Multivariate Regression Estimation with Errors-in-Variables: Asymptotic Normality for Mixing Processes*

Jianqing Fan†
Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

Elias Masry‡
Department of Electrical and
Computer Engineering
University of California at San Diego
La Jolla, CA 92093

July 14, 1991

Abstract

Errors-in-variables regression is the study of the association between covariates and responses where covariates are observed with errors. In this paper, we consider the estimation of multivariate regression functions for dependent data with errors in covariates. Nonparametric deconvolution technique is used to account for errors-in-variables. The asymptotic behavior of regression estimators depends on the smoothness of the error distributions, which are characterized as either ordinary smooth or super smooth. Asymptotic normality is established for both strongly mixing and ρ-mixing processes, when the error distribution function is either ordinary smooth or super smooth.

*‡ Supported by the Office of Naval Research under Grant N00014-90-J-1175.

† Completed while visiting the Department of Electrical and Computer Engineering, University of California at San Diego.

Abbreviated title. Errors-in-variables regression.

AMS 1991 Subject Classification. Primary 62G07. Secondary 62H10, 60F05.

Key words and phrases. Asymptotic normality, deconvolution, errors-in-variables, multivariate regression, mixing processes.

1 Introduction

In data analysis, it is customary to explore the association between covariates and responses via regression analysis. Let X° denote the covariate variable and Y be the response variable. The regression function is defined by $m(x) = E(Y|X^\circ = x)$ which is assumed to exist. This paper deals with the regression problem with errors-in-variables: We wish to estimate $m(x)$, but direct observations of the covariate X° are not available. Instead, due to the measuring mechanism or the nature of the environment, the covariate X° is measured with error ε: $X_j = X_j^\circ + \varepsilon_j$ so that X_j instead of X_j° is observed and one desires to explore the association between X° and Y based on the observation $(X_j, Y_j)_{j=1}^n$. This problem arises, for example, in medical and epidemiologic studies where risk factors are partially observed. See Prentice (1986) and Whittemore and Keller (1988).

In the i.i.d. case, the nonparametric errors-in-variables problem was studied by Fan and Truong (1990) and Fan, Truong and Wang (1990) where optimal rates of convergence and asymptotic normality are established. Let $\tilde{K}(\cdot)$ be a kernel function whose Fourier transform is given by

$$\tilde{\phi}_K(t) = \int_{-\infty}^{+\infty} \exp(itx)\tilde{K}(x)dx \tag{1.1}$$

and $\tilde{\phi}_\varepsilon(t)$ be the characteristic function of the error variable ε. Set

$$\tilde{W}_b(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-itx)\frac{\tilde{\phi}_K(t)}{\tilde{\phi}_\varepsilon(t/b)}dt, \tag{1.2}$$

which is a deconvolution kernel. Fan and Truong (1990) proposed the following kernel estimate for $m(x)$:

$$\hat{m}_n(x) = \frac{\sum_{j=1}^n Y_j \tilde{W}_{b_n}((x - X_j)/b_n)}{\sum_{j=1}^n \tilde{W}_{b_n}((x - X_j)/b_n)}, \tag{1.3}$$

where b_n is the bandwidth parameter. We remark that the deconvolution kernel $\tilde{W}_{b_n}$ is used to account for the fact that the covariates are observed with error. For more discussions on deconvolution, see Carroll and Hall (1988), Liu and Taylor (1989), Stefanski and Carroll (1990), Zhang (1990), and Fan (1991a, b, 1992) in the i.i.d. setting and Masry (1991a,b,c) for dependent observations.

Our goal in this paper is to establish the asymptotic normality for estimators of form (1.3) in the following more general setting.

- The processes $\{X_j^\circ\}$ and $\{Y_j\}$ are individually and jointly *dependent.*
- Multivariate regression from past *vector* data is considered.
- Estimation of general regression function of form

$$m(\boldsymbol{x};p) = m(x_1, \cdots, x_p; p) \equiv E\left(\psi(Y_p)\middle| X_1^\circ = x_1, \cdots, X_p^\circ = x_p\right) \tag{1.4}$$

  is studied, where $\psi(\cdot)$ is an arbitrary measurable function. These functions include the usual mean regression and conditional moment functions as well as conditional distribution functions.

We note that in this general setting, sharp almost sure convergence rates were established in Masry (1991d).

When the error variable $\varepsilon \equiv 0$, the errors-in-variables problem reduces to the ordinary nonparametric regression where covariates are observable. In that case, the deconvolution kernel (1.2) is just an ordinary kernel in which case the estimator (1.3) was proposed by Nadaraya (1964) and Watson (1964). The estimator (1.3) has been thoroughly studied with no errors in covariates. See, for example, Mack and Silverman (1982) and Härdle (1990) and references therein for i.i.d. observations, and Rosenblatt (1969), Robinson (1983, 1986), Collomb and Härdle (1986), Truong and Stone (1991), Truong (1991) and Roussas and Tran (1991), among others for dependent observations.

We now introduce the regression estimator in the more general setting mentioned above. Let $\{X_j^\circ\}_{j=-\infty}^{\infty}$ and $\{Y_j\}_{j=-\infty}^{\infty}$ be jointly stationary processes and $\{\varepsilon_j\}_{j=-\infty}^{\infty}$ be i.i.d. random variables, independent of the processes $\{X_j^\circ\}_{i=-\infty}^{\infty}$ and $\{Y_j\}_{j=-\infty}^{\infty}$. Denote the probability density and the characteristic function of the error variable ε by $\tilde{h}(x)$ and $\tilde{\phi}_\varepsilon(t)$, respectively. Set

$$X_j = X_j^\circ + \varepsilon_j, j = 0 \pm 1, \cdots.$$

Let $f^\circ(\boldsymbol{x};p) = f^\circ(x_1,\cdots,x_p;p)$ be the joint probability density function of the random variables $X_1^\circ,\cdots,X_p^\circ$, which is assumed to exist. Then the joint probability density function of $X_1,\cdots,X_p$ is given by

$$f(\boldsymbol{x};p) = \int_{R^p} f^\circ(\boldsymbol{x}-\boldsymbol{u};p)h(\boldsymbol{u})d\boldsymbol{u} \tag{1.5}$$

where

$$h(\boldsymbol{u}) = \prod_{j=1}^{p} \tilde{h}(u_j). \tag{1.6}$$

Let

$$\boldsymbol{X}_j^\circ = (X_{j+1}^\circ,\cdots,X_{j+p}^\circ), \qquad \boldsymbol{X}_j = (X_{j+1},\cdots,X_{j+p}). \tag{1.7}$$

For simplicity, we use product kernel for the multivariate nonparametric regression estimation. Let the kernel $\tilde{K}$ be a real-valued, even, and bounded density function on the real line satisfying $\tilde{K}(x) = O(|x|^{-1-\delta})$ for some $\delta > 0$ and let $\tilde{\phi}_K(t)$ be its Fourier transform. A basic assumption on the error distribution and the kernel function is that for every $b > 0$

$$\tilde{\phi}_\varepsilon(t) \neq 0, \forall t \in R; \qquad \tilde{\phi}_K(t)/\tilde{\phi}_\varepsilon(t/b) \in L_1 \cap L_\infty.$$

With $\tilde{W}_b$ defined by (1.2), set

$$K(\boldsymbol{x}) = \prod_{j=1}^{p} \tilde{K}(x_j); \qquad W_b(\boldsymbol{x}) = \prod_{j=1}^{p} \tilde{W}_b(x_j) \tag{1.8}$$

so that

$$\phi_K(\boldsymbol{t}) = \prod_{j=1}^{p} \tilde{\phi}_K(t_j). \tag{1.9}$$

Let $\{b_n\}$ be a sequence of positive numbers such that $b_n \to 0$ as $n \to \infty$. Given the observations $(X_j, Y_j)_{j=1}^n$, we estimate the regression function (1.4) by

$$\hat{m}_n(\boldsymbol{x};p) = \hat{R}_n(\boldsymbol{x};p)/\hat{f}_n^\circ(\boldsymbol{x};p), \tag{1.10}$$

where

$$\hat{R}_n(\boldsymbol{x};p) = \frac{1}{(n-p+1)b_n^p} \sum_{j=0}^{n-p} \psi(Y_{p+j}) W_{b_n}((\boldsymbol{x}-\boldsymbol{X}_j)/b_n) \tag{1.11}$$

and

$$\hat{f}_n^{\circ}(\boldsymbol{x};p) = \frac{1}{(n-p+1)b_n^p} \sum_{j=0}^{n-p} W_{b_n}\left((\boldsymbol{x} - \boldsymbol{X}_j)/b_n\right). \tag{1.12}$$

We remark that $\hat{f}_n^{\circ}(\boldsymbol{x};p)$ is a deconvolution density estimation of $f^{\circ}(\boldsymbol{x};p)$.

For considering the asymptotic normality of $\hat{m}(\boldsymbol{x};p)$, we define the centralizing parameter by

$$B_n(\boldsymbol{x};p) = \frac{E\hat{R}_n(\boldsymbol{x};p) - R(\boldsymbol{x};p) - m(\boldsymbol{x};p)(E\hat{f}_n^{\circ}(\boldsymbol{x};p) - f^{\circ}(\boldsymbol{x};p))}{E\hat{f}_n^{\circ}(\boldsymbol{x};p)}, \tag{1.13}$$

where $R(\boldsymbol{x};p) = m(\boldsymbol{x};p)f^{\circ}(\boldsymbol{x};p)$. We will see that $B_n(\boldsymbol{x};p)$ is the 'asymptotic bias' of the estimator $\hat{m}(\boldsymbol{x};p)$. With $\hat{R}_n$, $\hat{f}_n^{\circ}$ and B_n defined respectively by (1.11), (1.12) and (1.13), it is easy to verify that

$$\hat{m}_n(\boldsymbol{x};p) - m(\boldsymbol{x};p) - B_n(\boldsymbol{x};p) = \frac{\tilde{Q}_n(\boldsymbol{x};p) - B_n(\boldsymbol{x};p)(\hat{f}_n^{\circ}(\boldsymbol{x};p) - E\hat{f}_n^{\circ}(\boldsymbol{x};p))}{\hat{f}_n^{\circ}(\boldsymbol{x};p)},$$

where

$$\tilde{Q}_n(\boldsymbol{x};p) = \hat{R}_n(\boldsymbol{x};p) - E\hat{R}_n(\boldsymbol{x};p) - m(\boldsymbol{x};p)\left(\hat{f}_n^{\circ}(\boldsymbol{x};p) - E\hat{f}_n^{\circ}(\boldsymbol{x};p)\right).$$

It will be shown in Proposition 1.1 that $B_n(\boldsymbol{x};p) = o(1)$ under some mild conditions. Therefore, the dominated term in the numerator is $\tilde{Q}_n$:

$$\hat{m}_n(\boldsymbol{x};p) - m(\boldsymbol{x};p) - B_n(\boldsymbol{x};p) = \tilde{Q}_n(\boldsymbol{x};p)(1 + o_P(1))/\hat{f}_n^{\circ}(\boldsymbol{x};p). \tag{1.14}$$

Note that $\tilde{Q}_n$ is centralized and has the form of an average of a sequence of stationary random variables. Hence, we need first to establish asymptotic normality for $\tilde{Q}_n$, and then the asymptotic normality of $\hat{m}_n$ follows easily from (1.14).

The bias of the estimators $\hat{f}_n^{\circ}(\boldsymbol{x};p)$ and $\hat{R}_n(\boldsymbol{x};p)$ and the asymptotic value of $B_n(\boldsymbol{x};p)$ are given by the following proposition.

