f_\epsilon^{\text{ft}}(t) \propto e^{-\mu_\epsilon |t|^{\gamma_\epsilon}}$  for all  $t \in \mathbb{R}$ , then  $\frac{dm_3^\lambda(t)}{dt} \propto |t|^{\gamma_\epsilon - 1} e^{\mu_\epsilon |t|^{\gamma_\epsilon}}$ , which also belongs to  $\mathbb{W}$  for any integer  $\gamma_\epsilon \geq 1$ . Assumption M' (2), however, is stronger than Assumption OS (4) and SS (1) as it imposes further restrictions on the smoothness of  $f_\epsilon^{\text{ft}}$ ,  $f_{W^*}^{\text{ft}}$ , and  $\{E[XX'|W^* = \cdot]\{\beta(\cdot) - \beta(w^*)\} f_{W^*}(\cdot)\}^{\text{ft}}(t)$  and their derivatives by requiring  $\frac{dm_1^\lambda(t)}{dt}, \frac{dm_2^\lambda(t)}{dt} \in \mathbb{W}$ , which is used to derive the asymptotic distribution of  $\tilde{\beta}(w^*)$  as  $f_\epsilon$  is not directly known but estimated. Assumption M' (3) states that the variance of  $\tilde{\beta}(w^*)$  is of an order no less than any term in its asymptotic representation; this is also used in Schennach (2004, Assumption 14).

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

**Assumption OS'.**

(1): For some positive constants  $\alpha_w > 1$  and  $c_{w,1}^{\text{OS}} \geq c_{w,0}^{\text{OS}}$ , it holds

$$c_{w,0}^{\text{OS}} |t|^{-\alpha_w} \leq |f_{W^*}^{\text{ft}}(t)| \leq c_{w,1}^{\text{OS}} |t|^{-\alpha_w} \quad \text{for all } t \in \mathbb{R}.$$

(2):  $n^{-1} a_n^{-(6\alpha_w + 6\alpha_\epsilon + 2)} (\log a_n^{-1})^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

**(3):** *There exists a positive constant  $\varsigma$  such that for each  $\lambda \in \mathbb{R}^k$  with  $|\lambda| = 1$ , it holds*

$$\min \left\{ \begin{array}{l} n^{\frac{\varsigma}{2+\varsigma}} a_n^{4\alpha_w+2\alpha_\epsilon+4}, \\ n(\log a_n^{-1})^{-4} a_n^{6\alpha_w+4\alpha_\epsilon+6}, \\ n^{-1} a_n^{-2p} \end{array} \right\} \max_{l=1,2,3} \int |m_{n,l}^\lambda(t)|^2 dt \rightarrow \infty,$$

*as  $n \rightarrow \infty$ , where  $m_{n,l}^\lambda(t)$  for  $l = 1, 2, 3$  are defined in (B.12) in Appendix.*

Assumption OS' (1) assumes that  $f_{W^*}$  is ordinary smooth of order  $\alpha_w$ , which is introduced to guarantee the consistency of the estimated error characteristic function  $\hat{f}_\epsilon^{\text{ft}}$ . Since  $\hat{f}_\epsilon^{\text{ft}}$  is defined by the ratio of the empirical averages, to characterise its estimation error, we need both the upper and lower bounds of  $f_{W^*}^{\text{ft}}$ . Assumption OS' (2) guarantees that  $M_{XX}$  is consistently estimated when we use the estimated error characteristic function  $\hat{f}_\epsilon^{\text{ft}}$  in place of  $f_\epsilon^{\text{ft}}$ . Assumption OS' (3) contains further restrictions on the bandwidth in order to use the Lyapunov central limit theorem and to ensure the asymptotic negligibility of the estimation bias and the higher order terms in the estimation variance.

**Theorem 3.**

**(i):** *Under Assumptions  $M$ ,  $R$ , OS (1)-(3), and OS' (1)-(2), it holds*

$$|\tilde{\beta}(w^*) - \beta(w^*)|^2 = O_p \left( n^{-1} a_n^{-(4\alpha_w+2\alpha_\epsilon+4)} + n^{-2} a_n^{-(6\alpha_w+4\alpha_\epsilon+6)} (\log a_n^{-1})^4 + a_n^{2p} \right).$$

**(ii):** *Under Assumptions  $M$ ,  $M'$ ,  $R$ , OS (1)-(3), and OS' (1)-(3), it holds*

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

$$\text{where } V_n(w^*) = S(w^*)^{-1} \text{Var} \left( \sum_{l=1}^3 \zeta_{l,1} \right) S(w^*)^{-1} \text{ and } S(w^*) = E[XX' | W^* = w^*] f_{W^*}(w^*).$$

Theorem 3 (i) shows the  $L_2$ -risk of our estimator when the error characteristic function is estimated by Li and Vuong's (1998) estimator. The first term results from the estimation of the measurement error characteristic function and, under Assumption OS' (3), represents the dominant term. The second term is the approximation error from a linearisation needed to establish the asymptotic properties of the estimator. Finally, the third part is the usual bias term from the 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 Li and Vuong (1998). Interestingly, the convergence rate and asymptotic distribution are driven by the estimation error of the measurement error characteristic function; this is also seen in, for example, Schennach (2004) and Kurisu and Otsu (2019).

The convergence rates achieved in this setting are slower than those in Theorem 1 where the error distribution is known. Furthermore, Delaigle, Hall and Meister (2008) show for the nonparametric regression estimator that when the density of the measurement error is assumed to be unknown but symmetric - which allows for a simple estimator of the error characteristic function - the convergence rate is the same as if the error density was known (providing that the density of  $W^*$  is sufficiently smooth in relation to that of the measurement error density). We conjecture that a similar result holds in this varying coefficient setting so that the convergence rate given in Theorem 1 is maintained under a symmetric unknown error density. Theorem 3

shows that this rate deteriorates quite considerably when the symmetry assumption is relaxed, and the estimator of Li and Vuong (1998) is used for the error characteristic function. This is in line with previous research which shows a similar reduction in the convergence rate when symmetry is relaxed, and the error distribution is estimated, see, for example, Bonhomme and Robin (2010), Comte and Kappus (2015), and Kurisu and Otsu (2019).

We now turn to the case of super smooth measurement error and introduce the following additional assumptions.

**Assumption SS'.**

(1): For some positive constants  $\mu_w$ ,  $\gamma_w$ , and  $c_{w,0}^{\text{ss}} \leq c_{w,1}^{\text{ss}}$ , it holds

$$c_{w,0}^{\text{ss}} e^{-\mu_w |t|^{\gamma_w}} \leq f_{W^*}^{\text{ft}}(t) \leq c_{w,1}^{\text{ss}} e^{-\mu_w |t|^{\gamma_w}} \quad \text{for all } t \in \mathbb{R}.$$

(2):  $n^{-1} a_n^{-2} (\log a_n^{-1})^2 \exp(6\mu_w a_n^{-\gamma_w} + 4\mu_\epsilon a_n^{-\gamma_\epsilon}) \rightarrow 0$  as  $n \rightarrow \infty$ .

(3): There exists a positive constant  $\varsigma$  such that for each  $\lambda \in \mathbb{R}^k$  with  $|\lambda| = 1$ , it holds

$$\min \left\{ \begin{array}{l} n^{\frac{\varsigma}{2+\varsigma}} a_n^4 e^{-4\mu_w a_n^{-\gamma_w} - 2\mu_\epsilon a_n^{-\gamma_\epsilon}}, \\ n a_n^6 (\log a_n^{-1})^{-4} e^{-6\mu_w a_n^{-\gamma_w} - 4\mu_\epsilon a_n^{-\gamma_\epsilon}}, \\ n^{-1} a_n^{-2p} \end{array} \right\} \max_{l=1,2,3} \int |m_{n,l}^\lambda(t)|^2 dt \rightarrow \infty,$$

as  $n \rightarrow \infty$ , where  $m_{n,l}^\lambda(t)$  for  $l = 1, 2, 3$  are defined in (B.12) in Appendix.

Assumption SS' (1) assumes that  $f_{W^*}$  is supersmooth of order  $\gamma_w$ . Again, since  $\hat{f}_\epsilon^{\text{ft}}$  is defined by the ratio of the empirical averages, to characterise the estimation error, we need both the upper and lower bounds of  $f_{W^*}^{\text{ft}}$ . As in the ordinary smooth case, Assumption SS' (2) guarantees that  $M_{XX}$  is consistently estimated. Assumption SS' (3) contains similar further restrictions on the bandwidth as were used in the ordinary smooth setting.

**Theorem 4.**

(i): Under Assumptions  $M$ ,  $R$ ,  $SS$  (1)-(3), and  $SS'$  (1)-(2), if  $1 \leq \gamma_\epsilon \leq 2$ , then

$$|\hat{\beta}(w^*) - \beta(w^*)|^2 = O_p \left( n^{-1} a_n^{-4} e^{4\mu_w a_n^{-\gamma_w} + 2\mu_\epsilon a_n^{-\gamma_\epsilon}} + n^{-2} a_n^{-6} (\log a_n^{-1})^4 e^{6\mu_w a_n^{-\gamma_w} + 4\mu_\epsilon a_n^{-\gamma_\epsilon}} + a_n^{2p} \right).$$

If  $1/3 < \gamma_\epsilon < 1$ , then

$$|\hat{\beta}(w^*) - \beta(w^*)|^2 = O_p \left( n^{-1} a_n^{-4} e^{4\mu_w a_n^{-\gamma_w} + 2\mu_\epsilon a_n^{-\gamma_\epsilon}} + n^{-2} a_n^{-6} (\log a_n^{-1})^4 e^{6\mu_w a_n^{-\gamma_w} + 4\mu_\epsilon a_n^{-\gamma_\epsilon}} + a_n^{2p} \right).$$

(ii): Under Assumptions  $M$ ,  $M'$ ,  $R$ ,  $SS$  (1)-(3), and  $SS'$  (1)-(3),

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

where  $V_n(w^*) = S(w^*)^{-1} \text{Var}[\hat{T}_n^{\text{dom}}(w^*)] S(w^*)^{-1}$ , and  $\hat{T}_n^{\text{dom}}(w^*)$  is defined in Appendix.

Similar comments to Theorem 3 apply here. Again the dominant term comes from the estimation of the measurement error characteristic function, and Assumption SS' (3) ensures that the bias terms from the nonparametric estimation and linearisation are asymptotically negligible.

## 4. SIMULATION

In this section, we analyse the small sample properties of our deconvolution estimator using a Monte Carlo study. We focus on the estimator for the unknown measurement error case of Section 3 and consider the following data generating process

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

where  $(X_1, X_2)$  is drawn from  $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right)$  and independent of  $(W^*, U)$ ,  $U$  is drawn from  $N(0, 0.025)$  and independent of  $W^*$ .  $W^*$  is assumed unobservable, however we observe  $W = W^* + \epsilon_1$  and  $W^r = W^* + \epsilon_2$ , where  $(\epsilon_1, \epsilon_2)$  is mutually independent and independent of  $(X_1, X_2, W^*, U)$ . For the densities of  $W^*$  and  $(\epsilon_1, \epsilon_2)$ , we consider two cases. First, for the ordinary smooth setting,  $(\epsilon_1, \epsilon_2)$  have a zero mean Laplace distribution with standard deviation of  $1/3$ , and  $W^*$  also has a zero mean Laplace distribution with standard deviation of  $1$ . Second, for the supersmooth situation,  $(\epsilon_1, \epsilon_2)$  have a normal distribution with zero mean and standard deviation of  $1/3$ , and  $W^*$  is standard normal. We consider three different data generating processes based on  $\beta_1(\cdot)$  and  $\beta_2(\cdot)$ . In each case, we take  $\beta_1(w) = w$ , and for  $\beta_2(\cdot)$  we consider

$$\text{DGP1} : \beta_2(w) = 1,$$

$$\text{DGP2} : \beta_2(w) = w + w^2,$$

$$\text{DGP3} : \beta_2(w) = \cos(0.5\pi w).$$

Throughout this simulation study, we use the infinite-order flat-top kernel proposed by McMurray and Politis (2004), defined by its Fourier transform as

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

This kernel satisfies Assumptions OS (2) and SS (2) and exhibits more stable estimates than the sinc kernel and the kernel of Fan (1992) in our preliminary simulations (not reported). We provide results for two sample sizes,  $n = 250, 500$ , 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. We suggest the method proposed by Bissantz *et al.* (2007), which provides a sensible choice in our setting. In that paper, they explain that the dependence between the bandwidth and the  $L^\infty$  distance between a deconvolution kernel density estimate and the true density changes depending on whether the bandwidth is larger or smaller than the optimal choice. We see the same pattern in our varying coefficient context. Figure 1 (a) shows the relationship between the bandwidth and the  $L^\infty$  distance between our varying coefficient estimator for  $\beta_1(w) = w$  in DGP2 and the truth, while Figure 1 (b) shows the same relationship for our estimator of  $\beta_2(w) = w + w^2$  in DGP2.

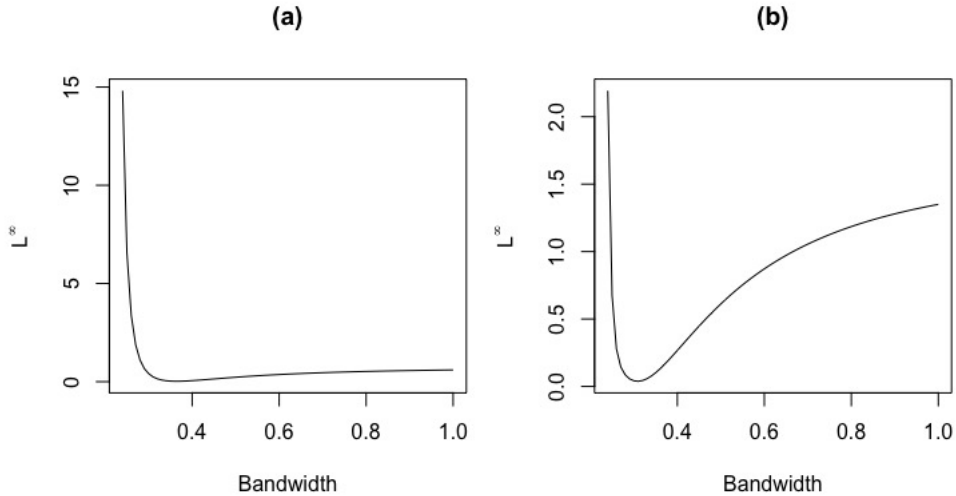


FIGURE 1. (a) shows the relationship between the bandwidth and the  $L^\infty$  distance between our varying coefficient estimator for  $\beta_1(w) = w$  in DGP2 and the truth. (b) shows the same relationship for  $\beta_2(w) = w + w^2$  in DGP2. In both plots, the sample size is 500, and the data is contaminated with Gaussian error.