Proposition 1.1.

a) *For almost all $\boldsymbol{x} \in R^p$, we have as $n \to \infty$*

$$E\hat{f}_n^{\circ}(\boldsymbol{x};p) \to f^{\circ}(\boldsymbol{x};p); \quad E\hat{R}_n(\boldsymbol{x};p) \to R(\boldsymbol{x};p); \quad \textit{and} \quad B_n(\boldsymbol{x};p) \to 0,$$

where $R(\boldsymbol{x};p) = m(\boldsymbol{x};p)f^{\circ}(\boldsymbol{x};p)$.

b) *If $f^\circ(\boldsymbol{x};p)$ and $R(\boldsymbol{x};p)$ are twice differentiable and their second partial derivatives are bounded and continuous on R^p and the kernel function $\tilde{K}$ satisfies $\int_{-\infty}^{+\infty} u^2\tilde{K}(u)du < \infty$, then as $n \to \infty$*

1. $b_n^{-2}\text{bias}(\hat{f}_n^\circ(\boldsymbol{x};p)) \to \frac{1}{2}\int_{R^p} \boldsymbol{u}G''_{f^\circ}(\boldsymbol{x};p)\boldsymbol{u}^T K(\boldsymbol{u})d\boldsymbol{u}$;
2. $b_n^{-2}\text{bias}(\hat{R}_n(\boldsymbol{x};p)) \to \frac{1}{2}\int_{R^p} \boldsymbol{u}G''_{R}(\boldsymbol{x};p)\boldsymbol{u}^T K(\boldsymbol{u})d\boldsymbol{u}$;
3. $b_n^{-2}B_n(\boldsymbol{x};p) \to B(\boldsymbol{x};p)$;

where $\boldsymbol{u}^T$ is the transpose of the row vector $\boldsymbol{u}$, the $p \times p$ matrices G'' are given by

$$G''_{f^\circ}(\boldsymbol{x};p) = \left(\frac{\partial^2 f^\circ(\boldsymbol{x};p)}{\partial x_i \partial x_j}\right), \qquad G''_{R}(\boldsymbol{x};p) = \left(\frac{\partial^2 R(\boldsymbol{x};p)}{\partial x_i \partial x_j}\right),$$

and

$$B(\boldsymbol{x};p) = \int_{R^p} \boldsymbol{u}\left(G''_R(\boldsymbol{x};p) - m(\boldsymbol{x};p)G''_{f^\circ}(\boldsymbol{x};p)\right)\boldsymbol{u}^T K(\boldsymbol{u})d\boldsymbol{u}/(2f^\circ(\boldsymbol{x};p)).$$

Proof. See Fan and Truong (1990) and Masry (1991d).

We remark that the above bias expressions do not dependent on the error distribution. However, the asymptotic variance and the optimal rates of convergence depend strongly on the smoothness of the error distributions. Fan (1991a) shows that such a dependence is an intrinsic part of the regression problem, not an artifact produced by the kernel method being used. Following Fan (1991a), we call a distribution

- super smooth of order β: if the characteristic function of the error distribution $\tilde{\phi}_\varepsilon(\cdot)$ satisfies

$$a_0|t|^{\beta_0}\exp(-a|t|^\beta) \le |\tilde{\phi}_\varepsilon(t)| \le a_1|t|^{\beta_1}\exp(-a|t|^\beta) \quad \text{as } t \to \infty, \tag{1.15}$$

  where a, a_0, a_1, β are positive constants and β_0, β_1 are constants;

- ordinary smooth of order β: if the characteristic function of the error distribution $\tilde{\phi}_\varepsilon(\cdot)$ satisfies

$$d_0|t|^{-\beta} \le |\tilde{\phi}_\varepsilon(t)| \le d_1|t|^{-\beta} \quad \text{as } t \to \infty,$$

  for positive constants d_0, d_1, β.

Note that the above conditions are imposed in the Fourier domain and only on the tail of the characteristic function. The faster the decay of the tail of the characteristic function, the smoother its corresponding density. The super smooth distributions include normal and Cauchy distributions and their mixtures; the ordinary smooth distributions include gamma and Laplace distributions.

It will be seen in the sequel that the technical conditions needed to establish the asymptotic normality of $\hat{m}_n(\boldsymbol{x};p)$, as well as the nature of the proofs, depend strongly on the type of the error distributions. Section 2 establishes asymptotic normality when the error distribution is ordinary smooth and Section 3 accomplishes the same for super smooth error distributions. Both strongly mixing and ρ-mixing processes are studied. Some technical proofs are given in the Appendix.

2 Asymptotic normality for ordinary smooth error distributions

2.1 Preliminaries

Let $\mathcal{F}_i^k$ be the σ-algebra of events generated by the random variables $\{X_j^\circ, \varepsilon_j, Y_j, i \le j \le k\}$ and $L_2(\mathcal{F}_i^k)$ denote the collection of all second-order random variables which are $\mathcal{F}_i^k$-measurable. The stationary processes $\{X_j^\circ, \varepsilon_j, Y_j\}$ are called strongly mixing (Rosenblatt, 1956) if

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |P(AB) - P(A)P(B)| = \alpha(k) \to 0 \quad \text{as} \quad k \to \infty$$

and are said to be uniformly mixing if

$$\sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |P(B|A) - P(B)| = \phi(k) \to 0 \quad \text{as} \quad k \to \infty,$$

and are called ρ-mixing (Kolmogorov and Rozanov, 1960) if

$$\sup_{U \in L_2(\mathcal{F}_{-\infty}^0), V \in L_2(\mathcal{F}_k^\infty)} \frac{|\text{cov}(U,V)|}{\text{var}^{1/2}(U)\text{var}^{1/2}(V)} = \rho(k) \to 0 \quad \text{as} \quad k \to \infty.$$

It is well known that these mixing coefficients satisfy

$$\alpha(k) \leq \frac{1}{4}\rho(k) \leq \frac{1}{2}\phi^{1/2}(k),$$

and thus the class of ρ-mixing processes is intermediate between strongly and uniformly mixing.

We begin by imposing some conditions on the kernel function, the error distribution, as well as the mixing coefficients.

Condition 2.1.

$\tilde{\phi}_\varepsilon(t)$ and $\tilde{\phi}_K(t)$ are twice continuously differentiable with bounded derivatives such that

i) $\tilde{\phi}_\varepsilon(t) \neq 0, \quad \forall t \in R,$

ii) $t^\beta \tilde{\phi}_\varepsilon(t) \to B$ as $t \to \infty$ for some $\beta \geq 1$ and $B > 0$,

iii) $\delta_{\beta,1} \int_{-\infty}^{+\infty} |t|^{\beta-2} |\tilde{\phi}_K(t)| dt < \infty; \quad \int_{-\infty}^{+\infty} |t|^{2\beta} |\tilde{\phi}_K(t)|^2 dt < \infty,$

iv) $\int_{-\infty}^{+\infty} |t|^{\beta-1} |\tilde{\phi}'_K(t)| dt < \infty; \quad \int_{-\infty}^{+\infty} |t|^{\beta} |\tilde{\phi}''_K(t)| dt < \infty,$

where $\delta_{\beta,1}$ is the Kronecker's delta.

Let $f_{\boldsymbol{X}_0, \boldsymbol{X}_l | Y_p, Y_{p+l}}(\boldsymbol{u}, \boldsymbol{v} | y_p, y_{p+l})$ be the conditional density of $(\boldsymbol{X}_0, \boldsymbol{X}_l)$ given $Y_p = y_p$ and $Y_{p+l} = y_{p+l}$, where $\boldsymbol{X}_j$ is given by (1.7), and when $1 \leq l < p$, the vector $(\boldsymbol{X}_0, \boldsymbol{X}_l)$ means $(X_1, \cdots, X_{l+p})$. Recall that $f(\cdot)$ is the joint density of $\boldsymbol{X}_j$ given by (1.5). Let $f(\boldsymbol{x}_0, \boldsymbol{x}_l)$ be the probability density function of $(\boldsymbol{X}_0, \boldsymbol{X}_l)$ with a similar meaning as above when $1 \leq l < p$. Denote by

$$V(\boldsymbol{x}; p) = E\left[(\psi(Y_p) - m(\boldsymbol{x}; p))^2 \middle| \boldsymbol{X}_0 = \boldsymbol{x} \right]. \tag{2.1}$$

We make the following assumptions on the processes involved.

Condition 2.2.

i) $E\left[|\psi(Y_p)|^\nu \middle| \boldsymbol{X}_0 = \boldsymbol{x} \right] \leq A_1$ for some $\nu > 2$.

ii) $|f(\boldsymbol{x};p)| \leq A_2$.

iii) $f_{\boldsymbol{X}_0,\boldsymbol{X}_l|Y_p,Y_{p+l}}(\boldsymbol{u},\boldsymbol{v}|y_p,y_{p+l}) \leq A_3$, for all $l \geq 1$.

iv) Either the processes $\{X_j^\circ, \varepsilon_j, Y_j\}$ are ρ-mixing with $\sum_{j=1}^{\infty} \rho(j) < \infty$; or are strongly mixing with $\sum_{l=1}^{\infty} l^a [\alpha(l)]^{1-2/\nu} < \infty$ for some $a > 1 - 2/\nu$.

v) $f(\boldsymbol{x};q) \leq A_4$ for all $1 \leq q \leq 2p$, and $f(\boldsymbol{x}_0, \boldsymbol{x}_l) \leq A_4$ for all $l \geq p$.

where $A_j (j = 1, \cdots, 4)$ are some positive constants.

We remark that Condition 2.2 ii), iv) and v) are imposed on the X-variable. By the convolution theorem, they are satisfied when the density $\tilde{h}(\cdot)$ of error variable ε is bounded. With B and β given in Condition 2.1, let

$$D = \frac{1}{2\pi |B|^2} \int_{-\infty}^{+\infty} |t|^{2\beta} |\tilde{\phi}_K(t)|^2 dt. \qquad (2.2)$$

We need the following lemmas.

Lemma 2.1 (Masry, 1991a). *Under Condition 2.1 and Condition 2.2 iv) and v), we have*

$$\lim_{n\to\infty} n b_n^{(2\beta+1)p} \mathrm{var}\left(\hat{f}_n^\circ(\boldsymbol{x};p)\right) = D^p f(\boldsymbol{x};p)$$

at points of continuity of $f(\boldsymbol{x};p)$. If in addition $nb_n^{p(2\beta+1)} \to \infty$, then by Proposition 1.1

$$\hat{f}_n^\circ(\boldsymbol{x};p) \xrightarrow{P} f^\circ(\boldsymbol{x};p), \qquad \text{as } n \to \infty.$$

Lemma 2.2 (Masry, 1991a). *Under Condition 2.1, we have the following result:*

a) *There exists a constant c such that*

$$\|\tilde{W}_{b_n}\|_1 \leq c/b_n^\beta; \qquad \|\tilde{W}_{b_n}\|_l \leq c/b_n^\beta, \quad 2 \leq l \leq \infty.$$

b) *For all points $\boldsymbol{x}$ of continuity of a bounded function $g(\cdot)$,*

$$\lim_{b_n \to 0} \frac{1}{b_n^p} \int_{R^p} \prod_{j=1}^{p} \left(b_n^{2\beta} \tilde{W}_{b_n}^2 \left(\frac{x_j - u_j}{b_n} \right) \right) g(\boldsymbol{u}) d\boldsymbol{u} = D^p g(\boldsymbol{x}),$$

where $\tilde{W}_{b_n}$ is the deconvolution kernel given by (1.2) and D is given by (2.2).