This pattern is not an artefact of the data generating process; similar results are obtained for different functional forms and different types of measurement error. Bissantz *et al.* (2007) explained that for large bandwidths, the estimator becomes very smooth, so that changes in the bandwidth produce only minor changes in the  $L^\infty$  distance from the truth. However, when the bandwidth is smaller than the optimal choice “those frequencies in the spectral domain where the empirical characteristic function is dominated by noise in the data contribute increasingly more to the estimator, producing increasingly strong artificial oscillations” (Bissantz *et al.*, 2007, pp. 494). As a result, small changes in the bandwidth produce large changes in the  $L^\infty$  distance.

Given the close similarity to the findings of Bissantz *et al.* (2007), we pursue the same bandwidth selection approach in this paper. This involves estimating the varying coefficient functions for a range of bandwidths of decreasing size and choosing the largest bandwidth such that the change in the  $L^\infty$  distance for two estimators using consecutive bandwidths is larger than some threshold,  $T$ . In particular, following Kato and Sasaki (2018) (who also use this bandwidth choice method), we take the set of bandwidths to be  $a_j = ja_0/J$  for  $j = 1, \dots, J$  where  $J = 4 \log n$  and  $a_0$  is a pilot bandwidth that is larger than the optimal choice.

There is no ‘correct’ choice for the pilot bandwidth. Fortunately, it is easy to determine whether a given choice is large enough by plotting the  $L^\infty$  distance for two estimators using consecutive bandwidths (see Figure 2). Furthermore, in our setting, if the threshold,  $T$ , is chosen large enough, this method is surprisingly adept at recovering the optimal bandwidth from any suitably large choice of  $a_0$ . As such, we take  $a_0 = 0.5$  and choose  $T = d_{J-1,J}^{(\infty)} \log n$  where  $d_{J-1,J}^{(\infty)}$  is the  $L^\infty$  distance between the estimate using the pilot bandwidth and the estimate using the next largest bandwidth. This departs slightly from Bissantz *et al.* (2007) and Kato and Sasaki (2018) who choose  $T = 2d_{J-1,J}^{(\infty)}$  and  $T = 0.4d_{J-1,J}^{(\infty)} \log n$ , respectively, and  $a_0$  is the normal-reference

bandwidth (Delaigle and Gijbels, 2004).<sup>1</sup> With regards to  $T$ , we find that a larger threshold is better able to find the optimal bandwidth in our setting. In the next paragraph, we discuss another reason why a larger  $T$  is preferable in our setting.

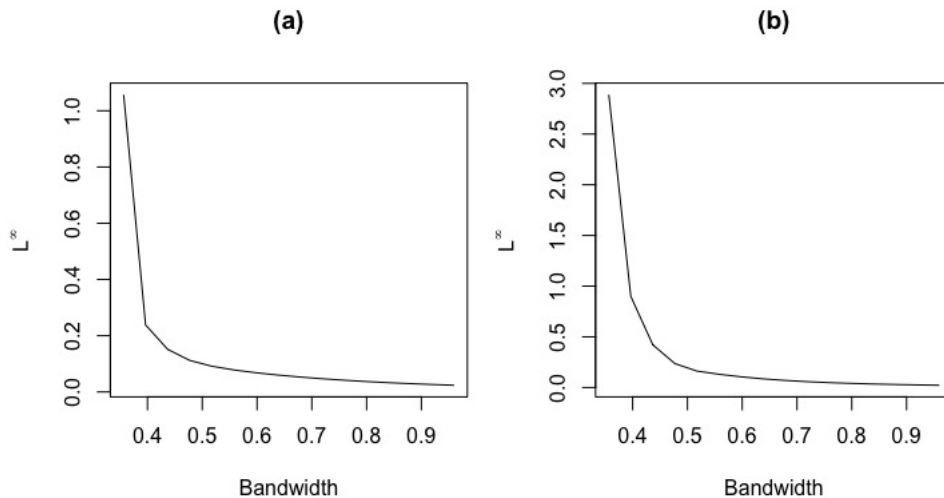


FIGURE 2. (a) Shows the relationship between the bandwidth and  $d_{j-1,j}^{(\infty)}$  for  $j = 1, \dots, 4 \ln(n)$  for the estimation of  $\beta_1(w) = w$  in DGP2. (b) shows the same relationship for  $\beta_2(w) = w + w^2$  in DGP2. In both plots, the sample size is 500, and the data is contaminated with Gaussian error.

In contrast to Bissantz *et al.* (2007) and Kato and Sasaki (2018) who are concerned with the estimation of a density, our focus is on a regression function. In particular, in this simulation study, we estimate two functions,  $\beta_1(\cdot)$  and  $\beta_2(\cdot)$ , and it may be that the  $L^\infty$  distance between each function and the truth is minimised by different bandwidths. However, given the form of our estimator, only one bandwidth can be chosen. Furthermore, since  $W^*$  is unobserved, it is not possible to choose the bandwidth to minimise the distance between  $E[Y|X, W^*]$  and its estimator.<sup>2</sup> We suggest to perform the above bandwidth selection procedure for each of  $\beta_1(\cdot)$  and  $\beta_2(\cdot)$ , respectively, then select the final bandwidth as the larger of the two. As seen in Figure 1, bandwidths that are smaller than the optimal can result in much poorer performance than larger bandwidths. Hence, we should be cautious of choosing a bandwidth smaller than the optimal. However, we must balance this with the knowledge that the bandwidth selected will be larger than optimal for one of the regression functions. Thus, it is important to select a bandwidth as small as possible without going beyond the optimal; a larger value for  $T$  is more likely to achieve this.

<sup>1</sup>There is no theory for selecting a normal-reference bandwidth for our estimator. Moreover, there is no reason to suppose this would result in a suitably large pilot bandwidth in any case.

<sup>2</sup>Although  $E[Y|X, W^*] = X_1 \hat{\beta}_1(W^*) + X_2 \hat{\beta}_2(W^*)$  can be calculated, the  $L^\infty$ -distance based bandwidth selection procedure cannot use this as  $\sup_{x,w} |E[Y|X = x, W^* = w; a_1] - E[Y|X = x, W^* = w; a_2]| = \sup_{x,w} |x_1 |\hat{\beta}_1(w; a_1) - \hat{\beta}_1(w; a_2)| + x_2 |\hat{\beta}_2(w; a_1) - \hat{\beta}_2(w; a_2)|| = \infty$  as  $X$  is normally distributed and is supported on  $\mathbb{R}$ .

To reduce the computational cost, the bandwidth selection procedure was conducted in 100 preliminary simulations, and the average value was chosen for all subsequent simulations.

**4.2. Results.** In Table 1, we report results for our varying coefficient estimator when the measurement error is assumed unknown. In particular, we calculate the median integrated squared error (MISE) over the range  $[-\sigma_{W^*}, \sigma_{W^*}]$ , and use a naive bootstrap to report the coverage of the confidence interval at  $w^* = 0$  with a nominal level of 95%.

DGP1		Ordinary Smooth		Supersmooth	
$n$	Function	MISE	Coverage	MISE	Coverage
250	$\beta_1(\cdot)$	0.09	94.0	0.24	97.2
	$\beta_2(\cdot)$	0.05	95.2	0.11	97.8
500	$\beta_1(\cdot)$	0.06	94.8	0.22	97.0
	$\beta_2(\cdot)$	0.03	94.4	0.06	96.2
DGP2					
250	$\beta_1(\cdot)$	0.15	94.6	0.50	96.6
	$\beta_2(\cdot)$	0.34	96.2	0.70	97.2
500	$\beta_1(\cdot)$	0.09	95.2	0.47	98.0
	$\beta_2(\cdot)$	0.26	95.8	0.68	91.4
DGP3					
250	$\beta_1(\cdot)$	0.10	93.2	0.37	97.2
	$\beta_2(\cdot)$	0.21	93.0	0.29	97.8
500	$\beta_1(\cdot)$	0.07	95.4	0.36	97.2
	$\beta_2(\cdot)$	0.17	93.0	0.21	97.2

The results are encouraging and seem to align with our theoretical findings. For each functional form and each sample size, the MISE is higher under supersmooth error than under ordinary smooth. This reflects the slower convergence rates of our estimator in the face of supersmooth error contamination found in Section 3. For the ordinary smooth case, the estimator performs well when the coefficient function is linear or constant ( $\beta_1(\cdot)$  in all DGPs and  $\beta_2(\cdot)$  in DGP1), but struggles more when faced with nonlinearity. This pattern is not as clear in the supersmooth setting. For example, in DGP3, the estimator is better able to estimate the nonlinear function,  $\beta_2(\cdot)$ , than the linear one. This is likely to be a result of the bandwidth selection mechanism choosing a bandwidth slightly smaller than optimal.

Indeed, the coverage probabilities also point to the possibility that the bandwidth choice for the supersmooth case is too small. In general, smaller bandwidths result in wider confidence intervals and, consequently, higher coverage probabilities. Whereas for ordinary smooth error, the coverage probabilities are close to the nominal level and suggest that perhaps the bandwidth selection procedure has worked well.

## 5. CONCLUSION

In this paper, we develop an estimator for the varying coefficient models 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 the case when the error density is known, and also the case when the error density is unknown but a repeated measurement of the noisy covariate is available. In both cases, we show that the proposed estimator is consistent and asymptotically normally distributed. In particular, when the error density is unknown, we use the approach of Li and Vuong (1998) based on Kotlarski's (1967) identity to estimate the error characteristic function, which allows the error density to be asymmetric. Although this approach significantly increases the applicability of the estimator, we show that it does reduce the convergence rate. Finally, the finite sample performance of the estimator is investigated by Monte Carlo simulation. We find that the estimator is more sensitive to the choice of bandwidth when the bandwidth is below the optimal level. As such, following Bissantz *et al.* (2007), we use a  $L^\infty$ -distance based approach to select the bandwidth. Under this choice, the estimator shows a good finite sample performance with only a mild requirement on the sample size.



APPENDIX A. PROOFS FOR SECTION 2

**Notation.** For the results in Section 2, define

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

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

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

**A.1. Proof of Theorem 1 (i).** Let  $S = E[XX'|W^* = w^*]f_{W^*}(w^*)$ ,  $\xi_j^{k_1}$  denote the  $k_1$ -th element of  $\xi_j$  for  $k_1 = 1, \dots, k$ , 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  $S$  for  $k_1, k_2 = 1, \dots, k$ , respectively. Observe that

$$|\hat{\beta}(w^*) - \beta(w^*)|^2 = |S_n^{-1}(T_n + B_n)|^2 \leq 2\lambda_{\max}(S_n^{-2})(|T_n|^2 + |B_n|^2).$$

Note that  $\lambda_{\max}(S_n^{-2}) = \{\lambda_{\min}(S_n)\}^{-2}$  and  $\lambda_{\min}(S_n) \geq \inf_{|\delta|=1} \delta'(S_n - S)\delta + \lambda_{\min}(S)$ . By  $\lambda_{\min}(S) > 0$  (Assumption M), if  $S_n \xrightarrow{p} S$ , we have

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

To show  $S_n \xrightarrow{p} S$ , note that

$$\begin{aligned}E[S_n^{k_1, k_2}] &= \frac{1}{2\pi a_n} \int e^{-itw^*/a_n} \{E[X_{k_1} X_{k_2} | W^* = \cdot] f_{W^*}(\cdot)\}^{\text{ft}}(t/a_n) K^{\text{ft}}(t) dt \\ &=_{(1)} \frac{1}{a_n} E \left[ X_{k_1} X_{k_2} K \left( \frac{w^* - W^*}{a_n} \right) \right] = \int \{E[X_{k_1} X_{k_2} | W^* = \cdot] f_{W^*}(\cdot)\} (w^* - a_n u) K(u) du \\ &=_{(2)} S^{k_1, k_2} + O(a_n^p),\end{aligned}\tag{A.3}$$

where (1) follows by Plancherel's isometry and (2) follows by the smoothness of  $E[X_{k_1} X_{k_2} | W^* = \cdot]$  and  $f_{W^*}$  (Assumption M) and the properties of the kernel functions  $K$  (Assumption OS (2)). Also, note that

$$\begin{aligned}\text{Var}(S_n^{k_1, k_2}) &\leq \frac{1}{na_n^2} E \left[ X_{k_1}^2 X_{k_2}^2 \mathbb{K}^2 \left( \frac{w^* - W}{a_n} \right) \right] \\ &= \frac{1}{na_n^2} \int \int \mathbb{K}^2 \left( \frac{w^* - u - \eta}{a_n} \right) \{E[X_{k_1}^2 X_{k_2}^2 | W^* = \cdot] f_{W^*}(\cdot)\} (u) f_\epsilon(\eta) du d\eta \\ &=_{(1)} \frac{1}{na_n^2} \int \mathbb{K}^2 \left( \frac{w^* - \tilde{u}}{a_n} \right) \int \{E[X_{k_1}^2 X_{k_2}^2 | W^* = \cdot] f_{W^*}(\cdot)\} (\tilde{u} - \eta) f_\epsilon(\eta) d\eta d\tilde{u} \\ &=_{(2)} O \left( \frac{1}{na_n^2} \int \mathbb{K}^2 \left( \frac{w^* - \tilde{u}}{a_n} \right) d\tilde{u} \right) =_{(3)} O \left( \frac{1}{na_n} \int \mathbb{K}^2(\tilde{u}) d\tilde{u} \right), \\ &=_{(4)} O(n^{-1} a_n^{-(1+2\alpha_\epsilon)}),\end{aligned}\tag{A.4}$$

where (1) follows by the change of variables  $\tilde{u} = u + \eta$ , (2) follows by the boundedness of  $E[X_{k_1}^2 X_{k_2}^2 | W^* = \cdot]$  and  $f_{W^*}$  (Assumption M), (3) follows by the change of variables  $\tilde{u} = \frac{w^* - \tilde{u}}{a_n}$ , (4) follows by Lemma 1. Thus, by (A.3), (A.4), and Assumption OS (3), we obtain  $S_n \xrightarrow{p} S$ .