To study the asymptotic normality for $\tilde{Q}_n$ and hence for $\hat{m}_n$, we put

$$\tilde{Z}_{n,j} = b_n^{-p} \left(\psi(Y_p) - m(\boldsymbol{x};p) \right) W_{b_n} \left((\boldsymbol{x} - \boldsymbol{X}_0)/b_n \right) - \mu_n, \tag{2.3}$$

where K and W_{b_n} are defined by (1.8) and

$$\mu_n = E \left[b_n^{-p} \left(\psi(Y_p) - m(\boldsymbol{x};p) \right) W_{b_n} \left((\boldsymbol{x} - \boldsymbol{X}_0)/b_n \right) \right].$$

By Proposition 1.1, μ_n is independent of the error distribution and goes to zero for almost all $\boldsymbol{x} \in R^p$:

$$\mu_n = E \left[b_n^{-p} \left(\psi(Y_p) - m(\boldsymbol{x};p) \right) K \left((\boldsymbol{x} - \boldsymbol{X}_0)/b_n \right) \right] = o(1). \tag{2.4}$$

Then, we have

$$\tilde{Q}_n = \frac{1}{n-p+1} \sum_{j=0}^{n-p} \tilde{Z}_{n,j}. \tag{2.5}$$

With $V(\boldsymbol{x};p)$ given by (2.1), let

$$\theta^2(\boldsymbol{x};p) = D^p V(\boldsymbol{x};p) f(\boldsymbol{x};p).$$

Lemma 2.3. *Under Conditions 2.1 and 2.2 i) – iii), we have at continuity points of $V(\boldsymbol{x};p)$ and $f(\boldsymbol{x};p)$ that*

$$\operatorname{var}(\tilde{Z}_{n,0}) = b_n^{-(2\beta+1)p} \theta^2(\boldsymbol{x};p)(1 + o(1)),$$

$$\sum_{l=1}^{n-1} \left| \operatorname{cov}(\tilde{Z}_{n,0}, \tilde{Z}_{n,l}) \right| = o\left(\operatorname{var}(\tilde{Z}_{n,0}) \right)$$

and

$$\operatorname{var}\left(\sum_{j=0}^{n-1} \tilde{Z}_{n,j} \right) = n b_n^{-(2\beta+1)p} \theta^2(\boldsymbol{x};p)(1 + o(1)),$$

where $\tilde{Z}_{n,j}$ is given by (2.3).

Proof . We first remark that the third result follows directly from the first two results together with the stationarity assumption:

$$\text{var}\left(\sum_{j=0}^{n-1} \tilde{Z}_{n,j}\right) = n\text{var}(\tilde{Z}_{n,0}) + 2n\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)\text{cov}(\tilde{Z}_{n,0},\tilde{Z}_{n,j}). \tag{2.6}$$

By conditioning on $\boldsymbol{X}_0$, we have from (2.3) and (2.4) that

$$\begin{aligned}
\text{var}(\tilde{Z}_{n,0}) &= b_n^{-2p} E\left([\psi(Y_p) - m(\boldsymbol{x};p)]^2 W_{b_n}^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}_0}{b_n}\right)\right) + o(1) \\
&= b_n^{-2p} E\left(V(\boldsymbol{X}_0;p) W_{b_n}^2\left(\frac{\boldsymbol{x}-\boldsymbol{X}_0}{b_n}\right)\right) + o(1) \\
&= b_n^{-(2\beta+1)p}\frac{1}{b_n^p}\int_{R^p}\prod_{j=1}^{p}\left(b_n^{2\beta}\tilde{W}_{b_n}^2\left(\frac{x_j-u_j}{b_n}\right)\right) V(\boldsymbol{u};p) f(\boldsymbol{u};p) d\boldsymbol{u} + o(1).
\end{aligned}$$

Applying Lemma 2.2 b) with $g(\boldsymbol{x}) = V(\boldsymbol{x};p)f(\boldsymbol{x};p)$ which is bounded by Condition 2.2 i) and ii), we obtain the first part of the result.

Next, with a sequence of integers $c_n \to \infty$ such that $c_n b_n^p \to 0$, we write

$$\sum_{l=1}^{n-1}\left|\text{cov}(\tilde{Z}_{n,0},\tilde{Z}_{n,l})\right| = \left(\sum_{l=1}^{p-1} + \sum_{l=p}^{c_n} + \sum_{l=c_n+1}^{n-1}\right)\left|\text{cov}(\tilde{Z}_{n,0},\tilde{Z}_{n,l})\right| \equiv J_1 + J_2 + J_3.$$

For $1 \le l \le p-1$, (2.3) and (2.4) lead to

$$\text{cov}(\tilde{Z}_{n,0},\tilde{Z}_{n,l}) = b_n^{-2p} E\left([\psi(Y_p)-m][\psi(Y_{p+l})-m] W_{b_n}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_0}{b_n}\right) W_{b_n}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_l}{b_n}\right)\right) + o(1).$$

Let $\boldsymbol{u}', \boldsymbol{u}'', \boldsymbol{u}'''$ denote respectively an $l, (p-l), l$ dimensional vector, where $\boldsymbol{u}''$ will represent the overlap part of the vector $\boldsymbol{X}_0$ and $\boldsymbol{X}_l$. Conditioning on (Y_p, Y_{p+l}) and using Condition 2.2 iii) and a change of variables, the above covariance is further bounded by

$$\begin{aligned}
&\left|b_n^{-2p} E\left\{[\psi(Y_p)-m][\psi(Y_{p+l})-m] E\left[W_{b_n}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_0}{b_n}\right) W_{b_n}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_l}{b_n}\right)\middle| Y_p, Y_{p+l}\right]\right\}\right| + o(1) \\
\le\ & \frac{A_3}{b_n^{2p}} E\Big|[\psi(Y_p)-m][\psi(Y_{p+l})-m]\Big| \left(b_n^{p+l}\int_{R^{p+l}} \left|W_{b_n}(\boldsymbol{u}',\boldsymbol{u}'') W_{b_n}(\boldsymbol{u}'',\boldsymbol{u}''')\right| d\boldsymbol{u}' d\boldsymbol{u}'' d\boldsymbol{u}'''\right) \\
=\ & \frac{A_3}{b_n^{p-l}} \|\tilde{W}_{b_n}\|_1^{2l} \|\tilde{W}_{b_n}\|_2^{2(p-l)} E\Big|[\psi(Y_p)-m][\psi(Y_{p+l})-m]\Big|,
\end{aligned}$$

uniformly in l. The last equality follows from the factorization (1.8). This together with Lemma 2.2 a) lead to

$$J_1 = \sum_{l=1}^{p-1} O\left(b_n^{-(2\beta+1)p+l}\right) = o\left(\operatorname{var}(\tilde{Z}_{n,0})\right).$$

For $p \le l \le c_n$, we have similarly that

$$\left|\operatorname{cov}(\tilde{Z}_{n,0}, \tilde{Z}_{n,l})\right| \le A_3 \|\tilde{W}_{b_n}\|_1^{2p} E\left|[\psi(Y_p) - m][\psi(Y_{p+l}) - m]\right| = O\left(b_n^{-2\beta p}\right),$$

uniformly in l. Therefore, by the choice of c_n, we have

$$J_2 = O\left(c_n b_n^{-(2\beta+1)p+p}\right) = o\left(b_n^{-(2\beta+1)p}\right) = o\left(\operatorname{var}(\tilde{Z}_{n,0})\right).$$

We now deal with J_3. We separate the argument into two cases, depending on whether the processes are ρ-mixing or strongly mixing. For ρ-mixing processes, we have

$$J_3 \le \sum_{l=c_n}^{n} \rho(l-p+1)\operatorname{var}(\tilde{Z}_{n,0}) = o\left(\operatorname{var}(\tilde{Z}_{n,0})\right),$$

by the summability of the mixing coefficients. For strongly mixing processes, we proceed as follows. By Davydov's Lemma [see Hall and Heyde (1980), Corollary A2], we have

$$\left|\operatorname{cov}(\tilde{Z}_{n,0}, \tilde{Z}_{n,l})\right| \le 8[\alpha(l-p+1)]^{1-2/\nu} \left(E|\tilde{Z}_{n,0}|^\nu\right)^{2/\nu}. \tag{2.7}$$

Conditioning on $\boldsymbol{X}_0$ and using Condition 2.2 i), we have

$$\begin{aligned} E|\tilde{Z}_{n,0}|^\nu &\le b_n^{-p\nu} E\left(\left|W_{b_n}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_0}{b_n}\right)\right|^\nu E\left[|\psi(Y_p)-m|^\nu|\boldsymbol{X}_0\right]\right) + o(1) \\ &\le C_1 b_n^{-p\nu} E\left|W_{b_n}\left(\frac{\boldsymbol{x}-\boldsymbol{X}_0}{b_n}\right)\right|^\nu, \end{aligned}$$

where C_1 is a positive constant. Therefore, by a change of variable and Condition 2.2 ii), we obtain

$$E|\tilde{Z}_{n,0}|^\nu \le C_1 A_2 b_n^{-p(\nu-1)} \|W_{b_n}\|_\nu^\nu = O\left(b_n^{-p(\nu-1+\beta\nu)}\right),$$

where the last equality follows from (1.8) and Lemma 2.1 a). This together with (2.7) entail

$$\begin{aligned} J_3 &\le \frac{C_2}{b_n^{2p(\beta+1-1/\nu)}} \sum_{l=c_n+1}^{\infty} [\alpha(l-p+1)]^{1-2/\nu} \\ &\le \frac{C_2}{c_n^a b_n^{2p(\beta+1-1/\nu)}} \sum_{l=c_n+1}^{\infty} l^a [\alpha(l-p+1)]^{1-2/\nu}, \end{aligned}$$

for some positive constant C_2. Choose $c_n = [b_n^{-p(1-2/\nu)/a}]$ so that $c_n b_n^p \to 0$ since $a > 1-2/\nu$. In view of Condition 2.2 iv), we have $J_3 = o(b_n^{-(2\beta+1)p})$. This completes the proof. □

Lemma 2.4 (Volkonskii and Rozanov, 1959). *Let $V_1, \cdots, V_L$ be random variables measurable with respect to the σ-algebras $\mathcal{F}_{i_1}^{j_1}, \cdots, \mathcal{F}_{i_L}^{j_L}$ respectively with $1 \le i_1 < j_1 < i_2 < \cdots < j_L \le n$, $i_{l+1} - j_l \ge w \ge 1$ and $|V_j| \le 1$ for $j = 1, \cdots, L$. Then*

$$\left| E\left(\prod_{j=1}^{L} V_j\right) - \prod_{j=1}^{L} E(V_j) \right| \le 16(L-1)\alpha(w),$$

where $\alpha(w)$ is the strongly mixing coefficient.

2.2 Main Results

The principle result of this section gives the asymptotic normality of the regression estimator (1.10) for both strongly mixing and ρ-mixing processes.

Condition 2.3.

Let $\{s_n\}$ be a sequence of positive integers, $s_n \to \infty$, such that $s_n = o((nb_n^p)^{1/2})$. For strongly mixing processes, $\alpha(k)$ satisfies $(nb_n^{-p})^{1/2}\alpha(s_n) \to 0$ as $n \to \infty$; for ρ-mixing processes, $\rho(k)$ satisfies $(nb_n^{-p})^{1/2}\rho(s_n) \to 0$ as $n \to \infty$.

Theorem 2.1. *Under Conditions 2.1 – 2.3 and $nb_n^{(2\beta+1)p} \to \infty$ as $n \to \infty$, we have*

$$\left(nb_n^{(2\beta+1)p}\right)^{1/2} [\hat{m}_n(\boldsymbol{x};p) - m(\boldsymbol{x};p) - B_n(\boldsymbol{x};p)] \xrightarrow{\mathcal{L}} N(0, \tau^2(\boldsymbol{x};p))$$

at continuity point of $f^\circ(\boldsymbol{x};p)$ and $V(\boldsymbol{x};p)$, where $\tau^2(\boldsymbol{x};p) = D^p V(\boldsymbol{x};p) f(\boldsymbol{x};p)/(f^\circ(\boldsymbol{x};p))^2$ and f, V and D are given respectively by (1.5), (2.1) and (2.2).

The following corollary follows from Theorem 2.1 and Part b) of Proposition 1.1 together with the choice of bandwidth.

Corollary 2.1 *Under Conditions 2.1– 2.3, if the functions $m(\boldsymbol{x};p)$ and $f(\boldsymbol{x};p)$ have* *bounded and continuous second partial derivatives on R^p, then*

$$\left(nb_n^{(2\beta+1)p}\right)^{1/2}[\hat{m}_n(\boldsymbol{x};p)-m(\boldsymbol{x};p)] \xrightarrow{\mathcal{L}} N(0,\tau^2(\boldsymbol{x};p)),$$

provided that $nb_n^{4+(2\beta+1)p} \to 0$ and that $\int_{-\infty}^{+\infty} u^2\tilde{K}(u)du < \infty$.

Proof of Theorem 2.1

The idea of the proof of Theorem 2.1 is as follows. We first establish the asymptotic normality for $\tilde{Q}_n$ and then use (1.14) to conclude the desired result. We employ the following big blocks and small blocks argument.