For  $|T_n|^2$ , note that

$$\begin{aligned} \text{Var}(\xi_j^{k_1}) &= E \left[ E \left[ X_{k_1}^2 |U + X' \{\beta(W^*) - \beta(w^*)\}|^2 | W^* \right] \mathbb{K}^2 \left( \frac{w^* - W}{a_n} \right) \right] \\ &=_{(1)} O \left( \frac{1}{a_n} \int \mathbb{K}^2(\tilde{u}) d\tilde{u} \right) =_{(2)} O(a_n^{-(1+2\alpha_\epsilon)}), \end{aligned}$$

where (1) follows by the boundedness of  $E[X_{k_1}^2 X_{k_2}^2 | W^* = \cdot]$ ,  $E[U^2 | X, W^* = \cdot]$ , and  $\beta$  as in Assumption M, and (2) follows by Lemma 1. Thus, by  $|T_n|^2 = O_p \left( n^{-1} \max_{k_1} \text{Var}(\xi_j^{k_1}) \right)$ , we obtain

$$|T_n(w^*)|^2 = O_p(n^{-1} a_n^{-(1+2\alpha_\epsilon)}). \quad (\text{A.5})$$

For  $|B_n|^2$ , note that

$$\begin{aligned} E[\xi_j] &=_{(1)} \frac{1}{a_n} E \left[ X X' [\beta(W^*) - \beta(w^*)] \mathbb{K} \left( \frac{w^* - W}{a_n} \right) \right] \\ &= \frac{1}{2\pi a_n} \int e^{-itw^*/a_n} \{E[XX'\beta(W^*) | W^* = \cdot] f_{W^*}(\cdot)\}^{\text{ft}}(t/a_n) K^{\text{ft}}(t) dt \\ &\quad - \frac{\beta(w^*)}{2\pi a_n} \int e^{-itw^*/a_n} \{E[XX' | W^* = \cdot] f_{W^*}(\cdot)\}^{\text{ft}}(t/a_n) K^{\text{ft}}(t) dt \\ &= \frac{1}{a_n} E \left[ XX'\beta(W^*) K \left( \frac{w^* - W^*}{a_n} \right) \right] - \frac{\beta(w^*)}{a_n} E \left[ XX' K \left( \frac{w^* - W^*}{a_n} \right) \right] \\ &= \frac{1}{a_n} \int \{E[XX' | W^*] f_{W^*} \beta\}(u) K \left( \frac{w^* - u}{a_n} \right) du - \frac{\beta(w^*)}{a_n} \int \{E[XX' | W^*] f_{W^*}\}(u) K \left( \frac{w^* - u}{a_n} \right) du \\ &=_{(2)} \int \{E[XX' | W^*] f_{W^*} \beta\}(w^* - a_n \tilde{u}) K(\tilde{u}) d\tilde{u} - \beta(w^*) \int \{E[XX' | W^*] f_{W^*}\}(w^* - a_n \tilde{u}) K(\tilde{u}) d\tilde{u} \\ &=_{(3)} O(a_n^p), \end{aligned} \quad (\text{A.6})$$

where (1) follows by  $E[U | X, W^*] = 0$ , (2) follows by the change of variable  $\tilde{u} = \frac{w^* - u}{a_n}$ , and (3) follows by the smoothness of  $E[X_{k_1} X_{k_2} | W^* = \cdot]$ ,  $f_{W^*}$ , and  $\beta$  and the properties of the kernel function  $K$  from Assumption OS (2). Thus, by  $|B_n|^2 \leq k \max_{k_1} E^2[\xi_j^{k_1}]$ , we obtain

$$|B_n|^2 = O(a_n^{2p}). \quad (\text{A.7})$$

Combining (A.2), (A.5), and (A.7), the conclusion follows.

**A.2. Proof of Theorem 1 (ii).** First, we show the asymptotic normality of  $T_n$ , i.e.,

$$\sqrt{na_n^{1+2\alpha_\epsilon}} T_n \xrightarrow{d} N(0, \Sigma). \quad (\text{A.8})$$

By the Cramér–Wold device, (A.8) is equivalent to

$$\sqrt{na_n^{1+2\alpha_\epsilon}} \lambda' T_n \xrightarrow{d} N(0, \lambda' \Sigma \lambda), \quad (\text{A.9})$$

for each  $\lambda \in \mathbb{R}^k$  with  $|\lambda| = 1$ . We first establish the normalised version of (A.9), i.e.,

$$\frac{\lambda' T_n}{\sqrt{\text{Var}(\lambda' T_n)}} \xrightarrow{d} N(0, 1). \quad (\text{A.10})$$

For (A.10), it suffices to check whether the Lyapunov condition is satisfied for some  $\varsigma > 0$ , i.e.,

$$\frac{E|\lambda'\xi_j - E[\lambda'\xi_j]|^2}{n^{\varsigma/2}\{Var(\lambda'\xi_j)\}^{1+\varsigma/2}} \rightarrow 0. \quad (\text{A.11})$$

Let  $g_{1,\varsigma}(u) = E[\lambda'X|^{2+\varsigma}\{|U|^{2+\varsigma} + |\{\beta(W^*) - \beta(w^*)\}'X|^{2+\varsigma}\}|W^* = u]f_{W^*}(u)$ . For the numerator of (A.11), we have

$$\begin{aligned} & E|\lambda'\xi_j - E[\lambda'\xi_j]|^{2+\varsigma} \leq_{(1)} 2^{2+\varsigma} E|\lambda'\xi_j|^{2+\varsigma} \\ & \leq_{(2)} 2^{3+2\varsigma} a_n^{-2-\varsigma} \int \int \left| \mathbb{K}\left(\frac{w^* - u - \eta}{a_n}\right) \right|^{2+\varsigma} g_{1,\varsigma}(u) f_\epsilon(\eta) dud\eta \\ & =_{(3)} 2^{3+2\varsigma} a_n^{-2-\varsigma} \int \left| \mathbb{K}\left(\frac{w^* - \tilde{u}}{a_n}\right) \right|^{2+\varsigma} \int g_{1,\varsigma}(\tilde{u} - \eta) f_\epsilon(\eta) d\eta d\tilde{u} \\ & =_{(4)} O\left(a_n^{-2-\varsigma} \int \left| \mathbb{K}\left(\frac{w^* - \tilde{u}}{a_n}\right) \right|^{2+\varsigma} d\tilde{u}\right) =_{(5)} O\left(a_n^{-1-\varsigma} \int |\mathbb{K}(\tilde{u})|^{2+\varsigma} d\tilde{u}\right) =_{(6)} O(a_n^{-1-2\alpha_\epsilon-\varsigma(1+\alpha_\epsilon)}), \end{aligned}$$

where (1) and (2) follow by Jensen's inequality, (3) follows by the change of variables  $\tilde{u} = u + \eta$ , (4) follows by the boundedness of  $\beta$ ,  $E[|U|^{2+\varsigma}|X, W^* = \cdot]$ , and  $E[|X_{k_1}|^{2+\varsigma}|W^* = \cdot]$  for  $k_1 = 1, \dots, k$  (Assumption OS (4)), (5) follows by the change of variables  $\tilde{u} = \frac{w^* - \tilde{u}}{a_n}$ , (6) follows by Assumption OS (2) and

$$\begin{aligned} \int |\mathbb{K}(\tilde{u})|^{2+\varsigma} d\tilde{u} &= \sup_{\tilde{u}} |\mathbb{K}(\tilde{u})|^\varsigma \int \mathbb{K}^2(\tilde{u}) d\tilde{u} =_{(1)} \left( \int \frac{|K^{\text{ft}}(t)|}{|f_\epsilon^{\text{ft}}(t/a_n)|} dt \right)^\varsigma O(a_n^{-2\alpha_\epsilon}) \\ &=_{(2)} O\left(a_n^{-2\alpha_\epsilon} \left( \inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|^\varsigma \right)^{-1}\right) =_{(3)} O\left(a_n^{-(2+\varsigma)\alpha_\epsilon}\right), \end{aligned}$$

where (1) follows by Lemma 1, (2) follows by Assumption OS (2), and (3) follows by Assumption OS (1).

Let  $g_2(u) = E[|\lambda'X\{U + X'\{\beta(W^*) - \beta(w^*)\}\}|^2|W^* = u]f_{W^*}(u)$ . For the denominator of (A.11), we have

$$\begin{aligned} E|\lambda'\xi_j|^2 &= \frac{1}{a_n^2} \int \int \mathbb{K}^2\left(\frac{w^* - u - \eta}{a_n}\right) g_2(u) f_\epsilon(\eta) dud\eta \\ &=_{(1)} \frac{1}{a_n^2} \int \mathbb{K}^2\left(\frac{w^* - \tilde{u}}{a_n}\right) \int g_2(\tilde{u} - \eta) f_\epsilon(\eta) d\eta d\tilde{u} \\ &=_{(2)} \frac{1}{2\pi c_\epsilon^{\text{os}2} a_n^{1+2\alpha_\epsilon}} \int g_2(w^* - \eta) f_\epsilon(\eta) d\eta \int |K^{\text{ft}}(t)|^2 |t|^{2\alpha_\epsilon} dt (1 + o(1)), \end{aligned}$$

where (1) follows by the change of variables  $\tilde{u} = u + \eta$ , (2) follows by the continuity of  $\int g_2(\tilde{u} - \eta) f_\epsilon(\eta) d\eta$  implied by Assumption M and Fan (1991, Lemma 2.1). Therefore, the Lyapunov condition in (A.11) is satisfied if  $na_n \rightarrow \infty$  as  $n \rightarrow \infty$ , which holds under Assumption OS (3).

To show (A.9), besides (A.10), we also need

$$na_n^{1+2\alpha_\epsilon} Var(T_n) \rightarrow \Sigma. \quad (\text{A.12})$$

For (A.12), decompose  $na_n^{1+2\alpha_\epsilon} Var(T_n) = V_{n,1} + V_{n,2}$ , where

$$V_{n,1} = a_n^{1+2\alpha_\epsilon} E[\xi_j \xi_j'], \quad V_{n,2} = -a_n^{1+2\alpha_\epsilon} E[\xi_j] E[\xi_j'].$$

For the convergence of  $V_{n,1}$ , it suffices to check the convergence of its elements. Let  $g_{3,k_1,k_2}(u) = E[X_{k_1}X_{k_2}\{U + X'\{\beta(W^*) - \beta(w^*)\}\}^2|W^* = u]f_{W^*}(u)$ . The convergence of its  $(k_1, k_2)$  element follows by

$$\begin{aligned}
& a_n^{2\alpha_\epsilon - 1} E \left[ X_{k_1} X_{k_2} \{U + X'\{\beta(W^*) - \beta(w^*)\}\}^2 \mathbb{K}^2 \left( \frac{w^* - W}{a_n} \right) \right] \\
&= a_n^{2\alpha_\epsilon - 1} \int \int \mathbb{K}^2 \left( \frac{w^* - u - \eta}{a_n} \right) g_3(u) f_\epsilon(\eta) du d\eta \\
&\stackrel{(1)}{=} a_n^{2\alpha_\epsilon - 1} \int \mathbb{K}^2 \left( \frac{w^* - \tilde{u}}{a_n} \right) \int g_3(\tilde{u} - \eta) f_\epsilon(\eta) d\eta d\tilde{u} \\
&\stackrel{(2)}{=} \frac{1}{2\pi c_\epsilon^{\text{os}^2}} \int g_{3,k_1,k_2}(w^* - \eta) f_\epsilon(\eta) d\eta \int |K^{\text{ft}}(t)|^2 |t|^{2\alpha_\epsilon} dt (1 + o(1)), \tag{A.13}
\end{aligned}$$

where (1) follows by the change of variables  $\tilde{u} = u + \eta$ , (2) follows by the continuity of  $\int g_{3,k_1,k_2}(\tilde{u} - \eta) f_\epsilon(\eta) d\eta$  implied by Assumption M and Fan (1991, Lemma 2.1).

For  $V_{n,2}$ , note that (A.6) implies  $a_n^{1/2 + \alpha_\epsilon} E[\xi_j] = O(a_n^{1/2 + \alpha_\epsilon + \varsigma})$ , which, by  $a_n \rightarrow 0$  as in Assumption OS (3), gives

$$V_{n,2} = o(1). \tag{A.14}$$

Thus, (A.13) and (A.14) imply

$$\Sigma_{k_1,k_1} = C \int g_{3,k_1,k_2}(w^* - \eta, z) f_\epsilon(\eta) d\eta,$$

where  $C = \frac{\int |K^{\text{ft}}(t)|^2 |t|^{2\alpha_\epsilon} dt}{2\pi c_\epsilon^{\text{os}^2}}$  is a constant that is independent of  $k_1$  and  $k_2$  but depends on  $K$  and  $f_\epsilon$ . Thus, by (A.10) and (A.12), we obtain (A.8).

Therefore, the conclusion follows by (A.8) and the asymptotic negligibility of  $B_n$  (i.e.,  $\sqrt{na_n^{1+2\alpha_\epsilon}} B_n \rightarrow 0$ ) which is an immediate result of (A.6) and Assumption OS (6).

**A.3. Proof of Theorem 2 (i).** In this proof, we leverage the approach of van Es and Uh (2004) and Uh (2003). Uh (2003, Theorem 4.7) shows that

$$\mathbb{K} \left( \frac{w^* - W_j}{a_n} \right) = \tilde{C} c(a_n) a_n^{\theta\gamma_\epsilon} \left[ \begin{array}{c} \cos(D_j) \Re \Psi_\theta \left( \frac{a_n^{\gamma_\epsilon - 1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \\ - \sin(D_j) \Im \Psi_\theta \left( \frac{a_n^{\gamma_\epsilon - 1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \end{array} \right] + O_p(a_n^{\gamma_\epsilon(1+\theta)} c(a_n)),$$

where  $\tilde{C} = \frac{\Gamma(\theta+1)}{c_\epsilon^{\text{ss}} \pi (\mu_\epsilon \gamma_\epsilon)^\theta}$ ,  $D_j = \frac{W_j - w^*}{a_n} \bmod 2\pi$ ,  $c(a_n) = \int_0^1 e^{\mu_\epsilon (s/a_n)^{\gamma_\epsilon}} ds = O(a_n^{\gamma_\epsilon} e^{\mu_\epsilon (1/a_n)^{\gamma_\epsilon}})$ ,  $\Re$  and  $\Im$  denote the real and imaginary parts of a complex function, respectively, and  $\Psi_\theta(t) = 1/(1-it)^{1+\theta}$  is the characteristic function of the Gamma( $\theta + 1$ ) distribution. Note that

$$\int \{\Re \Psi_\theta(t)^2 + \Im \Psi_\theta(t)^2\} dt = \int |\Psi_\theta|^2 dt = \int_{-\infty}^{\infty} \frac{1}{(1+t^2)^{1+\theta}} dt < \infty.$$

We consider three separate cases based on the value of  $\gamma_\epsilon$ , i.e., (a)  $1 < \gamma_\epsilon \leq 2$ , (b)  $\gamma_\epsilon = 1$ , and (c)  $1/3 < \gamma_\epsilon < 1$ .

*Case (a)  $1 < \gamma_\epsilon \leq 2$ .* As in the proof of Theorem 1, if  $S_n \xrightarrow{P} S$ , we can write

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

Also from the proof of Theorem 1, for  $k_1, k_2 = 1, \dots, k$ , we have

$$E[S_n^{k_1, k_2}] = S^{k_1, k_2} + O(a_n^p). \quad (\text{A.15})$$

Using van Es and Uh (2004, Lemma 3.1), we can write

$$\begin{aligned} & \text{Var}(S_n^{k_1, k_2}) \\ \leq_{(1)} & \frac{1}{n} \left( \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \right)^2 E \left[ E[X_{k_1}^2 X_{k_2}^2 | W^*] \left\{ \begin{array}{c} \cos(D_j) \Re\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \\ - \sin(D_j) \Im\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \end{array} \right\}^2 \right] (1 + o(1)) \\ \sim_{(2)} & \frac{1}{n} \left( \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \right)^2 E [E[X_{k_1}^2 X_{k_2}^2 | W^*] \{\cos(U) \Re\Psi_\theta(0) - \sin(U) \Im\Psi_\theta(0)\}^2] (1 + o(1)) \\ =_{(3)} & \frac{1}{n} \left( \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \right)^2 \frac{1}{2} E[X_{k_1}^2 X_{k_2}^2 | W^*] (1 + o(1)) \\ =_{(4)} & O \left( \frac{c(a_n)^2 a_n^{2(\theta\gamma_\epsilon-1)}}{n} \right) = O \left( n^{-1} a_n^{2(1+\theta)\gamma_\epsilon-2} e^{2\mu_\epsilon(1/a_n)^{\gamma_\epsilon}} \right), \end{aligned} \quad (\text{A.16})$$

where  $U$  denotes a uniform random variable on  $[0, 2\pi]$  which is independent of  $(W^*, X)$ . Step (1) uses Uh (2003, Theorem 4.7), (2) uses in van Es and Uh (2004, Lemma 3.1), (3) follows from the independence of  $U$ ,  $E[\cos(U)] = E[\sin(U)] = 0$ ,  $E[\cos^2(U)] = E[\sin^2(U)] = 1/2$ ,  $\cos(x) \sin(x) = \frac{1}{2} \sin(2x)$ , and  $\Re\Psi_\theta(0)^2 + \Im\Psi_\theta(0)^2 < \infty$ , and (4) follows from Assumptions SS (1) and (3).

Thus, by (A.15), (A.16), and Assumption SS (5), we obtain  $S_n \xrightarrow{p} S$ .

We now turn to  $|T_n|^2$ . First, define

$$\psi_j = \left( \begin{array}{c} \cos(D_j) \Re\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \\ - \sin(D_j) \Im\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \end{array} \right) X_j \{U_j + X'_j \{\beta(W_j^*) - \beta(w^*)\}\},$$

so that

$$T_n = \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \frac{1}{n} \sum_{j=1}^n \{\psi_j - E[\psi_j]\} + O_p \left( n^{-1/2} a_n^{\gamma_\epsilon(2+\theta)-1} e^{\mu_\epsilon(1/a_n)^{\gamma_\epsilon}} \right), \quad (\text{A.17})$$

by Uh (2003, Theorem 4.7) and Assumption SS (1).

Now, using van Es and Uh (2004, Lemma 3.1), we can write

$$\begin{aligned} \text{Var}(\psi_j^{k_1}) &= E \left[ \left\{ \cos(D_j) \Re\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) - \Im\Psi_\theta(D_j) \Im\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \right\}^2 \right. \\ &\quad \left. \times E[X_{k_1} X_{k_2} \{U + X' \{\beta(W_j^*) - \beta(w^*)\}\}^2 | W^*] \right] \\ &\sim E \left[ \begin{array}{c} \{\cos(U) \Re\Psi_\theta(0) - \sin(U) \Im\Psi_\theta(0)\}^2 \\ \times E[X_{k_1} X_{k_2} \{U + X' \{\beta(W_j^*) - \beta(w^*)\}\}^2 | W^*] \end{array} \right] (1 + o(1)) \\ &= \frac{1}{2} E[X_{k_1} X_{k_2} \{U + X' \{\beta(W_j^*) - \beta(w^*)\}\}^2 | W^*] \{\Re\Psi_\theta(0)^2 + \Im\Psi_\theta(0)^2\} (1 + o(1)) = O(1), \end{aligned}$$

where the final equality uses Assumptions SS (1) and (3), and  $\Re\Psi_\theta(0)^2 + \Im\Psi_\theta(0)^2 < \infty$ . Thus, we obtain

$$|T_n|^2 = O_p \left( \frac{c(a_n)^2 k}{n a_n^2} \max_{k_1} \text{Var}(\psi_j^{k_1}) \right) = O_p \left( n^{-1} a_n^{2\{(1+\theta)\gamma_\epsilon-1\}} e^{2\mu_\epsilon(1/a_n)^{\gamma_\epsilon}} \right). \quad (\text{A.18})$$

For  $|B_n|^2$ , note that the method of proof used to show  $|B_n|^2 = O(a_n^{2p})$  in the proof of Theorem 1 did not depend on the distribution of the measurement error. Hence,  $|B_n(w^*)|^2 = O(a_n^{2p})$  in the supersmooth case as well. Therefore, the conclusion follows by (A.18).

*Case (b)*  $\gamma_\epsilon = 1$ . Using the same arguments as for the case of  $1 < \gamma_\epsilon \leq 2$ , we have

$$\begin{aligned}
& \text{Var}(S_n^{k_1, k_2}) \\
& \sim \frac{1}{n} \left( \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \right)^2 E \left[ E[X_{k_1}^2 X_{k_2}^2 | W^*] \left\{ \begin{array}{c} \cos(U)\Re\Psi_\theta \left( \frac{1}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) \\ -\sin(U)\Im\Psi_\theta \left( \frac{1}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) \end{array} \right\}^2 \right] (1 + o(1)) \\
& \stackrel{(1)}{=} \frac{1}{n} \left( \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \right)^2 E \left[ E[X_{k_1}^2 X_{k_2}^2 | W^*] \left\{ \begin{array}{c} \frac{1}{2}\Re\Psi_\theta \left( \frac{1}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right)^2 \\ +\frac{1}{2}\Im\Psi_\theta \left( \frac{1}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right)^2 \end{array} \right\} \right] (1 + o(1)) \\
& \stackrel{(2)}{=} O \left( \frac{c(a_n)^2 a_n^{2(\theta\gamma_\epsilon-1)}}{n a_n^2} \right) = O \left( n^{-1} a_n^{2(1+\theta)\gamma_\epsilon-2} e^{2\mu_\epsilon(1/a_n)^{\gamma_\epsilon}} \right),
\end{aligned}$$

where (1) follows from the independence of  $U$ ,  $E[\cos(U)] = E[\sin(U)] = 0$ ,  $E[\cos^2(U)] = E[\sin^2(U)] = 1/2$ ,  $\cos(x)\sin(x) = \frac{1}{2}\sin(2x)$ , and (2) follows from the boundedness of  $\Re\Psi_\theta^2 + \Im\Psi_\theta^2$ ,  $E[X_{k_1}^2 X_{k_2}^2 | W^* = \cdot]$  for  $k_1, k_2 = 1, \dots, k$  as in Assumptions SS (1) and (3). Again, we obtain  $S_n \xrightarrow{p} S$ .

For  $|T_n|^2$ , very similar arguments give

$$\begin{aligned}
& \text{Var}(\psi_j^{k_1}) \\
& \sim E \left[ \left\{ \cos(U)\Re\Psi_\theta \left( \frac{1}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) - \sin(U)\Im\Psi_\theta \left( \frac{1}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) \right\}^2 \right. \\
& \quad \left. \times E[X_{k_1} X_{k_2} \{U + X' \{ \beta(W_j^*) - \beta(w^*) \} \}^2 | W^*] \right] (1 + o(1)) = O(1),
\end{aligned}$$

and we conclude

$$|T_n|^2 = O_p \left( n^{-1} a_n^{2(1+\theta)\gamma_\epsilon-2} e^{2\mu_\epsilon(1/a_n)^{\gamma_\epsilon}} \right). \quad (\text{A.19})$$

For  $|B_n|^2$ , again, the proof from the ordinary smooth case shows  $|B_n|^2 = O(a_n^{2p})$ . Therefore, the conclusion follows by (A.19).

*Case (c)*  $1/3 < \gamma_\epsilon < 1$ . Finally, for  $\gamma_\epsilon < 1$ , we have

$$\begin{aligned}
& \text{Var}(S_n^{k_1, k_2}) \\
& \stackrel{(1)}{=} \frac{1}{n} \left( \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \right)^2 E \left[ E[X_{k_1} X_{k_2} | W^*] \left\{ \begin{array}{c} \cos(D_j)\Re\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) \\ -\sin(D_j)\Im\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) \end{array} \right\}^2 \right] (1 + o(1)) \\
& \leq \stackrel{(2)}{=} \frac{1}{n} \left( \tilde{C}c(a_n)a_n^{\theta\gamma_\epsilon-1} \right)^2 E \left[ (E[X_{k_1} X_{k_2} | W^*])^2 \right]^{1/2} \\
& \quad \times E \left[ \left\{ \begin{array}{c} \cos(D_j)\Re\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) \\ -\sin(D_j)\Im\Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon\gamma_\epsilon}(W_j - w^*) \right) \end{array} \right\}^4 \right]^{1/2} (1 + o(1)) \\
& \sim \stackrel{(3)}{=} O \left( \frac{c(a_n)^2 a_n^{2(\theta\gamma_\epsilon-1)} a_n^{(1-\gamma_\epsilon)/2}}{n} \right) E \left[ (E[X_{k_1} X_{k_2} | W^*])^2 \right]^{1/2} \stackrel{(4)}{=} O \left( \frac{c(a_n)^2 a_n^{2(\theta\gamma_\epsilon-1)} a_n^{(1-\gamma_\epsilon)/2}}{n} \right),
\end{aligned}$$

where (1) was shown in the case of  $1 < \gamma_\epsilon \leq 2$ , (2) uses the Cauchy-Schwartz inequality, (3) follows from Uh (2003, Theorem 4.9), and (4) uses Assumptions SS (1) and (3).

For  $|T_n|^2$ , very similar arguments give  $Var(\psi_j^{k_1}) = O(a_n^{(1-\gamma_\epsilon)/2})$ , which implies

$$|T_n|^2 = O_p\left(n^{-1}a_n^{\{(2\theta+1)\gamma_\epsilon-1\}}e^{2\mu_\epsilon(1/a_n)^{\gamma_\epsilon}}\right). \quad (\text{A.20})$$

Finally,  $|B_n|^2 = O(a_n^{2p})$  as shown in the proof of Theorem 1. Therefore, the conclusion follows by (A.20).

Notice that in the case of  $\gamma_\epsilon < 1$ , the approximation error in (A.17) is of smaller order than the asymptotic variance as long as  $\gamma_\epsilon > 1/3$ .

**A.4. Proof of Theorem 2 (ii).** To show the asymptotic normality of  $T_n$ , we again use the Cramér–Wold device and establish the distribution for the normalised version, i.e.,

$$\frac{\lambda' T_n}{\sqrt{Var(\lambda' T_n)}} \xrightarrow{d} N(0, 1), \quad (\text{A.21})$$

for each  $\lambda \in \mathbb{R}^k$  with  $|\lambda| = 1$ . For (A.21), it suffices to check whether the Lyapunov condition is satisfied for some  $\varsigma > 0$ , i.e.,

$$\frac{E|\lambda' \psi_j - E[\lambda' \psi_j]|^{2+\varsigma}}{n^{\varsigma/2}\{Var(\lambda' \psi_j)\}^{1+2/\varsigma}} \rightarrow 0. \quad (\text{A.22})$$

As for the proof of Theorem 2 (i), we consider three cases based on the value of  $\gamma_\epsilon$ , i.e., (a)  $1 < \gamma_\epsilon \leq 2$ , (b)  $\gamma_\epsilon = 1$ , and (c)  $1/3 < \gamma_\epsilon < 1$ .

*Case (a)  $1 < \gamma_\epsilon \leq 2$ .* For the denominator of (A.22), using the same arguments as those used to bound  $Var(\psi_j^{k_1})$  in the proof of Theorem 2 (i), we have

$$\begin{aligned} Var(\lambda' \psi_j) &\sim \frac{1}{2^d} E[\lambda' X X' \lambda \{U + X' \{\beta(W^*) - \beta(w^*)\}\}^2 | W^*] \\ &\times \{\Re \Psi_\theta(0)^2 + \Im \Psi_\theta(0)^2\} = O(1), \end{aligned}$$

where the final equality uses Assumptions SS (3) and (4), and  $\{\Re \Psi_\theta(0)^2 + \Im \Psi_\theta(0)^2\} < \infty$ .

For the numerator of (A.22), using similar arguments as used to bound the variance, we have

$$E|\lambda' \psi_j - E[\lambda' \psi_j]|^{2+\varsigma} \leq 2^{2+\varsigma} E|\lambda' \psi_j|^{2+\varsigma} = O(1).$$

Since the Lyapunov condition is satisfied, the conclusion follows. Also the order of  $\Omega_n(w^*)^{-1/2}$  follows from (A.18).

*Case (b)  $\gamma_\epsilon = 1$ .* Again, using the same arguments as for bounding  $Var(\psi_j^{k_1})$  in the proof of Theorem 2 (i), we have

$$Var(\lambda' \psi_j) \sim E \left[ \begin{aligned} &E[\lambda' X X' \lambda \{U + X' \{\beta(W^*) - \beta(w^*)\}\}^2 | W^*] \\ &\times \frac{1}{2} \left\{ \Re \Psi_\theta \left( \frac{1}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right)^2 + \Im \Psi_\theta \left( \frac{1}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right)^2 \right\} \end{aligned} \right] = O(1),$$

where the final equality follows from the boundedness of  $\Re \Psi_\theta^2 + \Im \Psi_\theta^2$ ,  $\beta(\cdot)$ ,  $E[|U|^2 | X, W^* = \cdot]$ , and  $E[|X_{k_1}|^2 | W^* = \cdot]$  for  $k_1 = 1, \dots, k$  as in Assumption SS (3) and (4).