Set

$$Z_{n,j} = b_n^{(2\beta+1)p/2}\tilde{Z}_{n,j}; \qquad S_n = \sum_{j=0}^{n-1} Z_{n,j}.$$

Then, (2.5) leads to

$$\left(nb_n^{(2\beta+1)p}\right)^{1/2}\tilde{Q}_n = \sqrt{n/(n-p+1)}\frac{1}{\sqrt{n-p+1}}S_{n-p+1}. \tag{2.8}$$

In view of Lemma 2.3, we have

$$\mathrm{var}(Z_{n,0}) \to \theta^2(\boldsymbol{x};p); \quad \sum_{j=1}^{n-1}|\mathrm{cov}(Z_{n,0},Z_{n,j})| \to 0. \tag{2.9}$$

Partition the set $\{1,2,\cdots,n\}$ into $2k_n+1$ subsets with large blocks of size $r=r_n$ and small blocks of size $s=s_n$, where, with $[\cdot]$ denoting integer part,

$$k = k_n = [\frac{n}{r_n+s_n}]. \tag{2.10}$$

Define the random variables

$$\eta_j = \sum_{i=j(r+s)}^{j(r+s)+r-1} Z_{n,i}, \quad 0 \le j \le k-1, \tag{2.11}$$

$$\xi_j = \sum_{i=j(r+s)+r}^{(j+1)(r+s)-1} Z_{n,i}, \quad 0 \le j \le k-1, \tag{2.12}$$

and

$$\zeta_k = \sum_{i=k(r+s)}^{n-1} Z_{n,i}. \tag{2.13}$$

Then,

$$S_n = \sum_{j=0}^{k-1} \eta_j + \sum_{j=0}^{k-1} \xi_j + \eta_k \equiv S_n' + S_n'' + S_n'''. \tag{2.14}$$

We will show that as $n \to \infty$,

$$\frac{1}{n} E(S_n'')^2 \to 0, \quad \frac{1}{n} E(S_n''')^2 \to 0 \tag{2.15}$$

$$\left| E\left[\exp(itS_n')\right] - \prod_{j=0}^{k-1} E\left[\exp(it\eta_j)\right] \right| \to 0 \tag{2.16}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E\left(\eta_j^2\right) \to \theta^2(\boldsymbol{x}; p) \tag{2.17}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E\left(\eta_j^2 I\{|\eta_j| \geq \varepsilon\theta(\boldsymbol{x}; p)\sqrt{n}\}\right) \to 0 \tag{2.18}$$

for every ε. (2.15) implies that S_n'' and S_n''' are asymptotically negligible, (2.16) implies that the summands $\{\eta_j\}$ in S_n' are asymptotically independent, and (2.17) and (2.18) are the standard Lindeberg-Feller conditions for asymptotic normality of S_n' under independence. Expressions (2.15) – (2.18) entail the following asymptotic normality:

$$\frac{1}{\sqrt{n}} S_n \xrightarrow{\mathcal{L}} N\left(0, \theta^2(\boldsymbol{x}; p)\right) \tag{2.19}$$

so that by (2.8)

$$\left(nb_n^{(2\beta+1)p}\right)^{1/2} \tilde{Q}_n(\boldsymbol{x}; p) \xrightarrow{\mathcal{L}} N(0, \theta^2(\boldsymbol{x}; p)).$$

This together with (1.14) and Lemma 2.1 prove Theorem 2.1.

We now establish (2.15) – (2.18). The proof concentrates on the strongly mixing case as it is more involved. We remark on the differences for ρ-mixing processes.

We first choose the block sizes. Condition 2.3 implies that there exist constants $q_n \to \infty$ such that for strongly mixing processes,

$$q_n s_n = o\left((nb_n^p)^{1/2}\right); \quad q_n(n/b_n^p)^{1/2}\alpha(s_n) \to 0;$$

and for ρ-mixing processes

$$q_n s_n = o\left((nb_n^p)^{1/2}\right), \qquad q_n(nb_n^{-p})^{1/2}\rho(s_n) \to 0.$$

Define the large block size $r_n = [(nb_n^p)^{1/2}/q_n]$. Then, simple algebra shows the following properties:

$$s_n/r_n \to 0, \quad r_n/n \to 0, \quad r_n/(nb_n^p)^{1/2} \to 0, \tag{2.20}$$

and

$$\frac{n}{r_n}\alpha(s_n) \to 0. \tag{2.21}$$

[For ρ-mixing processes, (2.21) is proved via the inequality $\alpha(s_n) \le \rho(s_n)/4$.]

We now establish (2.15). Note that

$$E(S_n'')^2 = \sum_{j=0}^{k-1} \mathrm{var}(\xi_j) + \sum_{i=0}^{k-1} \sum_{\substack{j=0 \\ i\neq j}}^{k-1} \mathrm{cov}(\xi_i, \xi_j) \equiv F_1 + F_2. \tag{2.22}$$

Using stationarity and (2.9), we obtain that

$$\begin{aligned} \mathrm{var}(\xi_j) &= s\mathrm{var}(Z_{n,0}) + 2s\sum_{l=1}^{s-1}(1 - l/s)\mathrm{cov}(Z_{n,0}, Z_{n,l}) \\ &= s\mathrm{var}(Z_{n,0}) + O\left(s\sum_{l=1}^{s-1}|\mathrm{cov}(Z_{n,0}, Z_{n,l})|\right) \\ &= s\theta^2(\boldsymbol{x};p)(1 + o(1)). \end{aligned} \tag{2.23}$$

By (2.22),

$$F_1 = O(ks) = o(n),$$

since by (2.10) and (2.20)

$$r_n k_n/n = s_n/(r_n + s_n) \to 0.$$

Now, we consider F_2. We first note that with $m_j = j(r + s) + r$,

$$F_2 = \sum_{i=0}^{k-1} \sum_{\substack{j=0 \\ i\neq j}}^{k-1} \sum_{l_1=0}^{s-1}\sum_{l_2=0}^{s-1} \mathrm{cov}(Z_{n,m_i+l_1}, Z_{n,m_j+l_2}),$$

but since $i \neq j$, $|m_i - m_j + l_1 - l_2| \ge r$ so that

$$|F_2| \le 2\sum_{l_1=0}^{n-r-1} \sum_{l_2=l_1+r}^{n-1} |\mathrm{cov}(Z_{n,l_1}, Z_{n,l_2})|.$$

By stationarity and (2.9)

$$|F_2| \le 2n \sum_{j=r}^{n-1} |\mathrm{cov}(Z_{n,0}, Z_{n,j})| = o(n).$$

By (2.22), we have validated the first part of (2.15). For the second part of (2.15), using a similar argument together with (2.9), we obtain that

$$\begin{aligned}\frac{1}{n}E(S_n''')^2 &\le \frac{1}{n}(n-k(r+s))\mathrm{var}(Z_{n,0}) + 2\sum_{j=1}^{n-1}|\mathrm{cov}(Z_{n,0}, Z_{n,j})| \\ &\le \frac{r_n+s_n}{n}\theta(\boldsymbol{x};p) + o(1) \to 0.\end{aligned}$$

For (2.16) we proceed as follows. We note that η_a is a function of the random variables

$$\{X_{a(r+s)+1}, \cdots, X_{p+a(r+s)+r-1}, Y_{p+a(r+s)}, \cdots, Y_{p+a(r+s)+r-1}\}$$

or η_a is $\mathcal{F}_{i_a}^{j_a}$-measurable with $i_a = a(r+s)+1$ and $j_a = p+a(r+s)+r-1$. Hence, applying Lemma 2.4 with $V_j = \exp(it\eta_j)$, we have

$$\left| E\left(\prod_{j=0}^{k-1}\exp(it\eta_j)\right) - \prod_{j=0}^{k-1}E[\exp(it\eta_j)]\right| \le 16k\alpha(s_n+2-p) \sim 16\frac{n}{r_n}\alpha(s_n+2-p),$$

which tends to zero by (2.21).

We now show (2.17). By stationarity and (2.23) with s replaced by r, we have

$$\mathrm{var}(\eta_j) = \mathrm{var}(\eta_0) = r\theta^2(\boldsymbol{x};p)(1+o(1)).$$

This implies that

$$\frac{1}{n}\sum_{j=0}^{k-1}E(\eta_j^2) = \frac{k_n r_n}{n}\theta^2(\boldsymbol{x};p)(1+o(1)) \sim \frac{r_n}{r_n+s_n}\theta^2(\boldsymbol{x};p) \to \theta^2(\boldsymbol{x};p),$$

since $s_n/r_n \to 0$.

It remains to establish (2.18). We first prove (2.18) when $\psi(\cdot)$ is bounded. This would establish the asymptotic normality (2.19) for this particular case. The general case of ψ possibly unbounded is then established by using a truncation argument.

Assume that $|\psi(\cdot)| \le L$. Then by the definition of $Z_{n,j}$,

$$|Z_{n,j}| \le b_n^{(2\beta-1)p/2}\left\{(L+|m(\boldsymbol{x};p)|)\|W_{b_n}\|_\infty + |\mu_n|\right\},$$

and by Lemma 2.2 a) and the fact that $\mu_n \to 0$, we have

$$|Z_{n,j}| \leq C_3/b_n^{p/2},$$

for some constant C_3. This and (2.11) entail

$$\max_{0\leq j\leq k-1} |\eta_j|/\sqrt{n} \leq C_3 r_n/(nb_n^p)^{1/2}$$

which tends to by (2.20). Hence, when n is large the set $\{|\eta_j| \geq \theta(\boldsymbol{x};p)\varepsilon\sqrt{n}\}$ becomes an empty set. Hence, (2.18) follows, and consequently the asymptotic normality (2.19) holds for bounded $\psi(\cdot)$.

To complete the proof for the general case, we utilize the following truncation argument: Put

$$\psi_L = \psi(y)I\{|\psi(y)| \leq L\},$$

where L is a fixed truncation point. Correspondingly let

$$m_L(\boldsymbol{x};p) = E\left(\psi_L(Y_p)\Big|\boldsymbol{X}_0 = \boldsymbol{x}\right),$$

and

$$V_L(\boldsymbol{x};p) = E\left[(\psi_L(Y_p) - m_L(\boldsymbol{x};p))^2\Big|\boldsymbol{X}_0 = \boldsymbol{x}\right], \qquad \theta_L^2 = D^p V_L(\boldsymbol{x};p)f(\boldsymbol{x};p).$$

Put

$$\mu_{n,L} = E\left[b_n^{-p}\left(\psi_L(Y_p) - m_L(\boldsymbol{x};p)\right)W_{b_n}\left((\boldsymbol{x}-\boldsymbol{X}_0)/b_n\right)\right],$$

$$Z_{n,j}^L = b_n^{(2\beta+1)p/2}\Big[b_n^{-p}\left(\psi_L(Y_p) - m_L(\boldsymbol{x};p)\right)W_{b_n}\left((\boldsymbol{x}-\boldsymbol{X}_0)/b_n\right) - \mu_{n,L}\Big]$$

and

$$S_n^L = \sum_{j=0}^{n-1} Z_{n,j}^L, \qquad \tilde{S}_n^L = \sum_{j=0}^{n-1}(Z_{n,j} - Z_{n,j}^L). \tag{2.24}$$

Then, by the asymptotic normality for bounded $\psi(\cdot)$, we have

$$\frac{1}{\sqrt{n}}S_n^L \xrightarrow{\mathcal{L}} N(0,\theta_L^2). \tag{2.25}$$

In order to complete the proof, namely to establish (2.19) for the general case, it suffices to show that as first $n \to \infty$ and then $L \to \infty$ we have

$$\frac{1}{n}\text{var}\left(\tilde{S}_n^L\right) \longrightarrow 0. \tag{2.26}$$

Indeed,

$$\begin{aligned}
&\left|E\exp(itS_n/\sqrt{n}) - \exp(-t^2\theta(\boldsymbol{x};p)/2)\right| \\
=\ &\left|E\exp(it(S_n^L + \tilde{S}_n^L)/\sqrt{n}) - \exp(-t^2\theta_L^2/2) + \exp(-t^2\theta_L^2/2) - \exp(-t^2\theta^2/2)\right| \\
\le\ &\left|E\exp(itS_n^L/\sqrt{n}) - \exp(-t^2\theta_L^2/2)\right| + E\left|\exp(it\tilde{S}_n^L/\sqrt{n}) - 1\right| \\
&\quad + \left|\exp(-t^2\theta_L^2/2) - \exp(-t^2\theta^2/2)\right|
\end{aligned}$$

Letting $n \to \infty$, the first term goes to zero by (2.25) for every $L > 0$; the second term converges to zero by (2.26) as first $n \to \infty$ and then $L \to \infty$; the third term goes to zero as $L \to \infty$ by the dominated convergence theorem. Therefore, it remains to prove (2.26). Note that by (2.24) $\tilde{S}_n^L$ has the same structure as S_n^L except that the function ψ_L is replaced by $\psi - \psi_L$. Hence, by Lemma 2.3 (note the different scaling between $\tilde{Z}_{n,j}$ and $Z_{n,j}$), we have

$$\lim_{n\to\infty} \text{var}\left(\tilde{S}_n^L\right)/n = D^p f(\boldsymbol{x};p) E\left[(\psi(Y_p)I\{|\psi(Y_p)| > L\} - [m(\boldsymbol{x};p) - m_L(\boldsymbol{x};p)])^2 \middle| X_0 = \boldsymbol{x}\right].$$

By the dominated convergence theorem, the right hand side converges to 0 as $L \to \infty$. This establishes (2.26) and completes the proof of Theorem 1. □

3 Asymptotic normality for super smooth error distributions

In this section, we deal with super smooth error distributions, whose characteristic function decays exponentially fast. Recall that $\tilde{Q}_n(\boldsymbol{x};p)$ given by (2.5) is central to our discussion of asymptotic normality. Set

$$\sigma_0^2(n) = \text{var}(\tilde{Z}_{n,0}) = \text{var}\Big(b_n^{-p}[\psi(Y_p) - m(\boldsymbol{x};p)]W_{b_n}([\boldsymbol{x} - X_0]/b_n)\Big) \tag{3.1}$$

and

$$\sigma^2(n) = \mathrm{var}(\tilde{Q}_n(\boldsymbol{x}; p)). \tag{3.2}$$

Since the asymptotic rates and constants of $\sigma_0^2(n)$ and $\sigma^2(n)$ are not available, the technical arguments are more involved here than in the ordinary smooth case. We first derive both lower and upper bound for $\sigma_0^2(n)$ and then use these bounds to establish

$$\sigma^2(n) = \frac{1}{n}\sigma_0^2(n)(1 + o(1)).$$

These bounds are also useful in validating the Lindeberg-Feller condition for asymptotic normality.

3.1 Preliminaries

We make the following assumptions on the characteristic function $\tilde{\phi}_\varepsilon(t)$ of the error variable ε and on the Fourier transform $\tilde{\phi}_K(t)$ of a kernel function K.

Condition 3.1.

i) $\tilde{\phi}_\varepsilon(t) \neq 0$ for all $t \in R$. Moreover, expression (1.15) holds with $\beta_1 = \beta_0$.

ii) $\tilde{\phi}_K(t)$ has a finite support $(-d, d)$.

iii) There exist positive constants δ, a_2 and ℓ such that $|\tilde{\phi}_K(t)| \leq a_2(d-t)^\ell$, for $t \in (d-\delta, d)$.

iv) $\tilde{\phi}_K(t) \geq a_3(d-t)^\ell$ for $t \in (d-\delta, d)$, where a_3 is a positive constant.

v) With $\tilde{R}_\varepsilon(t)$ and $\tilde{I}_\varepsilon(t)$ being the real and the imaginary part of $\tilde{\phi}_\varepsilon(t)$, assume that either $\tilde{I}_\varepsilon(t) = o(\tilde{R}_\varepsilon(t))$ or $\tilde{R}_\varepsilon(t) = o(\tilde{I}_\varepsilon(t))$, as $t \to \infty$.

We remark that condition i) assumes that the error distribution is super smooth. Under such an assumption, by Fourier's inversion, the density $\tilde{h}(u)$ of the error variable ε is bounded and has bounded derivatives of all orders. This entails that the marginal density $f(\boldsymbol{x};p)$ given by (1.5) is bounded:

$$f(\boldsymbol{x};p) \leq M,$$

for some $M > 0$. We also remark that the conditional density $f_{\boldsymbol{X}_0|Y_p}(\boldsymbol{u}|y_p)$ of $\boldsymbol{X}_0$ given $Y_p = y_p$ exists and is bounded and continuous in view of the smoothness of $\tilde{h}(u)$ and the convolution theorem. Condition ii) is a sufficient condition for the existence of the deconvolution kernel (1.2) for $b > 0$. Note that $\tilde{\phi}_K(d) = 0$ since $\tilde{\phi}_K(\cdot)$ is continuous. Condition iii) describes the behavior of $\tilde{\phi}_K(t)$ in a neighborhood of $t = d$. Conditions iv) and v) are used to develop lower bounds. Condition v) says that at the tail, the characteristic function $\tilde{\phi}_\varepsilon(\cdot)$ is either purely real or purely imaginary.

The following lemma gives both lower and upper bounds on the norms of the deconvolution kernel function (1.2). The proof is given in the appendix.

Lemma 3.1. *Under Condition 3.1 i) – iii), we have as $b_n \to 0$,*

$$||\tilde{W}_{b_n}||_\infty = O\left(b_n^{(\ell+1)\beta+\beta_0}\left(\log(\frac{1}{b_n})\right)^\ell \exp(a(d/b_n)^\beta)\right)$$

and

$$||\tilde{W}_{b_n}||_2 = O\left(b_n^{(\ell+0.5)\beta+\beta_0}\left(\log(\frac{1}{b_n})\right)^\ell \exp(a(d/b_n)^\beta)\right).$$

If moreover Condition 3.1 iv) and v) hold, then we have

$$|\tilde{W}_{b_n}(x)| \geq a_4\tilde{H}(x)b_n^{(\ell+1)\beta+\beta_0}\exp(a(d/b_n)^\beta),$$

for some $a_4 > 0$ uniformly in x on a bounded interval, where

$$\tilde{H}(x) = \begin{cases} |\cos dx|, & \text{if } \tilde{I}_\varepsilon(t) = o(\tilde{R}_\varepsilon(t)) \\ |\sin dx|, & \text{if } \tilde{R}_\varepsilon(t) = o(\tilde{I}_\varepsilon(t)). \end{cases}$$

The following two lemmas establish a lower bound on $\sigma_0^2(n)$ and the identity $\sigma^2(n) = \frac{1}{n}\sigma_0^2(n)(1 + o(1))$. We impose the following conditions.

Condition 3.2.

i) $E|\psi(Y_1)|^\nu < \infty$ for some $\nu > 2$.

ii) The processes $\{X_j^\circ, \varepsilon_j, Y_j\}$ are either ρ-mixing with $\sum_{j=1}^\infty \rho(j) < \infty$ or strongly mixing with $\sum_{j=1}^\infty j^\lambda [\alpha(j)]^{1-2/\nu} < \infty$ for some $\lambda > 0$.

Lemma 3.2. *Under Condition 3.1, we have for large n*

$$\sigma_0^2(n) \geq a_5 b_n^{2p[(\ell+1)\beta+\beta_0-0.5]} \exp\left(2ap(d/b_n)^\beta\right),$$

and

$$\sigma_0^2(n) \leq a_6 b_n^{2p[(\ell+1)\beta+\beta_0-1]} \left(\log\frac{1}{b_n}\right)^{2\ell p} \exp\left(2ap(d/b_n)^\beta\right),$$

for some constants $a_5, a_6 > 0$.

Proof. Since μ_n is bounded [see (2.4)], we have by conditioning on Y_p and using a change of variables that

$$\begin{aligned} \sigma_0^2(n) &= b_n^{-2p} E\left([\psi(Y_p) - m(\boldsymbol{x};p)] W_{b_n}((\boldsymbol{x} - \boldsymbol{X}_0)/b_n)\right)^2 + O(1) \\ &= b_n^{-p} E\left([\psi(Y_p) - m(\boldsymbol{x};p)]^2 \int_{R^p} W_{b_n}^2(\boldsymbol{u}) f_{\boldsymbol{X}_0|Y_p}(\boldsymbol{x} - b_n\boldsymbol{u}|Y_p)d\boldsymbol{u}\right) + O(1) \\ &\geq b_n^{-p} E\left([\psi(Y_p) - m(\boldsymbol{x};p)]^2 \int_{[-1,1]^p} W_{b_n}^2(\boldsymbol{u}) f_{\boldsymbol{X}_0|Y_p}(\boldsymbol{x} - b_n\boldsymbol{u}|Y_p)d\boldsymbol{u}\right) + O(1) \end{aligned}$$

and by Lemma 3.1 and the factorization (1.8) of W_{b_n}

$$\begin{aligned} \sigma_0^2(n) \geq\ & a_4^{2p} b_n^{2p[(\ell+1)\beta+\beta_0-0.5]} \exp\left(2ap(d/b_n)^\beta\right) \\ & \times E\left([\psi(Y_p) - m(\boldsymbol{x};p)]^2 \int_{[-1,1]^p} \prod_{j=1}^p \tilde{H}^2(u_j) f_{\boldsymbol{X}_0|Y_p}(\boldsymbol{x} - b_n\boldsymbol{u}|Y_p)d\boldsymbol{u}\right) + O(1). \end{aligned}$$

By the continuity of $f_{\boldsymbol{X}_0|Y_p}$, we then have

$$\begin{aligned} \sigma_0^2(n) \geq\ & a_4^{2p} b_n^{2p[(\ell+1)\beta+\beta_0-0.5]} \exp\left(2ap(d/b_n)^\beta\right) \left(\int_{-1}^1 \tilde{H}(u)du\right)^p \\ & E\left([\psi(Y_p) - m(\boldsymbol{x};p)]^2 f_{\boldsymbol{X}_0|Y_p}(\boldsymbol{x}|Y_p)\right)(1 + o(1)). \end{aligned}$$

The first conclusion follows. The second conclusion follows immediately from the bound on $\|\tilde{W}_{b_n}\|_\infty$ in Lemma 3.1. $\square$.

Lemma 3.3. *Under Conditions 3.1 and 3.2, we have*

$$\sum_{j=1}^{n-1} \left|\text{cov}(\tilde{Z}_{n,0}, \tilde{Z}_{n,j})\right| = o\left(\sigma_0^2(n)\right)$$

and

$$\sigma^2(n) = \frac{1}{n}\sigma_0^2(n)(1 + o(1)),$$

where $\sigma^2(n)$ and $\sigma_0^2(n)$ are given by (3.1) and (3.2).