For the numerator of the Lyapunov condition, we use similar arguments to show  $E[|\lambda'\psi_j|^{2+\varsigma}] = O(1)$ . Hence, again the Lyapunov condition is satisfied and the conclusion follows. Also the order of  $\Omega_n(w^*)^{-1/2}$  follows from (A.19).

*Case (c)*  $1/3 < \gamma_\epsilon < 1$ . For the denominator of A.22, we again use the results of the proof of Theorem 2 (i) to show

$$\begin{aligned} \text{Var}(\lambda'\psi_j) &= O(a_n^{(1-\gamma_\epsilon)/2}) E \left[ E[\lambda'X X'\lambda \{U + X'\{\beta(W^*) - \beta(w^*)\}\}^2 | W^*] \right]^{1/2} \\ &= O(a_n^{(1-\gamma_\epsilon)/2}), \end{aligned}$$

where the final equality uses the boundedness of  $\beta$ ,  $E[|U|^2 | X, W^*]$ , and  $E[|X_{k_1}|^2 | W^*]$  for  $k_1 = 1, \dots, k$  from Assumption SS (4).

For  $\varrho > 2 + \varsigma$ , we have

$$\begin{aligned} E[|\lambda'\psi_j|^{2+\varsigma}] &\leq_{(1)} E \left[ \left\{ \begin{array}{l} \cos(D_j) \Re \Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \\ - \sin(D_j) \Im \Psi_\theta \left( \frac{a_n^{\gamma_\epsilon-1}}{\mu_\epsilon \gamma_\epsilon} (W_j - w^*) \right) \end{array} \right\}^{\frac{(2+\varsigma)\varrho}{\varrho-2-\varsigma}} \right]^{(\varrho-2-\varsigma)/\varrho} \\ &\quad \times E [(\lambda'X \{U + X'\{\beta(W^*) - \beta(w^*)\}\})^\varrho]^{(2+\varsigma)/\varrho} \\ &=_{(2)} O \left( a_n^{(1-\gamma_\epsilon)(\varrho-2-\varsigma)/\varrho} \right) E [(\lambda'X \{U + X'\{\beta(W^*) - \beta(w^*)\}\})^\varrho]^{(2+\varsigma)/\varrho} \\ &=_{(3)} O \left( a_n^{(1-\gamma_\epsilon)(\varrho-2-\varsigma)/\varrho} \right), \end{aligned}$$

where (1) uses Hölder's inequality, (2) uses van Es and Uh (2004, Lemma 3.1), and (3) follows from similar arguments as used to bound the variance of  $\psi_j(w^*)$  in earlier parts of this proof. Taking  $\varrho = 4$ , the Lyapunov condition is satisfied if  $na_n^2 \rightarrow 0$  which is true under Assumption SS (5). This leads to the conclusion. Also the order of  $\Omega_n(w^*)^{-1/2}$  follows from (A.20).

## APPENDIX B. PROOFS FOR SECTION 3

**Notation.** Define

$$\begin{aligned} \hat{\mu}_1(t) &= \frac{1}{n} \sum_{j=1}^n e^{itW_j}, & \hat{\mu}_2(t) &= \frac{1}{n} \sum_{j=1}^n W_j^r e^{itW_j}, \\ \hat{\mu}_3(t) &= \frac{1}{n} \sum_{j=1}^n X_j [U_j + X_j' \{\beta(W_j^*) - \beta(w^*)\}] e^{itW_j}, \\ \mu_l(t) &= E[\hat{\mu}_l(t)], & \text{for } l &= 1, 2, 3, \\ \delta_l(t) &= \hat{\mu}_l(t) - \mu_l(t), & \text{for } l &= 1, 2, 3. \end{aligned}$$

Let  $\hat{S}_n$  and  $\hat{T}_n$  denote the objects obtained by replacing  $\mathbb{K}$  by  $\hat{\mathbb{K}}$  in  $S_n$  and  $T_n$ , respectively. Also define

$$\hat{T}_n^{\text{dom}} = \frac{1}{n} \sum_{j=1}^n \sum_{l=1}^3 (\zeta_{l,j} - E[\zeta_{l,j}]), \quad R_n = \hat{T}_n - \hat{T}_n^{\text{dom}},$$



where

$$\begin{aligned}
\zeta_{1,j} &= -\frac{1}{2\pi} \int \left\{ \frac{iA_n(t, w^*)\mu_2(t)}{\mu_1^2(t)} + e^{-itw^*} K^{\text{ft}}(ta_n) \frac{\mu_3(t) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right)}{\mu_1^2(t)} \right\} e^{itW_j} dt, \\
\zeta_{2,j} &= \frac{1}{2\pi} \int \frac{iA_n(t, w^*)}{\mu_1(t)} W_j^r e^{itW_j} dt, \\
\zeta_{3,j} &= \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \frac{\exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right)}{\mu_1(t)} X_j[U_j + X_j'\{\beta(W_j^*) - \beta(w^*)\}] e^{itW_j} dt, \\
A_n(t, w^*) &= \begin{cases} \int_t^\infty e^{-isw^*} K^{\text{ft}}(sa_n) \frac{\mu_3(s)}{\mu_1(s)} \exp\left(\int_0^s \frac{i\mu_2(u)}{\mu_1(u)} du\right) ds & t \geq 0 \\ \int_t^{-\infty} e^{-isw^*} K^{\text{ft}}(sa_n) \frac{\mu_3(s)}{\mu_1(s)} \exp\left(\int_0^s \frac{i\mu_2(u)}{\mu_1(u)} du\right) ds & t < 0. \end{cases}
\end{aligned} \tag{B.1}$$

Then the estimator  $\tilde{\beta}(w^*)$  is written as

$$\tilde{\beta}(w^*) = \beta(w^*) + \hat{S}_n^{-1}(\hat{T}_n^{\text{dom}} + R_n + B_n).$$

To understand  $\hat{T}_n^{\text{dom}}$ , note that

$$\hat{T}_n = \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \left\{ \exp\left(\int_0^t \frac{i\hat{\mu}_2(s)}{\hat{\mu}_1(s)} ds\right) \frac{\hat{\mu}_3(t)}{\hat{\mu}_1(t)} - \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) \frac{\mu_3(t)}{\mu_1(t)} \right\} dt,$$

and its Fréchet derivative as a functional of  $(\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3)$  at  $(\mu_1, \mu_2, \mu_3)$  in the direction of  $(\delta_1, \delta_2, \delta_3)$  is

$$\begin{aligned}
&\frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \frac{\mu_3(t)}{\mu_1(t)} \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) \int_0^t \left\{ \frac{i\delta_2(s)}{\mu_1(s)} - \frac{i\mu_2(s)\delta_1(s)}{\mu_1^2(s)} \right\} ds dt \\
&+ \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \left\{ \frac{\delta_3(t)}{\mu_1(t)} - \frac{\mu_3(t)\delta_1(t)}{\mu_1^2(t)} \right\} \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) dt.
\end{aligned}$$

Then  $\hat{T}_n^{\text{dom}}$  is the Fréchet derivative of  $\hat{T}_n$ , which follows by reorganising the terms with respect to  $\delta_l$  for  $l = 1, 2, 3$  and using the fact that  $\int_{-\infty}^\infty \int_0^t f(t, s) ds dt = \int_0^\infty \int_t^\infty f(s, t) ds dt + \int_{-\infty}^0 \int_t^{-\infty} f(s, t) ds dt$  for any absolutely integrable function  $f$ .

Note that  $\zeta_{3,j} = \xi_j$  defined in (A.1). So  $\frac{1}{n} \sum_{j=1}^n (\zeta_{3,j} - E[\zeta_{3,j}]) = T_n$  characterises the estimation variance when  $f_\epsilon$  is known, and  $\frac{1}{n} \sum_{j=1}^n \sum_{l=1}^2 (\zeta_{l,j} - E[\zeta_{l,j}])$  characterises the extra variance caused by the estimation error of  $f_\epsilon$  in  $\hat{T}_n^{\text{dom}}$ . Since  $\hat{T}_n^{\text{dom}}$  dominates  $R_n$  in  $\hat{T}_n$ ,  $\frac{1}{n} \sum_{j=1}^n \sum_{l=1}^2 (\zeta_{l,j} - E[\zeta_{l,j}])$  is the dominant component of  $\hat{T}_n - T_n$ , and it is non-negligible compared to  $T_n$  for the asymptotics of  $\tilde{\beta}(w^*)$ .

**B.1. Proof of Theorem 3 (i).** Observe that

$$|\tilde{\beta}(w^*) - \beta(w^*)|^2 \leq 3\lambda_{\max}(\hat{S}_n^{-2})(|\hat{T}_n^{\text{dom}}|^2 + |R_n|^2 + |B_n|^2).$$

Also note that  $\lambda_{\max}(\hat{S}_n^{-2}) = \{\lambda_{\min}(\hat{S}_n)\}^{-2}$  and  $\lambda_{\min}(\hat{S}_n) \geq \inf_{|\delta|=1} \delta'(\hat{S}_n - S_n)\delta + \lambda_{\min}(S_n)$ . By Assumption M and (A.7), if

$$\hat{S}_n - S_n \xrightarrow{p} 0, \tag{B.2}$$

we have

$$|\tilde{\beta}(w^*) - \beta(w^*)|^2 = O_p(|\hat{T}_n^{\text{dom}}|^2 + |R_n|^2 + a_n^{2p}). \tag{B.3}$$

For (B.2), note that

$$\begin{aligned}
|\hat{S}_n^{k_1, k_2} - S_n^{k_1, k_2}| &= \left| \frac{1}{na_n} \sum_{j=1}^n X_{k_1, j} X_{k_2, j} \left\{ \hat{\mathbb{K}} \left( \frac{w^* - W_j}{a_n} \right) - \mathbb{K} \left( \frac{w^* - W_j}{a_n} \right) \right\} \right| \\
&= O_p \left( a_n^{-1} E \left| X_{k_1} X_{k_2} \left\{ \hat{\mathbb{K}} \left( \frac{w^* - W}{a_n} \right) - \mathbb{K} \left( \frac{w^* - W}{a_n} \right) \right\} \right| \right) \\
&\stackrel{(1)}{=} O_p \left( a_n^{-1} \int \left| \hat{\mathbb{K}} \left( \frac{w^* - u}{a_n} \right) - \mathbb{K} \left( \frac{w^* - u}{a_n} \right) \right| f_W(u) du \right) \\
&\stackrel{(2)}{=} O_p \left( \int |\hat{\mathbb{K}}(\check{u}) - \mathbb{K}(\check{u})| f_W(w^* - a_n \check{u}) d\check{u} \right) \stackrel{(3)}{=} O_p \left( n^{-1/2} a_n^{-(3\alpha_w + 3\alpha_\epsilon + 2)} \log a_n^{-1} \right),
\end{aligned}$$

for  $k_1, k_2 = 1, \dots, k$ , where (1) follows by  $E[|X_{k_1} X_{k_2}| |W^*] < \infty$  implied by Assumption M, (2) follows by the change of variable  $\check{u} = \frac{w^* - u}{a_n}$ , and (3) follows by Lemma 2. Then (B.2) follows by Assumption OS' (2).

For  $|\hat{T}_n^{\text{dom}}|^2$ , since  $|\hat{T}_n^{\text{dom}}|^2 = O\left(n^{-1} \sum_{l=1}^3 E|\zeta_{l,1}|^2\right)$ , it suffices to check the magnitude of  $E|\zeta_{l,1}^{k_1}|^2$  for  $l = 1, 2, 3$  and  $k_1 = 1, \dots, k$ . For  $E|\zeta_{2,1}^{k_1}|^2$ , we have

$$\begin{aligned}
E|\zeta_{2,1}^{k_2}| &= E \left| \frac{iW^r}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \frac{\mu_3^{k_1}(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \int_0^t \frac{e^{isW}}{\mu_1(s)} ds dt \right|^2 \\
&\leq \frac{E|W^r|^2}{(2\pi)^2} \left( \int |K^{\text{ft}}(ta_n)| \left| \frac{\mu_3^{k_1}(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \right| |t| \sup_{|s| \leq |t|} \frac{1}{|\mu_1(s)|} dt \right)^2 \\
&= O \left( a_n^{-4} \left( \sup_{|t| \leq a_n^{-1}} \left| \frac{\mu_3^{k_1}(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \right| \right)^2 \left( \sup_{|t| \leq a_n^{-1}} \frac{1}{|\mu_1(s)|} \right)^2 \right) \\
&= O(a_n^{-(2\alpha_w + 2\alpha_\epsilon + 4)}), \tag{B.4}
\end{aligned}$$

where the last equality follows by

$$\begin{aligned}
\sup_{|t| \leq a_n^{-1}} \left| \frac{\mu_3^{k_1}(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \right| &= \sup_{|t| \leq a_n^{-1}} \left| \left\{ E[XX' \{\beta(W^*) - \beta(w^*)\} |W^*]^{k_1} f_{W^*} \right\}^{\text{ft}}(t) \right| = O(1), \\
\sup_{|t| \leq a_n^{-1}} \frac{1}{|\mu_1(s)|} &= O \left( \frac{1}{\inf_{|t| \leq a_n^{-1}} |f_{W^*}^{\text{ft}}(t)| \inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|} \right) = O(a_n^{-(\alpha_w + \alpha_\epsilon)}).
\end{aligned}$$

By similar arguments, we have  $E|\zeta_{1,1}^{k_1}|^2 = O(a_n^{-(4\alpha_w + 2\alpha_\epsilon + 4)})$  and  $E|\zeta_{3,1}^{k_1}|^2 = O(a_n^{-(2\alpha_\epsilon + 2)})$ , which together with (B.4) imply

$$|\hat{T}_n^{\text{dom}}|^2 = O_p(n^{-1} a_n^{-(4\alpha_w + 2\alpha_\epsilon + 4)}). \tag{B.5}$$