Proof. Let $\tilde{I}_{n,j} = |\text{cov}(\tilde{Z}_{n,0}, \tilde{Z}_{n,j})|$ and

$$c_n = [\exp(ap(d/b_n)^\beta)]. \tag{3.3}$$

Then

$$\sum_{j=1}^{n-1} \left|\text{cov}(\tilde{Z}_{n,0}, \tilde{Z}_{n,j})\right| = \left(\sum_{j=1}^{p-1} + \sum_{j=p}^{c_n} + \sum_{j=c_n+1}^{n-1}\right) \tilde{I}_{n,j} \equiv J_1 + J_2 + J_3. \tag{3.4}$$

We now deal with each of the above three terms. For $1 \le j \le p-1$, by (2.3), we write

$$\tilde{I}_{n,j} = b_n^{-2p} E\Big([\psi(Y_p) - m(\boldsymbol{x};p)][\psi(Y_{p+j}) - m(\boldsymbol{x};p)] W_{b_n}([\boldsymbol{x} - \boldsymbol{X}_0]/b_n) W_{b_n}([\boldsymbol{x} - \boldsymbol{X}_j]/b_n)\Big) + O(1),$$

since μ_n is bounded by (2.4). Put

$$(\boldsymbol{X}^\circ)' = (X_1^\circ, \cdots, X_j^\circ); \qquad \boldsymbol{\varepsilon}' = (\varepsilon_1, \cdots, \varepsilon_j)$$

$$(\boldsymbol{X}^\circ)'' = (X_{j+1}^\circ, \cdots, X_p^\circ); \qquad \boldsymbol{\varepsilon}'' = (\varepsilon_{j+1}, \cdots, \varepsilon_p)$$

$$(\boldsymbol{X}^\circ)''' = (X_{p+1}^\circ, \cdots, X_{p+j}^\circ); \qquad \boldsymbol{\varepsilon}''' = (\varepsilon_{p+1}, \cdots, \varepsilon_{p+j})$$

and $\boldsymbol{X}' = (\boldsymbol{X}^\circ)' + \boldsymbol{\varepsilon}'$, $\boldsymbol{X}'' = (\boldsymbol{X}^\circ)'' + \boldsymbol{\varepsilon}''$ and $\boldsymbol{X}''' = (\boldsymbol{X}^\circ)''' + \boldsymbol{\varepsilon}'''$. Then, by conditioning on $(\boldsymbol{X}^\circ)''$, $\boldsymbol{\varepsilon}'', Y_p$ and Y_{p+j}, we have

$$\tilde{I}_{n,j} = b_n^{-2p} E\left\{q_2(\boldsymbol{X}'')q_4(Y_p, Y_{p+j}) E\left[q_1(\boldsymbol{X}')q_3(\boldsymbol{X}''')\Big|(\boldsymbol{X}^\circ)'', \boldsymbol{\varepsilon}'', Y_p, Y_{p+j}\right]\right\} + O(1), \tag{3.5}$$

where

$$q_1(\boldsymbol{X}') = \prod_{l=1}^{j} \tilde{W}_{b_n}\left(\frac{x_l - X_l}{b_n}\right)$$
$$q_2(\boldsymbol{X}'') = \prod_{l=j+1}^{p} \tilde{W}_{b_n}\left(\frac{x_l - X_l}{b_n}\right) \prod_{l=j+1}^{p} \tilde{W}_{b_n}\left(\frac{x_{l-j} - X_l}{b_n}\right)$$
$$q_3(\boldsymbol{X}''') = \prod_{l=p+1}^{p+j} \tilde{W}_{b_n}\left(\frac{x_{l-j} - X_l}{b_n}\right)$$

and

$$q_4(Y_p, Y_{p+j}) = [\psi(Y_p) - m(\boldsymbol{x};p)][\psi(Y_{p+j}) - m(\boldsymbol{x};p)].$$

By (1.2) and Fubinis's theorem, the inner conditional expectation is

$$\begin{aligned}
J_5 &\equiv E\left[q_1(\boldsymbol{X}')q_3(\boldsymbol{X}''')\Big|(\boldsymbol{X}^\circ)'',\ \boldsymbol{\varepsilon}'', Y_p, Y_{p+j}\right] \\
&= \frac{1}{(2\pi)^{2j}} \int_{R^{2j}} E\left[\exp\left(-\frac{i}{b_n}\sum_{l=1}^{j} t_l(x_l - X_l^\circ - \varepsilon_l)\right.\right. \\
&\qquad \left.\left. -\frac{i}{b_n}\sum_{l=p+1}^{p+j} t_l(x_{l-j} - X_l^\circ - \varepsilon_l)\right) \Big|(\boldsymbol{X}^\circ)'',\ \boldsymbol{\varepsilon}'', Y_p, Y''_{p+j}\right] \\
&\qquad \times \prod_{l=1}^{j}[\tilde{\phi}_K(t_l)/\tilde{\phi}_\varepsilon(t_l/b_n)] \prod_{l=p+1}^{p+j}[\tilde{\phi}_K(t_l)/\tilde{\phi}_\varepsilon(t_l/b_n)]d\boldsymbol{t}'d\boldsymbol{t}''' \\
&= \frac{1}{(2\pi)^{2j}} \int_{R^{2j}} \exp\left(-\frac{i}{b_n}\sum_{l=1}^{j} t_l x_l - \frac{i}{b_n}\sum_{l=p+1}^{p+j} t_l x_{l-j}\right) \phi\left(\boldsymbol{t}', \boldsymbol{t}'''\Big|(\boldsymbol{X}^\circ)'', Y_p, Y_{p+l}\right) \\
&\qquad \times \prod_{l=1}^{j} \tilde{\phi}_K(t_l) \prod_{l=p+1}^{p+j} \tilde{\phi}_K(t_l) d\boldsymbol{t}'d\boldsymbol{t}''',
\end{aligned}$$

where $\phi\left(\boldsymbol{t}', \boldsymbol{t}'''\Big|(\boldsymbol{X}^\circ)'', Y_p, Y_{p+j}\right)$ is the conditional characteristic function of $(\boldsymbol{X}^\circ)', (\boldsymbol{X}^\circ)'''$ given $\{(\boldsymbol{X}^\circ)'', Y_p, Y_{p+j}\}$. Therefore,

$$|J_5| \le \frac{1}{(2\pi)^{2j}} ||\tilde{\phi}_K||_1^{2j}.$$

Consequently, by (3.5), for $1 \le j \le p-1$, we have

$$\begin{aligned}
|\tilde{I}_{n,j}| &\le \frac{||\tilde{\phi}_K||_1^{2j}}{(2\pi)^{2j} b_n^{2p}} E\Big|q_2(\boldsymbol{X}'')q_4(Y_p, Y_{p+j})\Big| + O(1) \\
&\le \frac{C_1}{b_n^{2p}} ||\tilde{W}_{b_n}||_\infty^{2(p-j)}, \qquad (3.6)
\end{aligned}$$

for some positive constant C_1. The same argument yields

$$|\tilde{I}_{n,j}| \leq \frac{C_2}{b_n^{2p}}, \qquad j \geq p. \tag{3.7}$$

Thus, by (3.4) and (3.6), we have

$$J_1 \leq \frac{C_1}{b_n^{2p}} \sum_{j=1}^{p-1} \|\tilde{W}_{b_n}\|_\infty^{2(p-l)} = O\left(b_n^{-2p}\|\tilde{W}_{b_n}\|_\infty^{2(p-1)}\right).$$

This together with the upper bound on $\|\tilde{W}_{b_n}\|_\infty$ in Lemma 3.1 and the lower bound on $\sigma_0^2(n)$ in Lemma 3.2 show that

$$\frac{J_1}{\sigma_0^2(n)} \leq \frac{C_3}{b_n^{p+2(\ell+1)\beta+\beta_0}} \left(\log \frac{1}{b_n}\right)^{2\ell(p-1)} \exp(-2a(d/b_n)^\beta) = o(1), \tag{3.8}$$

where C_3 is a positive constant. For J_2, (3.7) leads to

$$J_2 \leq c_n \frac{C_2}{b_n^{2p}}.$$

Hence, by the choice of c_n given in (3.3) together with the lower bound on $\sigma_0^2(n)$ given by Lemma 3.2, we have

$$\frac{J_2}{\sigma_0^2(n)} = O\left(b_n^{-2p[(\ell+1)\beta+\beta_0+0.5]} \exp(-ap(d/b_n)^\beta)\right) = o(1), \tag{3.9}$$

Finally, we consider J_3. For ρ-mixing processes, we have

$$J_3 \leq \sigma_0^2(n) \sum_{j=c_n+1}^{\infty} \rho(j-p+1) = o\Big(\sigma_0^2(n)\Big).$$

From this together with (3.4), (3.8) and (3.9), we have proved the first conclusion for the ρ-mixing processes. For strongly mixing processes, we first note that by (2.3)

$$E|\tilde{Z}_{n,0}|^\nu \leq 2^\nu \frac{\|\tilde{W}_{b_n}\|_\infty^{\nu p}}{b_n^{\nu p}} E|\psi(Y_p) - m(x;p)|^\nu + O(1). \tag{3.10}$$

Then employing Davydov's lemma, we obtain that

$$\begin{aligned} |\tilde{I}_{n,j}| &\leq 8[\alpha(j-p+1)]^{1-2/\nu} \left(E|\tilde{Z}_{n,0}|^\nu\right)^{2/\nu} \\ &\leq C_4[\alpha(j-p+1)]^{1-2/\nu} \frac{\|\tilde{W}_{b_n}\|_\infty^{2p}}{b_n^{2p}} \end{aligned}$$

for some constant C_4. Thus,

$$J_3 \leq 2^{\nu+1} C_4 \frac{\|\tilde{W}_{b_n}\|_\infty^{2p}}{b_n^{2p}} \sum_{j=c_n}^{\infty} [\alpha(j-p+1)]^{1-2/\nu} \leq C_4 \frac{\|\tilde{W}_{b_n}\|_\infty^{2p}}{c_n^{\lambda} b_n^{2p}} \sum_{j=c_n}^{\infty} j^{\lambda} [\alpha(j-p+1)]^{1-2/\nu}.$$

Using again the upper bounded on $\|\tilde{W}_{b_n}\|_\infty$ and the lower bound of $\sigma_0^2(n)$ given by Lemma 3.2, we have

$$\frac{J_3}{\sigma_0^2(n)} \leq O\left(c_n^{-\lambda} b_n^{-p} [\log \frac{1}{b_n}]^{2\ell p} \sum_{j=c_n}^{\infty} j^{\lambda} [\alpha(j-p+1)]^{1-2/\nu} \right) = o(1).$$

Combining this with (3.4), (3.8) and (3.9) proves the first part of the Lemma for strongly mixing processes. The second conclusion follows directly from (2.6) and the first one. □

Lemma 3.4. *Under Conditions 3.1, we have*

$$\operatorname{var}\left(\hat{f}_n^{\circ}(\boldsymbol{x};p)\right) = O\left(\frac{1}{n} b_n^{2p[(\ell+1)\beta+\beta_0-1]} \left(\log(\frac{1}{b_n})\right)^{2p\ell} \exp\left(2ap(d/b_n)^{\beta}\right) \right).$$

Moreover, if $b_n \to 0$ such that $b_n > \gamma d (2ap/\log n)^{1/\beta}$ for some $\gamma > 1$, then

$$\hat{f}_n^{\circ}(\boldsymbol{x};p) \xrightarrow{P} f^{\circ}(\boldsymbol{x};p)$$

at the continuity points of $f^{\circ}(\boldsymbol{x};p)$.

Proof. The same argument as in the proof of Lemma 3.3 leads to

$$\begin{aligned} \operatorname{var}\left(\hat{f}_n^{\circ}(\boldsymbol{x};p)\right) &= O\left(\frac{1}{n} E \left| \frac{1}{b_n^p} W_{b_n}\left(\frac{\boldsymbol{x} - \boldsymbol{X}_0}{b_n} \right) \right|^2 \right) \\ &= O\left(\frac{1}{n b_n^{2p}} \|\tilde{W}_{b_n}\|_\infty^{2p} \right). \end{aligned}$$

By using the upper bound on $\|\tilde{W}_{b_n}\|_\infty$ given in Lemma 3.1, we obtain the first result. For the second result, by the assumption on the bandwidth, we have as $n \to \infty$

$$\operatorname{var}\left(\hat{f}_n^{\circ}(\boldsymbol{x};p)\right) \to 0,$$

and by Proposition 1.1, we have $\hat{f}_n^{\circ}(\boldsymbol{x};p) \to f^{\circ}(\boldsymbol{x};p)$ in quadratic-mean. Hence, the second conclusion follows. □.

We remark that a similar conclusion to Lemma 3.4 was proved in Masry (1991a). The current result is slightly stronger and broader.

3.2 Main Results

The goal of this section is to establish the asymptotic normality for the regression estimator (1.10). To this end, we first discuss the asymptotic normality for $\tilde{Q}_n$, the dominating term in the numerator of (1.14), and then use Lemma 3.4 to show the asymptotic normality for $\hat{m}(\boldsymbol{x};p)$ via (1.14). We need the following conditions.

Condition 3.3A.