For  $|R_n|^2$ , decompose  $R_n = \sum_{l=1}^{14} R_{n,l}$ , where

$$\begin{aligned}
R_{n,1} &= -\frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) \frac{\delta_3(t)\delta_1(t)}{\mu_1(t)[\mu_1(t) + \delta_1(t)]} dt, \\
R_{n,2} &= \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) \frac{\mu_3(t)\delta_1^2(t)}{\mu_1^2(t)[\mu_1(t) + \delta_1(t)]} dt, \\
R_{n,3} &= -\frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) \int_0^t \frac{i\delta_2(s)\delta_1(s)}{\mu_1(s)[\mu_1(s) + \delta_1(s)]} ds \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,4} &= \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) \int_0^t \frac{i\mu_2(s)\delta_1^2(s)}{\mu_1^2(s)[\mu_1(s) + \delta_1(s)]} ds \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,5} &= \frac{1}{4\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left(\int_0^t \frac{i\delta_2(s)}{\mu_1(s)} ds\right)^2 \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,6} &= \frac{1}{4\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left(\int_0^t \frac{i\mu_2(s)\delta_1(s)}{\mu_1^2(s)} ds\right)^2 \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,7} &= \frac{1}{4\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left(\int_0^t \frac{i\delta_2(s)\delta_1(s)}{\mu_1(s)[\mu_1(s) + \delta_1(s)]} ds\right)^2 \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,8} &= \frac{1}{4\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left(\int_0^t \frac{i\mu_2(s)\delta_1^2(s)}{\mu_1^2(s)[\mu_1(s) + \delta_1(s)]} ds\right)^2 \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,9} &= -\frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left\{ \begin{array}{l} \int_0^t \frac{i\delta_2(s)}{\mu_1(s)} ds \\ \times \int_0^t \frac{i\mu_2(s)\delta_1(s)}{\mu_1^2(s)} ds \end{array} \right\} \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,10} &= -\frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left\{ \begin{array}{l} \int_0^t \frac{i\delta_2(s)}{\mu_1(s)} ds \\ \times \int_0^t \frac{i\delta_2(s)\delta_1(s)}{\mu_1(s)[\mu_1(s) + \delta_1(s)]} ds \end{array} \right\} \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,11} &= \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left\{ \begin{array}{l} \int_0^t \frac{i\delta_2(s)}{\mu_1(s)} ds \\ \times \int_0^t \frac{i\mu_2(s)\delta_1^2(s)}{\mu_1^2(s)[\mu_1(s) + \delta_1(s)]} ds \end{array} \right\} \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,12} &= \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left\{ \begin{array}{l} \int_0^t \frac{i\mu_2(s)\delta_1(s)}{\mu_1^2(s)} ds \\ \times \int_0^t \frac{i\delta_2(s)\delta_1(s)}{\mu_1(s)[\mu_1(s) + \delta_1(s)]} ds \end{array} \right\} \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,13} &= -\frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left\{ \begin{array}{l} \int_0^t \frac{i\mu_2(s)\delta_1(s)}{\mu_1^2(s)} ds \\ \times \int_0^t \frac{i\mu_2(s)\delta_1^2(s)}{\mu_1^2(s)[\mu_1(s) + \delta_1(s)]} ds \end{array} \right\} \frac{\mu_3(t)}{\mu_1(t)} dt, \\
R_{n,14} &= -\frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left\{ \begin{array}{l} \int_0^t \frac{i\delta_2(s)\delta_1(s)}{\mu_1(s)[\mu_1(s) + \delta_1(s)]} ds \\ \times \int_0^t \frac{i\mu_2(s)\delta_1^2(s)}{\mu_1^2(s)[\mu_1(s) + \delta_1(s)]} ds \end{array} \right\} \frac{\mu_3(t)}{\mu_1(t)} dt,
\end{aligned}$$

for some  $|\phi_2(t)| \leq \left| \int_0^t i \left\{ \frac{\hat{\mu}_2(s)}{\hat{\mu}_1(s)} - \frac{\mu_2(s)}{\mu_1(s)} \right\} ds \right|$ . By similar arguments as used in Kurisu and Otsu (2019, Lemma 1), for  $l = 1, 2, 3$ , we have

$$\sup_{|t| \leq a_n^{-1}} |\delta_l(t)| = O_p(n^{-1/2} \log a_n^{-1}). \quad (\text{B.6})$$

For  $R_{n,1}$ , we write

$$\begin{aligned}
|R_{n,1}^{k_1}| &= \left| \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right) \frac{\delta_3^{k_1}(t)\delta_1(t)}{\mu_1(t)\{\mu_1(t) + \delta_1(t)\}} dt \right| \\
&= \left| \frac{1}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \frac{\delta_3^{k_1}(t)\delta_1(t)}{f_{W^*}^{\text{ft}}(t)\{f_\epsilon^{\text{ft}}(t)\}^2 + \delta_1(t)f_\epsilon^{\text{ft}}(t)} dt \right| \\
&\stackrel{(1)}{=} O_p\left(\frac{a_n^{-1} \sup_{|t|\leq a_n^{-1}} |\delta_3^{k_1}(t)| \sup_{|t|\leq a_n^{-1}} |\delta_1(t)|}{\left(\inf_{|t|\leq a_n^{-1}} |f_{W^*}^{\text{ft}}(t)| \left(\inf_{|t|\leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|\right)\right)^2}\right) \stackrel{(2)}{=} O_p\left(n^{-1} a_n^{-(\alpha_w+2\alpha_\epsilon+1)} (\log a_n^{-1})^2\right),
\end{aligned}$$

where (1) follows by Assumption OS' (2) and (2) follows by Assumption OS (1), Assumption OS' (1), and (B.6), which implies  $|R_{n,1}| = O_p\left(n^{-1} a_n^{-(\alpha_w+2\alpha_\epsilon+1)} (\log a_n^{-1})^2\right)$ . By similar arguments, we have  $|R_{n,2}| = O\left(n^{-1} a_n^{-(2\alpha_w+2\alpha_\epsilon+1)} (\log a_n^{-1})^2\right)$ ,  $|R_{n,3}| = O\left(n^{-1} a_n^{-(2\alpha_w+2\alpha_\epsilon+2)} (\log a_n^{-1})^2\right)$ , and  $|R_{n,4}| = O\left(n^{-1} a_n^{-(3\alpha_w+2\alpha_\epsilon+2)} (\log a_n^{-1})^2\right)$ .

For  $R_{n,5}$ , we write

$$\begin{aligned}
|R_{n,5}^{k_1}| &= \left| \frac{1}{4\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds + \phi_2(t)\right) \left(\int_0^t \frac{i\delta_2(s)}{\mu_1(s)} ds\right)^2 \frac{\mu_3^{k_1}(t)}{\mu_1(t)} dt \right| \\
&= \left| \frac{1}{4\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) e^{\phi_2(t)} \left(\int_0^t \frac{i\delta_2(s)}{\mu_1(s)} ds\right)^2 \left\{E[\lambda' X X' \{\beta(W^*) - \beta(w^*)\} | W^*]^{k_1} f_{W^*}\right\}^{\text{ft}}(t) dt \right| \\
&= O_p\left(\frac{a_n^{-3} \sup_{|t|\leq a_n^{-1}} e^{|\phi_2(t)|} \left(\sup_{|t|\leq a_n^{-1}} |\delta_2(t)|\right)^2}{\left(\inf_{|t|\leq a_n^{-1}} |f_{W^*}^{\text{ft}}(t)| \inf_{|t|\leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|\right)^2}\right) = O_p\left(n^{-1} a_n^{-(2\alpha_w+2\alpha_\epsilon+3)} (\log a_n^{-1})^2\right),
\end{aligned}$$

where the last equality follows by Assumption OS (1), Assumption OS' (1), (B.6), and

$$\sup_{|t|\leq a_n^{-1}} e^{|\phi_2(t)|} \leq \exp\left(\sup_{|t|\leq a_n^{-1}} \left|\int_0^t i \left\{\frac{\hat{\mu}_2(s)}{\hat{\mu}_1(s)} - \frac{\mu_2(s)}{\mu_1(s)}\right\} ds\right|\right) = O_p\left(\exp\left(n^{-1/2} a_n^{-(2\alpha_w+\alpha_\epsilon+1)} \log a_n^{-1}\right)\right) = O_p(1).$$

This implies  $|R_{n,5}| = O_p\left(n^{-1} a_n^{-(2\alpha_w+2\alpha_\epsilon+3)} (\log a_n^{-1})^2\right)$ . By similar arguments, we have

$$\begin{aligned}
|R_{n,6}| &= O\left(n^{-1} a_n^{-(2\alpha_w+\alpha_\epsilon+3)} (\log a_n^{-1})^2\right), & |R_{n,7}| &= O\left(n^{-2} a_n^{-(4\alpha_w+4\alpha_\epsilon+3)} (\log a_n^{-1})^4\right), \\
|R_{n,8}| &= O\left(n^{-2} a_n^{-(6\alpha_w+4\alpha_\epsilon+3)} (\log a_n^{-1})^4\right), & |R_{n,9}| &= O\left(n^{-1} a_n^{-(3\alpha_w+2\alpha_\epsilon+3)} (\log a_n^{-1})^2\right), \\
|R_{n,10}| &= O\left(n^{-3/2} a_n^{-(3\alpha_w+3\alpha_\epsilon+3)} (\log a_n^{-1})^3\right), & |R_{n,11}| &= O\left(n^{-3/2} a_n^{-(4\alpha_w+3\alpha_\epsilon+3)} (\log a_n^{-1})^3\right), \\
|R_{n,12}| &= O\left(n^{-3/2} a_n^{-(4\alpha_w+3\alpha_\epsilon+3)} (\log a_n^{-1})^3\right), & |R_{n,13}| &= O\left(n^{-3/2} a_n^{-(5\alpha_w+3\alpha_\epsilon+3)} (\log a_n^{-1})^3\right), \\
|R_{n,14}| &= O\left(n^{-2} a_n^{-(5\alpha_w+4\alpha_\epsilon+3)} (\log a_n^{-1})^4\right).
\end{aligned}$$

Since  $\alpha_w > 1$  as given in Assumption OS' (1), Assumption OS' (2) implies

$$|R_n| = O\left(n^{-1} a_n^{-(3\alpha_w+2\alpha_\epsilon+3)} (\log a_n^{-1})^2\right). \tag{B.7}$$

The conclusion then follows by (B.5), (B.7), and (B.3).

**B.2. Proof of Theorem 3 (ii).** First, for each  $\lambda \in \mathbb{R}^k$  with  $|\lambda| = 1$ , we show

$$\frac{\lambda' \hat{T}_n^{\text{dom}}}{\sqrt{\text{Var}(\lambda' \hat{T}_n^{\text{dom}})}} \xrightarrow{d} N(0, 1). \quad (\text{B.8})$$

For (B.8), it suffices to check the Lyapunov condition for some  $\varsigma > 0$ , i.e.,

$$\frac{E \left| \sum_{l=1}^3 \lambda' \zeta_{l,1} - E[\sum_{l=1}^3 \lambda' \zeta_{l,1}] \right|^{2+\varsigma}}{n^{\varsigma/2} \{ \text{Var}(\sum_{l=1}^3 \lambda' \zeta_{l,1}) \}^{1+\varsigma/2}} \rightarrow 0. \quad (\text{B.9})$$

For the numerator of (B.9), it suffices to check the upper bound of each  $E|\lambda' \zeta_{l,1}|^{2+\varsigma}$  for  $l = 1, 2, 3$ .

Note that

$$\begin{aligned} E|\lambda' \zeta_{2,1}|^{2+\varsigma} &= E \left| \frac{iW^r}{2\pi} \int e^{-itw^*} K^{\text{ft}}(ta_n) \frac{\lambda' \mu_3(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \int_0^t \frac{e^{isW}}{\mu_1(s)} ds dt \right|^{2+\varsigma} \\ &\leq \frac{E|W^r|^{2+\varsigma}}{(2\pi)^{2+\varsigma}} \left( \int |K^{\text{ft}}(ta_n)| \left| \frac{\lambda' \mu_3(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \right| |t| \sup_{|s| \leq |t|} \frac{1}{|\mu_1(s)|} dt \right)^{2+\varsigma} \\ &= O \left( a_n^{-2(2+\varsigma)} \left( \sup_{|t| \leq a_n^{-1}} \left| \frac{\lambda' \mu_3(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \right| \right)^{2+\varsigma} \left( \sup_{|t| \leq a_n^{-1}} \frac{1}{|\mu_1(s)|} \right)^{2+\varsigma} \right) \\ &= O \left( a_n^{-(\alpha_w + \alpha_\epsilon + 2)(2+\varsigma)} \right), \end{aligned} \quad (\text{B.10})$$

where the last equality follows by

$$\begin{aligned} \sup_{|t| \leq a_n^{-1}} \left| \frac{\lambda' \mu_3(t) \exp\left(\int_0^t \frac{i\mu_2(u)}{\mu_1(u)} du\right)}{\mu_1(t)} \right| &= \sup_{|t| \leq a_n^{-1}} \left| \{E[\lambda' X X' \{\beta(W^*) - \beta(w^*)\} | W^*] f_{W^*}\}^{\text{ft}}(t) \right| = O(1), \\ \sup_{|t| \leq a_n^{-1}} \frac{1}{|\mu_1(s)|} &= O \left( \frac{1}{\inf_{|t| \leq a_n^{-1}} |f_{W^*}^{\text{ft}}(t)| \inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|} \right) = O(a_n^{-(\alpha_w + \alpha_\epsilon)}). \end{aligned}$$

By similar arguments, we have  $E|\lambda' \zeta_{1,1}|^{2+\varsigma} = O(a_n^{-(2\alpha_w + \alpha_\epsilon + 2)(2+\varsigma)})$  and  $E|\lambda' \zeta_{3,1}|^{2+\varsigma} = O(a_n^{-(\alpha_\epsilon + 1)(2+\varsigma)})$ , which together with (B.10) imply

$$E \left| \sum_{l=1}^3 \lambda' \zeta_{n,l,1} - E[\sum_{l=1}^3 \lambda' \zeta_{n,l,1}] \right|^{2+\varsigma} = O \left( a_n^{-(2\alpha_w + \alpha_\epsilon + 2)(2+\varsigma)} \right). \quad (\text{B.11})$$

For the denominator of (B.9), by Assumption M' (3), it suffices to check the lower bound of each  $E|\lambda' \zeta_{l,1}|^2$  for  $l = 1, 2, 3$ . Define

$$\begin{aligned} m_{n,1}^\lambda(t) &= \frac{i\lambda' A_n(t, w^*) \mu_2(t)}{\mu_1^2(t)} + e^{-itw^*} K^{\text{ft}}(ta_n) \frac{\lambda' \mu_3(t) \exp\left(\int_0^t \frac{i\mu_2(s)}{\mu_1(s)} ds\right)}{\mu_1^2(t)}, \\ m_{n,2}^\lambda(t) &= \frac{\lambda' A_n(t, w^*)}{\mu_1(t)}, \quad m_{n,3}^\lambda(t) = \frac{K^{\text{ft}}(ta_n)}{f_\epsilon^{\text{ft}}(t)}. \end{aligned} \quad (\text{B.12})$$