Assume $nb_n^{p\gamma} \to \infty$ as $n \to \infty$ for some $\gamma > 1$. Let $\{s_n\}$ be a sequence of positive integers defined by $s_n = [(nb_n^{p\gamma})^{1/2}]$. For strongly mixing processes, $\alpha(k)$ satisfies $(nb^{-p\gamma})^{1/2}\alpha(s_n) \to 0$ as $n \to \infty$; for ρ-mixing processes, $\rho(k)$ satisfies $(nb_n^{-p\gamma})^{1/2}\rho(s_n) \to 0$ as $n \to \infty$.

Condition 3.3B.

Assume that $n^{\nu-2}b_n^{p\gamma\nu} \to \infty$ as $n \to \infty$ for some $\gamma > 1$, where ν is given in Condition 3.2. Let $\{s_n\}$ be a sequence of positive integers given by

$$s_n = [(n^{\nu-2}b_n^{p\gamma\nu})^{\frac{1}{2(\nu-1)}}].$$

For strongly mixing processes, $\alpha(k)$ satisfies $(nb_n^{-p\gamma})^{1/2}\alpha(s_n) \to 0$ as $n \to \infty$; for ρ-mixing processes, $\rho(k)$ satisfies $(nb_n^{-p\gamma})^{1/2}\rho(s_n) \to 0$ as $n \to \infty$.

We remark that Condition 3.3A is weaker than Condition 3.3B, since s_n given in Condition 3.3A is larger. Note that Condition 3.3A is very similar to Condition 2.3 in the ordinary smooth case. For bounded ψ, Theorem 3.1 shows that the asymptotic normality holds under this weaker condition. Interesting examples of bounded ψ include estimating conditional cdf and estimating mean regression function when responses are bounded (e.g. binary response).

Theorem 3.1. *Under Conditions 3.1 and 3.2, we have*

A. If $\psi(\cdot)$ is bounded and Condition 3.3 A holds, then

$$\sqrt{n}\tilde{Q}_n(\boldsymbol{x};p)/\sigma_0(n) \xrightarrow{\mathcal{L}} N(0,1).$$

B. If Condition 3.3 B holds, then the above asymptotic normality holds.

Proof. We first prove part A; the proof of part B is outlined following the proof of part A.

We first normalize $\tilde{Z}_{n,j}$ in (2.3) as follows:

$$Z_{n,j} = \tilde{Z}_{n,j}/\sigma_0(n), \qquad S_n = \sum_{j=0}^{n-1} Z_{n,j}$$

so that

$$\mathrm{var}\Big(Z_{n,j}\Big) = 1. \tag{3.11}$$

In view of Lemma 3.3, we have

$$\sum_{j=1}^{n-1}\Big|\mathrm{cov}(Z_{n,0}, Z_{n,j})\Big| \to 0, \qquad \mathrm{var}(S_n) = \frac{1}{n}\Big(1 + o(1)\Big). \tag{3.12}$$

With such a normalization, (2.5) leads to

$$\sqrt{n}\tilde{Q}_n(\boldsymbol{x}; p)/\sigma_0(n) = \sqrt{n/(n-p+1)}\frac{1}{\sqrt{n-p+1}}S_{n-p+1}.$$

Thus, it suffices to show that

$$n^{-1/2}S_n \xrightarrow{\mathcal{L}} N(0,1). \tag{3.13}$$

We now employ the big and small block arguments as in the proof of Theorem 2.1. Let γ_1 be a real number satisfying $1 < \gamma_1 < \gamma$, where γ is given in Condition 3.3 A. Let the small block size $s = s_n$, the big block size $r = r_n$ and the number of blocks $k = k_n$ be

$$s_n = [(nb_n^{p\gamma})^{1/2}]; \qquad r_n = [(nb_n^{p\gamma_1})^{1/2}]; \qquad k_n = [\frac{n}{r_n + s_n}].$$

Then, it is easy to verify as $n \to \infty$ that

$$\frac{s_n}{r_n} \longrightarrow 0; \quad \frac{r_n}{n} \longrightarrow 0; \quad \frac{r_n}{(nb_n^p)^{1/2}}\left(\log\frac{1}{b_n}\right)^{\ell p} \longrightarrow 0 \tag{3.14}$$

and that

$$\frac{n}{r_n}\alpha(s_n) \to 0. \tag{3.15}$$

Define respectively η_j, ξ_j, ζ_j, and S'_n, S''_n, S'''_n as in (2.11) – (2.14). To prove (3.13), we need only to verify that as $n \to \infty$,

$$\frac{1}{n}E(S''_n)^2 \to 0, \quad \frac{1}{n}E(S'''_n)^2 \to 0 \tag{3.16}$$

$$\left| E\left[\exp(itS'_n)\right] - \prod_{j=0}^{k-1} E\left[\exp(it\eta_j)\right] \right| \to 0 \tag{3.17}$$

$$\frac{1}{n}\sum_{j=0}^{k-1} E\left(\eta_j^2\right) \to 1 \tag{3.18}$$

$$\frac{1}{n}\sum_{j=0}^{k-1} E\left(\eta_j^2 I\{|\eta_j| \geq \varepsilon\sqrt{n}\}\right) \to 0 \tag{3.19}$$

for every ε.

As in (2.23), using (3.11) and (3.12), we have

$$\text{var}(\xi_j) = s(1 + o(1)). \tag{3.20}$$

One can similarly show that with F_1 and F_2 given by (2.22), we have

$$F_1 = O(k_n s_n) = o(n), \quad F_2 = o(n).$$

Thus $\frac{1}{n}E(S'')^2 \to 0$. Note that by (3.11) and (3.12)

$$\frac{1}{n}E(S'''_n)^2 \leq \frac{1}{n}(n - k(r+s)) + \sum_{j=1}^{n-1} |\text{cov}(Z_{n,0}, Z_{n,j})| \to 0.$$

Hence, (3.16) holds.

Next, using Lemma 2.4 and (3.15) we have

$$\left| E\left[\exp(itS'_n)\right] - \prod_{j=0}^{k-1} E\left[\exp(it\eta_j)\right] \right| \leq 16k_n\alpha(s_n + 2 - p) \sim 16\frac{n}{r_n}\alpha(s_n + 2 - p) \to 0,$$

establishing (3.17).

Now, using (3.20) with s replaced by r, we obtain

$$E\eta_0^2 = r_n(1 + o(1)) \tag{3.21}$$

so that by stationarity

$$\frac{1}{n}\sum_{j=0}^{k-1} E\left(\eta_j^2\right) = \frac{k_n r_n}{n}(1+o(1)) \to 1.$$

This establishes (3.18).

Finally, we verify the Lindeberg condition (3.19). Since $\psi(Y)$ is bounded, we have

$$|\eta_0| \le r_n \sup_{1\le j\le r_n} |\tilde{Z}_{n,j}|/\sigma_0(n) \le C_1 r_n \left(\|\tilde{W}_{b_n}\|_\infty / b_n\right)^p / \sigma_0(n),$$

where C_1 is a positive constant. By using the upper bound on $\|\tilde{W}_{b_n}\|_\infty$ in Lemma 3.1 and the lower bound on $\sigma_0(n)$ in Lemma 3.2, we obtain

$$|\eta_0|/\sqrt{n} = O\left(\frac{r_n}{(nb_n^p)^{1/2}}\left(\log\frac{1}{b_n}\right)^{\ell p}\right),$$

which tends to zero by (3.14). Hence (3.19) holds for bounded ψ since the set $\{|\eta_0| \ge \varepsilon\sqrt{n}\}$ is asymptotically empty. This completes the proof of the first part.

For part B, we proceed as follows. Define the small block size s_n as in Condition 3.3 B, and let

$$r_n = [(n^{\nu-2} b_n^{p\gamma_1\nu})^{\frac{1}{2(\nu-1)}}] \qquad \text{and} \qquad k_n = [\frac{n}{r_n+s_n}],$$

where $1 < \gamma_1 < \gamma$. Then, it is easy to show that

$$\frac{s_n}{r_n} \to 0; \quad \frac{r_n}{n} \to 0; \quad b_n^{-p}\left(\log\frac{1}{b_n}\right)^{2\ell p} \frac{r_n^{2(1-1/\nu)}}{n^{1-2/\nu}} \to 0 \tag{3.22}$$

and that (3.15) holds. The same argument as in part A) establishes (3.16) – (3.18). Thus, we need only to show (3.19). Since $k_n r_n/n \to 1$, it suffices to show, by stationarity, that

$$F_3 \equiv \frac{1}{r_n} E\left(\eta_0^2 I\{|\eta_0| \ge \varepsilon\sqrt{n}\}\right) \to 0 \tag{3.23}$$

as $n\to\infty$. Let $A_n = \{|\eta_0| > \varepsilon\sqrt{n}\}$. Then by Jensen's inequality,

$$\left(\sum_{j=1}^{r_n} Z_{n,j}/r_n\right)^2 \le \sum_{j=1}^{r_n} Z_{n,j}^2/r_n$$

so that

$$F_3 \le \sum_{j=1}^{r_n} E Z_{n,j}^2 I_{A_n}.$$

Let $\bar{p} = \nu/2$ and $\bar{q} = v/(\nu - 2)$ be its conjugate number so that

$$\frac{1}{\bar{p}} + \frac{1}{\bar{q}} = 1,$$

where ν was given in Condition 3.2. By Hölder's inequality and stationarity, we obtain

$$F_3 \leq r_n \left(E|Z_{n,0}|^{2\bar{p}}\right)^{1/\bar{p}} \left(P(A_n)\right)^{1/\bar{q}}$$

and by Tchebychev's inequality

$$F_3 \leq r_n \left(E|Z_{n,0}|^{2\bar{p}}\right)^{1/\bar{p}} \left(\frac{E|\eta_0|^2}{\varepsilon^2 n}\right)^{1/\bar{q}}.$$

By (3.10), Lemma 3.1 and Lemma 3.2,

$$\left(E|Z_{n,0}|^{2\bar{p}}\right)^{1/\bar{p}} = \left(\frac{E|\tilde{Z}_{n,0}|^{\nu}}{\sigma_0^{\nu}(n)}\right)^{2/\nu} = O\left(b_n^{-p}\left(\log\frac{1}{b_n}\right)^{2\ell p}\right).$$

Using this, together with (3.21), we have

$$F_3 = O\left(b_n^{-p}\left(\log\frac{1}{b_n}\right)^{2\ell p} r_n^{2(1-1/\nu)} n^{-(1-2/\nu)}\right)$$

which tends to zero by (3.22). Hence, (3.23) holds and this completes the proof. □

By using (1.14) together with Lemma 3.4 and Theorem 3.1, we have

Theorem 3.2. *If the assumptions of Theorem 3.1 hold, then for $\boldsymbol{x} \in R^p$ such that $f^{\circ}(\boldsymbol{x};p) > 0$ we have*

$$\sqrt{n}\frac{\hat{m}_n(\boldsymbol{x};p) - m(\boldsymbol{x};p) - B_n(\boldsymbol{x};p)}{\sigma_0(n)} \xrightarrow{\mathcal{L}} N\left(0, \frac{1}{f^{\circ 2}(\boldsymbol{x};p)}\right),$$

provided that $b_n \to 0$ such that $b_n \geq \gamma d(2ap/\log n)^{1/\beta}$ some $\gamma > 1$, where $\sigma_0^2(n)$ is given by (3.1).

We remark that for the ordinary smooth case, the asymptotic normality [see Theorem 2.1 and Corollary 2.1] admits a classical form. However, for super smooth case, the asymptotic normality does not have a classical form due to the unavailability of the asymptotic rate and constant of $\sigma_0^2(n)$. Note that the explicit upper and lower bound on $\sigma_0^2(n)$ are given in Lemma 3.2. These bounds are nearly sharp.