For  $E|\lambda'\zeta_{2,1}|^2$ , first note that

$$\begin{aligned}
E|\lambda'\zeta_{2,1}|^2 &= \frac{1}{4\pi^2} \int \left| \int \frac{\lambda' A_n(t, w^*)}{\mu_1(t)} e^{itu} dt \right|^2 E[W^{r^2} | W = u] f_W(u) du \\
&\geq \frac{1}{4\pi^2} \int_{u \in I_2} \left| \int \frac{\lambda' A_n(t, w^*)}{\mu_1(t)} e^{itu} dt \right|^2 E[W^{r^2} | W = u] f_W(u) du \\
&\geq \frac{c_2}{4\pi^2} \int_{u \in I_2} \left| \int \frac{\lambda' A_n(t, w^*)}{\mu_1(t)} e^{itu} dt \right|^2 du,
\end{aligned} \tag{B.13}$$

for any finite interval  $I_2$  not reduced to a point and some positive constant  $c_2$  which depends on  $I_2$ , where the last inequality follows by Assumption M' (1). Also note that there exists a positive constant  $c_2^{\text{kot}}$  such that

$$\left| \int \frac{\lambda' A(t, w^*)}{\mu_1(t)} e^{itu} dt \right|^2 \leq c_2^{\text{kot}} |u|^{-2},$$

for  $u \in \mathbb{R} \setminus I_2$ . This is an immediate result of

$$\lim_{|u| \rightarrow \infty} u \int \frac{\lambda' A(t, w^*)}{\mu_1(t)} e^{itu} dt = 0,$$

which follows by Assumption M' (2), Schennach (2004, Lemma 10), and

$$\int \frac{d}{dt} \left( \frac{\lambda' A(t, w^*)}{\mu_1(t)} \right) e^{itu} dt = -iu \int \frac{\lambda' A(t, w^*)}{\mu_1(t)} e^{itu} dt.$$

Then, for all  $n$  large enough, we have

$$\int_{u \in \mathbb{R} \setminus I_2} \left| \int \frac{\lambda' A_n(t, w^*)}{\mu_1(t)} e^{itu} dt \right|^2 du \leq c_2^{\text{kot}} \int_{u \in \mathbb{R} \setminus I_2} |u|^{-2} du < \infty, \tag{B.14}$$

which together with (B.13) imply that for all  $n$  large enough, there exists a positive constant  $C_2^{\text{kot}}$  such that

$$E|\lambda'\zeta_{2,1}|^2 \geq C_2^{\text{kot}} \int \left| \int \frac{\lambda' A_n(t, w^*)}{\mu_1(t)} e^{itu} dt \right|^2 du = 2\pi C_2^{\text{kot}} \int |m_{n,2}^\lambda(t)|^2 dt, \tag{B.15}$$

where the equality follows by the Parseval's identity. By similar arguments, for some positive constants  $C_1^{\text{kot}}$  and  $C_3^{\text{kot}}$ , we have  $E|\lambda'\zeta_{1,1}|^2 \geq 2\pi C_1^{\text{kot}} \int |m_{n,1}^\lambda(t)|^2 dt$  and  $E|\lambda'\zeta_{3,1}|^2 \geq 2\pi C_3^{\text{kot}} \int |m_{n,3}^\lambda(t)|^2 dt$ , which together with (B.15) and Assumption M' (3) imply that for some positive constant  $C^{\text{kot}}$ , we have

$$\text{Var} \left( \sum_{l=1}^3 \lambda' \zeta_{l,1} \right) \geq C^{\text{kot}} \max_{l=1,2,3} \int |m_{n,l}^\lambda(t)|^2 dt. \tag{B.16}$$

Thus, by Assumption OS' (3), (B.9) holds true, which implies

$$\{\text{Var}(\hat{T}_n^{\text{dom}})\}^{-1/2} \hat{T}_n^{\text{dom}} \xrightarrow{d} N(0, I_k).$$

Also note that by (B.7), (A.7), and (B.16), the higher order terms in  $\hat{T}_n$  and the bias terms are asymptotically negligible, i.e.

$$\frac{\lambda'(R_n + B_n)}{\sqrt{\text{Var}(\lambda' \hat{T}_n^{\text{dom}})}} = \frac{n^{1/2} \lambda'(R_n + B_n)}{\sqrt{\text{Var} \left( \sum_{l=1}^3 \lambda' \zeta_{l,1} \right)}} = o_p(1), \tag{B.17}$$

which follow by Assumption OS' (3). Thus, the conclusion follows by (B.8) and (B.17).

**B.3. Proof of Theorem 4 (i).** We know from the proof of Theorem 2, if

$$\hat{S}_n - S_n \xrightarrow{p} 0, \quad (\text{B.18})$$

then for  $1/3 \leq \gamma_\epsilon \leq 2$ ,

$$|\tilde{\beta}(w^*) - \beta(w^*)|^2 = O_p(|\hat{T}_n^{\text{dom}}|^2 + |R_n|^2 + a_n^{2p}). \quad (\text{B.19})$$

For (B.18), using the same arguments as in the proof of Theorem 3 (i), we have

$$\begin{aligned} |\hat{S}_n^{k_1, k_2} - S_n^{k_1, k_2}| &= O_p \left( \int |\hat{\mathbb{K}}(\tilde{u}) - \mathbb{K}(\tilde{u})| f_W(w^* - a_n \tilde{u}) d\tilde{u} \right) \\ &= O_p \left( n^{-1/2} a_n^{-2} (\log a_n^{-1}) e^{3\mu_w a_n^{-\gamma_w} + 3\mu_\epsilon a_n^{-\gamma_\epsilon}} \right), \end{aligned}$$

where the final equality follows by Lemma 3. Then (B.18) follows by Assumption SS' (4).

For  $|\hat{T}_n^{\text{dom}}|^2$ , since  $|\hat{T}_n^{\text{dom}}|^2 = O \left( n^{-1} \sum_{l=1}^3 E|\zeta_{l,1}|^2 \right)$ , we again check the magnitude of  $E|\zeta_{l,1}^{k_1}|^2$  for  $l = 1, 2, 3$  and  $k_1 = 1, \dots, k$ . For  $E|\zeta_{2,1}^{k_1}|^2$ , we have

$$E|\zeta_{2,1}^{k_2}|^2 \stackrel{(1)}{=} O \left( a_n^{-4} \left( \sup_{|t| \leq a_n^{-1}} \frac{1}{|\mu_1(s)|} \right)^2 \right) \stackrel{(2)}{=} O \left( a_n^{-4} e^{2\mu_w a_n^{-\gamma_w} + 2\mu_\epsilon a_n^{-\gamma_\epsilon}} \right), \quad (\text{B.20})$$

where (1) is taken directly from the proof of Theorem 3 (i), and (2) follows from

$$\sup_{|t| \leq a_n^{-1}} \frac{1}{|\mu_1(s)|} = O \left( \frac{1}{\inf_{|t| \leq a_n^{-1}} |f_{W^*}^{\text{ft}}(t)| \inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|} \right) = O \left( e^{\mu_w a_n^{-\gamma_w} + \mu_\epsilon a_n^{-\gamma_\epsilon}} \right),$$

where the first equality is also taken from the proof of Theorem 3 (i).

By similar arguments, we have  $E|\zeta_{1,1}^{k_1}| = O \left( a_n^{-4} e^{4\mu_w a_n^{-\gamma_w} + 2\mu_\epsilon a_n^{-\gamma_\epsilon}} \right)$  and  $E|\zeta_{3,1}^{k_1}|^2 = O \left( a_n^{-2} e^{2\mu_\epsilon a_n^{-\gamma_\epsilon}} \right)$ , which, together with (B.20), imply

$$|\hat{T}_n^{\text{dom}}|^2 = O_p \left( n^{-1} a_n^{-4} e^{4\mu_w a_n^{-\gamma_w} + 2\mu_\epsilon a_n^{-\gamma_\epsilon}} \right). \quad (\text{B.21})$$

For  $|R_n|^2$ , we again decompose this as  $R_n = \sum_{l=1}^{14} R_{n,l}$ , where  $R_{n,l}$  for  $l = 1, \dots, 14$  are defined in the proof of Theorem 3 (i). By similar arguments as used in Kurisu and Otsu (2019, Lemma 1), for  $l = 1, 2, 3$ , we have

$$\sup_{|t| \leq a_n^{-1}} |\delta_l(t)| = O_p(n^{-1/2} \log a_n^{-1}). \quad (\text{B.22})$$

For  $R_{n,1}$ , we write

$$\begin{aligned} |R_{n,1}^{k_1}| &\stackrel{(1)}{=} O_p \left( \frac{a_n^{-1} \sup_{|t| \leq a_n^{-1}} |\delta_3^{k_1}(t)| \sup_{|t| \leq a_n^{-1}} |\delta_1(t)|}{\inf_{|t| \leq a_n^{-1}} |f_{W^*}^{\text{ft}}(t)| \left( \inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)| \right)^2} \right) \\ &\stackrel{(2)}{=} O_p \left( n^{-1} a_n^{-1} (\log a_n^{-1})^2 e^{\mu_w a_n^{-\gamma_w} + 2\mu_\epsilon a_n^{-\gamma_\epsilon}} \right), \end{aligned}$$

where (1) follows directly from the proof of Theorem 3 (i), and (2) follows from Lemma 3 and (B.22), which implies

$$|R_{n,1}| = O_p \left( n^{-1} a_n^{-1} (\log a_n^{-1})^2 e^{\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right).$$

By similar arguments, we have  $|R_{n,2}| = O \left( n^{-1} a_n^{-1} (\log a_n^{-1})^2 e^{2\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right)$ ,

$$|R_{n,3}| = O \left( n^{-1} a_n^{-2} (\log a_n^{-1})^2 e^{2\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right), |R_{n,4}| = O \left( n^{-1} a_n^{-2} (\log a_n^{-1})^2 e^{3\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right).$$

For  $R_{n,5}$ , we can write

$$\begin{aligned} |R_{n,5}^{k_1}| &= O_p \left( \frac{a_n^{-3} \sup_{|t| \leq a_n^{-1}} e^{|\phi_2(t)|} \left( \sup_{|t| \leq a_n^{-1}} |\delta_2(t)| \right)^2}{\left( \inf_{|t| \leq a_n^{-1}} |f_{W^*}^{\text{ft}}(t)| \inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)| \right)^2} \right) \\ &= O_p \left( n^{-1} a_n^{-3} (\log a_n^{-1})^2 e^{2\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right), \end{aligned}$$

where the last equality follows by Lemma 3, (B.22), and the fact that  $\sup_{|t| \leq a_n^{-1}} e^{|\phi_2(t)|} = O_p(1)$  as shown in the proof of Theorem 3 (i). This implies  $|R_{n,5}| = O_p \left( n^{-1} a_n^{-3} (\log a_n^{-1})^2 e^{2\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right)$ .

By similar arguments, we have

$$\begin{aligned} |R_{n,6}| &= O \left( n^{-1} a_n^{-3} (\log a_n^{-1})^2 e^{2\mu_w a_n^{-\gamma w} + \mu_\epsilon a_n^{-\gamma \epsilon}} \right), & |R_{n,7}| &= O \left( n^{-2} a_n^{-3} (\log a_n^{-1})^4 e^{4\mu_w a_n^{-\gamma w} + 4\mu_\epsilon a_n^{-\gamma \epsilon}} \right), \\ |R_{n,8}| &= O \left( n^{-2} a_n^{-3} (\log a_n^{-1})^4 e^{6\mu_w a_n^{-\gamma w} + 4\mu_\epsilon a_n^{-\gamma \epsilon}} \right), & |R_{n,9}| &= O \left( n^{-1} a_n^{-3} (\log a_n^{-1})^2 e^{3\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right), \\ |R_{n,10}| &= O \left( n^{-3/2} a_n^{-3} (\log a_n^{-1})^3 e^{3\mu_w a_n^{-\gamma w} + 3\mu_\epsilon a_n^{-\gamma \epsilon}} \right), & |R_{n,11}| &= O \left( n^{-3/2} a_n^{-3} (\log a_n^{-1})^3 e^{4\mu_w a_n^{-\gamma w} + 3\mu_\epsilon a_n^{-\gamma \epsilon}} \right), \\ |R_{n,12}| &= O \left( n^{-3/2} a_n^{-3} (\log a_n^{-1})^3 e^{4\mu_w a_n^{-\gamma w} + 3\mu_\epsilon a_n^{-\gamma \epsilon}} \right), & |R_{n,13}| &= O \left( n^{-3/2} a_n^{-3} (\log a_n^{-1})^3 e^{5\mu_w a_n^{-\gamma w} + 3\mu_\epsilon a_n^{-\gamma \epsilon}} \right), \\ |R_{n,14}| &= O \left( n^{-2} a_n^{-3} (\log a_n^{-1})^4 e^{5\mu_w a_n^{-\gamma w} + 4\mu_\epsilon a_n^{-\gamma \epsilon}} \right). \end{aligned}$$

Combining these results, we obtain

$$|R_n| = O \left( n^{-1} a_n^{-3} (\log a_n^{-1})^2 e^{3\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}} \right). \quad (\text{B.23})$$

The conclusion then follows by (B.19), (B.21), and (B.23).

**B.4. Proof of Theorem 4 (ii).** First, for each  $\lambda \in \mathbb{R}^k$  with  $|\lambda| = 1$ , we show

$$\frac{\lambda' \hat{T}_n^{\text{dom}}}{\sqrt{\text{Var}(\lambda' \hat{T}_n^{\text{dom}})}} \xrightarrow{d} N(0, 1). \quad (\text{B.24})$$

For (B.24), it suffices to check the Lyapunov condition for some  $\varsigma > 0$ , i.e.,

$$\frac{E \left| \sum_{l=1}^3 \lambda' \zeta_{l,1} - E[\sum_{l=1}^3 \lambda' \zeta_{l,1}] \right|^{2+\varsigma}}{n^{\varsigma/2} \left\{ \text{Var} \left( \sum_{l=1}^3 \lambda' \zeta_{l,1} \right) \right\}^{1+\varsigma/2}} \rightarrow 0. \quad (\text{B.25})$$



For the numerator of (B.25), it suffices to check the upper bound of each  $E|\lambda'\zeta_{l,1}|^{2+\varsigma}$  for  $l = 1, 2, 3$ . For  $E|\lambda'\zeta_{2,j}|^{2+\varsigma}$ , we have

$$\begin{aligned} E|\lambda'\zeta_{2,1}|^{2+\varsigma} & \stackrel{(1)}{=} O\left(a_n^{-2(2+\varsigma)}\left(\sup_{|t|\leq a_n^{-1}}\frac{1}{|\mu_1(s)|}\right)^{2+\varsigma}\right) \\ & = O\left(a_n^{-2(2+\varsigma)}e^{(2+\varsigma)(\mu_w a_n^{-\gamma w} + \mu_\epsilon a_n^{-\gamma\epsilon})}\right), \end{aligned} \quad (\text{B.26})$$

where (1) follows directly from the proof of Theorem 3 (ii), and (2) uses

$$\sup_{|t|\leq a_n^{-1}}\frac{1}{|\mu_1(s)|} = O\left(e^{\mu_w a_n^{-\gamma w} + \mu_\epsilon a_n^{-\gamma\epsilon}}\right),$$

which was shown previously in the proof of Theorem 4 (i).

By similar arguments, we have

$$E|\lambda'\zeta_{1,1}|^{2+\varsigma} = O\left(a_n^{-2(2+\varsigma)}e^{(2+\varsigma)(2\mu_w a_n^{-\gamma w} + \mu_\epsilon a_n^{-\gamma\epsilon})}\right), \quad E|\lambda'\zeta_{3,1}|^{2+\varsigma} = O\left(a_n^{-(2+\varsigma)}e^{(2+\varsigma)\mu_\epsilon a_n^{-\gamma\epsilon}}\right).$$

Together with (B.26), these imply

$$E\left|\sum_{l=1}^3\lambda'\zeta_{l,1} - E\left[\sum_{l=1}^3\lambda'\zeta_{l,1}\right]\right|^{2+\varsigma} = O\left(a_n^{-2(2+\varsigma)}e^{(2+\varsigma)(2\mu_w a_n^{-\gamma w} + \mu_\epsilon a_n^{-\gamma\epsilon})}\right). \quad (\text{B.27})$$

For the denominator of (B.25), by Assumption SS' (3), it suffices to check the lower bound of each  $E|\lambda'\zeta_{l,1}|^2$  for  $l = 1, 2, 3$ . As shown in the proof of Theorem 3 (ii),  $E|\lambda'\zeta_{2,1}|^2$  can be bounded from below as

$$E|\lambda'\zeta_{2,1}|^2 \geq 2\pi C_2 \int |m_{n,2}^\lambda(t)|^2 dt, \quad (\text{B.28})$$

for some positive constant  $C_2$ . Equally, using similar arguments, for some positive constants  $C_1$  and  $C_3$ , we have

$$E|\lambda'\zeta_{1,1}|^2 \geq 2\pi C_1 \int |m_{n,1}^\lambda(t)|^2 dt, \quad E|\lambda'\zeta_{3,1}|^2 \geq 2\pi C_3 \int |m_{n,3}^\lambda(t)|^2 dt.$$

Together with Assumption SS' (3), these imply that for some positive constant  $C$ , we have

$$\text{Var}\left(\sum_{l=1}^3\lambda'\zeta_{l,1}\right) \geq C \max_{l=1,2,3} \int |m_{n,l}^\lambda(t)|^2 dt. \quad (\text{B.29})$$

Then, by Assumption SS' (4), (B.25) holds true, which implies

$$\text{Var}(\hat{T}_n^{\text{dom}})^{-1/2}\hat{T}_n^{\text{dom}} \xrightarrow{d} N(0, I_k).$$

Also note that by (B.23) and (B.29), the higher order terms in  $\hat{T}_n$  and the bias terms are asymptotically negligible, i.e.,

$$\frac{\lambda'(R_n + B_n)}{\sqrt{\text{Var}(\lambda'\hat{T}_n^{\text{dom}})}} = \frac{n^{1/2}\lambda'(R_n + B_n)}{\sqrt{\text{Var}\left(\sum_{l=1}^3\lambda'\zeta_{l,1}\right)}} = o_p(1), \quad (\text{B.30})$$

which follow by Assumption SS' (4). Thus, the conclusion follows by (B.24) and (B.30).

APPENDIX C. LEMMAS

**Lemma 1.** *Under Assumptions M and OS (1)-(2), it holds  $\int |\mathbb{K}(x)|^2 dx = O(a_n^{-2\alpha_\epsilon})$ .*

*Proof.* Note that

$$\int |\mathbb{K}(x)|^2 dx = O\left(\int \left|\frac{K^{\text{ft}}(t)}{f_\epsilon^{\text{ft}}(t/a_n)}\right|^2 dt\right) = O\left(\sup_{|t|\leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|^{-2}\right),$$

where the first equality follows by Parseval's identity and the second equality follows by Assumption OS (2). Then, the conclusion follows by Assumption OS (1).  $\square$

**Lemma 2.** *Under Assumptions M, R, OS (1)-(3), and OS' (1)-(2), it holds*

$$\sup_x |\hat{\mathbb{K}}(x) - \mathbb{K}(x)| = O\left(n^{-1/2} a_n^{-(3\alpha_w + 3\alpha_\epsilon + 2)} \log a_n^{-1}\right).$$

*Proof.* By the change of variables combined with  $|e^{-itx}| = 1$  and boundedness of  $K^{\text{ft}}$  (Assumption OS (2)), we have

$$|\hat{\mathbb{K}}(x) - \mathbb{K}(x)| = \left| \frac{1}{2\pi} \int e^{-itx} \left\{ \frac{1}{\hat{f}_\epsilon^{\text{ft}}(t/a_n)} - \frac{1}{f_\epsilon^{\text{ft}}(t/a_n)} \right\} K^{\text{ft}}(t) dt \right| = O\left(a_n^{-1} \sup_{|t|\leq a_n^{-1}} \left| \frac{1}{\hat{f}_\epsilon^{\text{ft}}(t)} - \frac{1}{f_\epsilon^{\text{ft}}(t)} \right|\right),$$

By the definition of  $\hat{f}_\epsilon^{\text{ft}}$ , note that

$$\begin{aligned} & \sup_{|t|\leq a_n^{-1}} \left| \frac{1}{\hat{f}_\epsilon^{\text{ft}}(t)} - \frac{1}{f_\epsilon^{\text{ft}}(t)} \right| = \sup_{|t|\leq a_n^{-1}} \left| \frac{\hat{f}_{W^*}^{\text{ft}}(t) f_W^{\text{ft}}(t) - f_{W^*}^{\text{ft}}(t) \hat{f}_W^{\text{ft}}(t)}{\hat{f}_W^{\text{ft}}(t) f_W^{\text{ft}}(t)} \right| \\ & \leq \sup_{|t|\leq a_n^{-1}} \frac{|\hat{f}_{W^*}^{\text{ft}}(t) - f_{W^*}^{\text{ft}}(t)|}{|\hat{f}_W^{\text{ft}}(t)|} + \sup_{|t|\leq a_n^{-1}} \frac{|\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)|}{|\hat{f}_W^{\text{ft}}(t)| |f_\epsilon^{\text{ft}}(t)|} \\ & \leq \frac{\sup_{|t|\leq a_n^{-1}} |\hat{f}_{W^*}^{\text{ft}}(t) - f_{W^*}^{\text{ft}}(t)|}{\inf_{|t|\leq a_n^{-1}} |f_W^{\text{ft}}(t)| - \sup_{|t|\leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)|} + \frac{\sup_{|t|\leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)| \left(\inf_{|t|\leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|\right)^{-1}}{\inf_{|t|\leq a_n^{-1}} |f_W^{\text{ft}}(t)| - \sup_{|t|\leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)|}, \end{aligned}$$

where the first inequality follows from the triangle inequality. Thus, the conclusion follows by Assumption OS' (2) and

$$\sup_{|t|\leq a_n^{-1}} |\hat{f}_{W^*}^{\text{ft}}(t) - f_{W^*}^{\text{ft}}(t)| = O_p\left(n^{-1/2} a_n^{-(2\alpha_w + 2\alpha_\epsilon + 1)} \log a_n^{-1}\right), \quad (\text{C.1})$$

$$\sup_{|t|\leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)| = O_p(n^{-1/2} \log a_n^{-1}), \quad (\text{C.2})$$

$$\left(\inf_{|t|\leq a_n^{-1}} |f_W^{\text{ft}}(t)|\right)^{-1} = O(a_n^{-(\alpha_w + \alpha_\epsilon)}), \quad (\text{C.3})$$

$$\left(\inf_{|t|\leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|\right)^{-1} = O(a_n^{-\alpha_\epsilon}), \quad (\text{C.4})$$

where (C.1) follows by Kurisu and Otsu (2019, Theorem (i)), (C.2) follows by Kurisu and Otsu (2019, Lemma 1), (C.3) follows by Assumption OS (1) and Assumption OS' (1), and (C.4) follows by Assumption OS (1).  $\square$

**Lemma 3.** Under Assumption M, SS (1)-(2), and SS' (1)-(2), it holds that

$$\sup_x |\hat{\mathbb{K}}(x) - \mathbb{K}(x)| = O\left(n^{-1/2} a_n^{-2} (\log a_n^{-1}) e^{3\mu_w a_n^{-\gamma w} + 3\mu_\epsilon a_n^{-\gamma \epsilon}}\right).$$

*Proof.* From the proof of Lemma 2, we know

$$\begin{aligned} |\hat{\mathbb{K}}(x) - \mathbb{K}(x)| &= O\left(a_n^{-1} \frac{\sup_{|t| \leq a_n^{-1}} |\hat{f}_{W^*}^{\text{ft}}(t) - f_{W^*}^{\text{ft}}(t)|}{\inf_{|t| \leq a_n^{-1}} |f_W^{\text{ft}}(t)| - \sup_{|t| \leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)|}\right) \\ &\quad + O\left(a_n^{-1} \frac{\sup_{|t| \leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)| \left(\inf_{|t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|\right)^{-1}}{\inf_{|t| \leq a_n^{-1}} |f_W^{\text{ft}}(t)| - \sup_{|t| \leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)|}\right). \end{aligned}$$

Thus, the conclusion follows by Assumption SS' (2) and

$$\sup_{|t| \leq a_n^{-1}} |\hat{f}_{W^*}^{\text{ft}}(t) - f_{W^*}^{\text{ft}}(t)| = O_p\left(n^{-1/2} a_n^{-1} (\log a_n^{-1}) e^{2\mu_w a_n^{-\gamma w} + 2\mu_\epsilon a_n^{-\gamma \epsilon}}\right), \quad (\text{C.5})$$

$$\sup_{|t| \leq a_n^{-1}} |\hat{f}_W^{\text{ft}}(t) - f_W^{\text{ft}}(t)| = O_p(n^{-1/2} \log a_n^{-1}), \quad (\text{C.6})$$

$$\left(\inf_{t \in |t| \leq a_n^{-1}} |f_W^{\text{ft}}(t)|\right)^{-1} = O\left(e^{\mu_w a_n^{-\gamma w} + \mu_\epsilon a_n^{-\gamma \epsilon}}\right), \quad (\text{C.7})$$

$$\left(\inf_{t \in |t| \leq a_n^{-1}} |f_\epsilon^{\text{ft}}(t)|\right)^{-1} = O\left(e^{\mu_\epsilon a_n^{-\gamma \epsilon}}\right), \quad (\text{C.8})$$

where (C.5) follows by Kurisu and Otsu (2019, Theorem (ii)), (C.6) follows by Kurisu and Otsu (2019, Lemma 1), (C.7) follows by Assumption SS (1) and SS' (1), and (C.8) follows by Assumption SS (1).  $\square$

## REFERENCES

- [1] Adusumilli, K. and T. Otsu (2018) Nonparametric instrumental regression with errors in variables, *Econometric Theory*, 34, 1256-1280.
- [2] Bissantz, N., Dümbgen, L., Holzmann, H. and A. Munk (2007) Nonparametric confidence bands in deconvolution density estimation, *Journal of the Royal Statistical Society*, B 69, 483-506.
- [3] Bonhomme, S. and Robin, J. M. (2010) Generalized non-parametric deconvolution with an application to earnings dynamics, *Review of Economic Studies*, 77, 491-533.
- [4] Chung, K. L. (1974) *A Course in Probability Theory*, Academic Press, London.
- [5] Cleveland, W. S., Grosse, E. and W. M. Shyu (1991) Local regression models, in *Statistical Models in S*, pp. 309-376, Wadsworth & Brooks, Pacific Grove.
- [6] Comte, F. and J. Kappus (2015) Density deconvolution from repeated measurement without symmetry assumption on the errors, *Journal of Multivariate Analysis*, 140, 31-46.
- [7] Dong, H. and T. Otsu (2019) Nonparametric estimation of additive model with errors-in-variables, Working paper.
- [8] Fan, J. (1992) Deconvolution with supersmooth distributions, *Canadian Journal of Statistics*, 20, 155-169.
- [9] 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.
- [10] Hastie, T. and R. Tibshirani (1993) Varying-coefficient models, *Journal of the Royal Statistical Society*, B 55, 757-779.
- [11] 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.
- [12] 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.
- [13] Kotlarski, I. (1967) On characterizing the gamma and the normal distribution, *Pacific Journal of Mathematics*, 20, 69-76.
- [14] Kurisu, D. and T. Otsu (2019) On uniform convergence of deconvolution estimator from repeated measurements, Working paper.
- [15] Li, L. and T. Greene (2008) Varying coefficients model with measurement error, *Biometrics*, 64, 519-526.
- [16] Li, Q., Huang, C. J., Li, D. and T.-T. Fu (2002) Semiparametric smooth coefficient models, *Journal of Business & Economic Statistics*, 20, 412-422.
- [17] Ma, S. and P. X. K. Song (2015) Varying index coefficient models, *Journal of the American Statistical Association*, 110, 341-356.
- [18] 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.
- [19] Meister, A (2009) *Deconvolution Problems in Nonparametric Statistics*, Springer.
- [20] McMurry, T. L. and D. N. Politis (2004) Nonparametric regression with infinite order flat-top kernels, *Journal of Nonparametric Statistics*, 16, 549-562.
- [21] Otsu, T. and L. Taylor (2019) Specification testing for errors-in-variables models, Working paper.
- [22] Robinson, P. M. (1988) Root-N-consistent semiparametric regression, *Econometrica*, 931-954.
- [23] Schennach, S. M. (2004) Nonparametric regression in the presence of measurement error, *Econometric Theory*, 20, 1046-1093.
- [24] Schennach, S. M. (2019) Convolution without independence, *Journal of Econometrics*, 211, 308-318.
- [25] Uh, H. W. (2003) *Kernel Deconvolution*, PhD thesis, University of Amsterdam.
- [26] 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.

- [27] Yao, F., Zhang, F. and S. C. Kumbhakar (2019) Semiparametric smooth coefficient stochastic frontier model with panel data, forthcoming in *Journal of Business & Economic Statistics*.
- [28] 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.
- [29] 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`