Remark 3.1. Under Condition 3.3B, there is a tradeoff between the order of the moment ν in $E|\psi(Y_1)|^\nu < \infty$ and the rate of decay of the mixing coefficients: the larger ν, the weaker the condition on the mixing coefficients. For example, if $b_n \geq \gamma d(2ap/\log n)^{1/\beta}$ as in Theorem 3.1 and the strongly mixing coefficient satisfies $\alpha(j) = O(j^{-c})$, then Conditions 3.2 and 3.3 B hold when

$$c > \frac{\nu}{\nu - 2}.$$

Appendix: Proof of Lemma 3.1

Let c be a genetic constant and

$$\gamma_n = \lambda b_n^\beta \log \frac{1}{b_n}, \tag{A.1}$$

where λ is a positive constant. Then, by (1.2) we obtain

$$\begin{aligned} \|\tilde{W}_{b_n}\|_\infty &\leq \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{|\tilde{\phi}_K(t)|}{|\tilde{\phi}_\varepsilon(t/b_n)|}dt \\ &= \frac{1}{\pi}\left(\int_0^{d-\gamma_n} + \int_{d-\gamma_n}^{d}\right)\frac{|\tilde{\phi}_K(t)|}{|\tilde{\phi}_\varepsilon(t/b_n)|}dt \\ &\equiv I_1 + I_2. \end{aligned} \tag{A.2}$$

We first deal with I_1. With M large but fixed, condition 3.1 i) leads to

$$\begin{aligned} I_1 &= \left(\int_0^{Mb_n} + \int_{Mb_n}^{d-\gamma_n}\right)\frac{|\tilde{\phi}_K(t)|}{|\tilde{\phi}_\varepsilon(t/b_n)|}dt \\ &\leq \frac{Mb_n}{\min_{0\leq t\leq M}|\tilde{\phi}_\varepsilon(t)|} + c\int_{Mb_n}^{d-\gamma_n}\left(\frac{t}{b_n}\right)^{-\beta_0}\exp\left(a(t/b_n)^\beta\right)dt \\ &\leq cb_n^{\beta_0}\int_{Mb_n}^{d-\gamma_n} t^{-\beta_0}\exp\left(a(t/b_n)^\beta\right)dt. \end{aligned} \tag{A.3}$$

By taking derivative with respect to t, it is easy to show that the integrand in (A.3) is an increasing function of t when $t \geq Mb_n$ and hence is bounded by its value at the point

$t = d - \gamma_n$. Therefore, we have

$$I_1 = O\left(b_n^{\beta_0} \exp\left(a(d/b_n)^\beta (1 - \gamma_n/d)^\beta\right)\right).$$

Taylor's expansion gives

$$(1 - \gamma_n/d)^\beta = 1 - \beta\gamma_n/d + O\left(\gamma_n^2\right) = 1 - \beta\gamma_n/d + o\left(b_n^\beta\right).$$

This together with (A.1) lead to

$$I_1 = O\left(b_n^{\beta_0 + \beta\lambda a d^{\beta-1}} \exp\left(a(d/b_n)^\beta\right)\right). \tag{A.4}$$

Next, we consider I_2. We first note that for $t \in [d - \gamma_n, d]$

$$(d-t)^\ell \leq \gamma_n^\ell; \qquad t^{-\beta_0 - (\beta - 1)} < c.$$

This together with Condition 3.1 iii) and (1.15) entail that

$$\begin{aligned} I_2 &\leq c \int_{d-\gamma_n}^{d} (d-t)^\ell (t/b_n)^{-\beta_0} \exp\left(a(t/b_n)^\beta\right) dt \\ &\leq c\gamma_n^\ell b_n^{\beta_0} \int_{d-\gamma_n}^{d} t^{\beta-1} \exp\left(a(t/b_n)^\beta\right) dt \\ &= O\left(b_n^{\beta_0+\beta+\ell\beta} \left(\log\frac{1}{b_n}\right)^\ell \exp\left(a(d/b_n)^\beta\right)\right). \end{aligned}$$

By choosing a large value of the constant λ, the upper bound of I_2 dominates I_1. Hence, by (A.2), we obtain the first conclusion. The second conclusion follows from the Parseval's identity:

$$\|\tilde{W}_{b_n}\|_2^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\tilde{\phi}_K(t)|^2 / |\tilde{\phi}_\varepsilon(t/b_n)|^2 dt$$

and a similar argument.

Now, we establish the third conclusion. First, write

$$\tilde{W}_{b_n}(x) = \frac{1}{2\pi}\left(\int_{-(d-\gamma_n)}^{d-\gamma_n} + \int_{[-(d-\gamma_n), d-\gamma_n]^c}\right) \exp(itx)\tilde{\phi}_K(t)/\tilde{\phi}_\varepsilon(t/b_n) dt \equiv J_1 + J_2. \tag{A.5}$$

By (A.4),

$$|J_1| \leq I_1 = O\left(b_n^{\beta_0 + \beta\lambda a d^{\beta-1}} \exp\left(a(d/b_n)^\beta\right)\right). \tag{A.6}$$

Next, by symmetry, we have

$$J_2 = \frac{1}{\pi}\int_{d-\gamma_n}^{d} \tilde{\phi}_K(t)\left(\cos tx \frac{\tilde{R}_\varepsilon(t/b_n)}{|\tilde{\phi}_\varepsilon(t/b_n)|^2} + \sin ty \frac{\tilde{I}_\varepsilon(t/b_n)}{|\tilde{\phi}_\varepsilon(t/b_n)|^2}\right) dt.$$

Without loss of generality, we treat the case that $\tilde{I}_\varepsilon(t/b_n) = o(|\tilde{R}_\varepsilon(t/b_n)|)$. In this case,

$$\begin{aligned} J_2 &= \frac{1}{\pi}\int_{d-\gamma_n}^{d} \tilde{\phi}_K(t)\cos tx \frac{\tilde{R}_\varepsilon(t/b_n)}{|\tilde{\phi}_\varepsilon(t/b_n)|^2} dt(1+o(1)) \\ &= \frac{1}{\pi}\left(\int_{d-\gamma_n}^{d-b_n^\beta} + \int_{d-b_n^\beta}^{d}\right)\tilde{\phi}_K(t)\cos tx \frac{\tilde{R}_\varepsilon(t/b_n)}{|\tilde{\phi}_\varepsilon(t/b_n)|^2} dt \\ &\equiv J_{2,1} + J_{2,2}. \end{aligned} \tag{A.7}$$

We remark that $\tilde{R}_\varepsilon(t/b_n)$ can not change its sign on the interval $[d-\gamma_n, d]$; otherwise, $\tilde{R}_\varepsilon(t/b_n)$ would have a root, say, t_n, which implies $\tilde{\phi}_\varepsilon(t_n/b_n) = \tilde{R}_\varepsilon(t_n/b_n) + i\tilde{I}_\varepsilon(t_n/b_n) = 0$ [since we assume $\tilde{I}_\varepsilon(t/b_n) = o(\tilde{R}_\varepsilon(t/b_n))$] and contradicts with the assumption that $\phi_\varepsilon(t) \neq 0$. Also, by Condition 3.1 vi), $\tilde{\phi}_K(t) > 0$ on the interval $(d-\gamma_n, d)$. For the point $x = (k+0.5)\pi/d$, $k = 0, \pm1, \pm2, \cdots$, the third conclusion follows naturally since $\cos(dx) = 0$. When $x \neq (k+0.5)\pi/d$, on the interval $t \in [d-\gamma_n, d]$, we have

$$\cos(tx) = \cos(dx)\Big(1 + o(1)\Big)$$

uniformly in x on a bounded interval. Thus, the function $\cos(tx)$ can not change its sign on $[d-\gamma_n, d]$. These imply that $J_{2,1}$ and $J_{2,2}$ have the same signs, say positive. Thus (A.7) entails

$$|J_2| \geq |J_{2,2}|.$$

Using the tail condition (1.15) and Condition 3.1 iv), we obtain

$$\begin{aligned} |J_2| &\geq c\left|\cos(dx)\Big(1+o(1)\Big)\right| \int_{d-b_n^\beta}^{d} (d-t)^\ell (t/b_n)^{-\beta_0} \exp\Big(a(t/b_n)^\beta\Big)\, dt \\ &\geq c|\cos(dx)| \left(\frac{d-b_n^\beta}{b_n}\right)^{-\beta_0} \exp\left(a\left(\frac{d-b_n^\beta}{b_n}\right)^{-\beta}\right) \int_{d-b_n^\beta}^{d} (d-t)^\ell dt \\ &\geq c|\cos dx| b_n^{\beta_0 + (\ell+1)\beta} \exp\Big(a(d/b_n)^\beta (1 - b_n^\beta/d)^\beta\Big). \end{aligned}$$

The second inequality follows from the fact that the function $t^{-\beta_0}\exp(a(t/b_n)^\beta)$ is an increasing function when $t \in [d-b_n^\beta, d]$. Using the fact that for small x,

$$(1-x)^\beta \geq 1 - \frac{\beta}{2}x,$$

we have

$$J_2 \geq c|\cos dx| b_n^{\beta_0+(\ell+1)\beta} \exp\left(a(d/b_n)^\beta\right).$$

This together with (A.5) and (A.6) lead to the desired lower bound by choosing a large value of λ so that J_2 dominates J_1. □

References

[1] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83**, 1184-1186.

[2] Collomb, G. and Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stoch. Processes Appl.*, **23**, 77-89.

[3] Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist*, to appear.

[4] Fan, J. (1991b). Asymptotic normality for deconvolving kernel density estimators. *Sankhyā*, to appear.

[5] Fan, J. (1992). Deconvolution with supersmoother distributions. *Canad. J. Statist.*, to appear.

[6] Fan, J. and Truong, Y. (1990). Nonparametric regression with errors-in-variables. Tentatively accepted by *Ann. Statist.*

[7] Fan, J., Truong, Y. and Wang, Y. (1990). Measurement errors regression: A nonparametric approach. *Institute of Statistics Mimeo Series, # 2042*, University of North Carolina, Chapel Hill.

[8] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications.* Academic Press, New York.

[9] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Boston.

[10] Kolmogorov, A.N. and Rozanov, Yu. A. (1960). On strong mixing conditions for stationary Gaussian processes. *Theorey Prob. Appl.*, **52**, 204-207.

[11] Liu, M. C. and Taylor, R. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.*, **17**, 399-410.

[12] Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie view. Gebiete*, **61**, 405-415.

[13] Masry, E. (1991a). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory*, **37**, 1105-1115.

[14] Masry, E. (1991b). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. Submitted for publication.

[15] Masry, E. (1991c). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. Submitted for publication.

[16] Masry, E. (1991d). Regression estimation with errors-in-variables for stationary processes. Submitted for publication.

[17] Nadaraya, E. A. (1964). On estimating regression. *Theor. Prob. Appl.*, **9**, 141-142.

[18] Prentice, R. L. (1986). Binary regression using an extended Beta-Binomial distribution, with discussion of correlation induced by covariate measurement errors. *J. Amer. Statist. Assoc.*, **81**, 321-327.

[19] Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Series Anal.*, **4**, 185-297.

[20] Robinson, P. M. (1986). On the consistency and finite sample properties of nonparametric kernel time series regression, auto regression, and density estimators. *Ann. Inst. Statist. Math.*, **38**, 539-549.

[21] Rosenblatt, M. (1956). A central limit theorem and strong mixing conditions. *Proc. Nat. Acad. Sci.*, **4**, 43-47.

[22] Rosenblatt, M. (1969). Conditional probability density and regression estimates. In *Multivariate Analysis II*, ed. Krishnaiah, 25-31.

[23] Roussas, G. G. and Tran, L. T. (1991). Asymptotic normality of the recursive kernel regression estimate under dependence conditions. *Ann. Statist.*, to appear.

[24] Stefanski, L. and Carroll, R. J. (1991) . Deconvolving kernel density estimators. *Statistics*, **21**, 169-184.

[25] Truong, Y. K. and C. J. Stone (1991). Nonparametric function estimation involving time series. *Annals of Statistics*, to appear.

[26] Truong, Y. K. (1991). Nonparametric curve estimation with time series errors. *J. Statistical Planning and Inference*, to appear.

[27] Volkonskii, V.A. and Rozanov, Yu. A. (1959). Some limit theorems for random functions. *Theory Prob. Appl.*, **4**, 178-197.

[28] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A.*, **26**, 359-372.

[29] Whittemore, A. S. and Keller, J. B. (1988). Approximations for errors in variables regression. *J. Amer. Statist. Assoc.*, **83**, 1057-1066.

[30] Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, **18**, 806-830.