Oliver B. Linton and Yoon-Jae Whang

Nonparametric estimation with aggregated data

Originally published in Econometric theory, 18 (2). pp. 420-468 © 2002 Cambridge University Press.

Econometric Theory, **18**, 2002, 420–468. Printed in the United States of America.
DOI: 10.1017.S0266466602182089

NONPARAMETRIC ESTIMATION WITH AGGREGATED DATA

Oliver Linton
London School of Economics

Yoon-Jae Whang
EWHA University

We introduce a kernel-based estimator of the density function and regression function for data that have been grouped into family totals. We allow for a common intrafamily component but require that observations from different families be independent. We establish consistency and asymptotic normality for our procedures. As usual, the rates of convergence can be very slow depending on the behavior of the characteristic function at infinity. We investigate the practical performance of our method in a simple Monte Carlo experiment.

1. INTRODUCTION

Grouped or aggregated data occur in many contexts in economics. Data aggregated by family, by region, and by other levels are often all that is available to the empirical researcher. If the object of interest is the underlying individual relationship, then grouping can imply some consequences for estimation and inference, depending on the model. Inference based on linear models is little affected by sort of grouping we consider, because it is a linear operation. The slope parameters of the aggregated model are the same as in the disaggregated model, and the usual least squares estimators are consistent. The worst thing that can happen is some heteroskedasticity when the groups are not of equal number, in which case one must correct the standard errors and/or improve efficiency by weighting. However, nonlinear models, and in particular nonparametric models, suffer considerable problems in the presence of grouping, because the grouped data regression function can have almost any relationship with the ungrouped regression function. Standard estimation procedures are no longer consistent and require considerable modification.

We thank Eugene Choo and Joel Horowitz for helpful discussions and two referees for comments. We thank Chang-sik Kim for excellent research assistance and the National Science Foundation and the Korean Research Foundation for financial support. We also thank Joel Horowitz for generously providing his GAUSS programs for computing the deconvolution part of our estimators. The second author thanks the Cowles Foundation for its generous hospitality during his recent visits. Address correspondence to: Oliver Linton, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom; e-mail: lintono@lse.ac.uk; and Yoon-Jae Whang, Department of Economics, Ewha University, Seoul, 120-750, Korea; e-mail: whang@mm.ewha.ac.kr.

  0266-4666/02 $9.50

We propose methods for estimating a nonparametric regression function and nonparametric density function based on aggregated data. We allow for a within "family" component but assume that the data are independent across families. Our estimators are based on the deconvolution methods of Fan (1991, 1992), Fan and Masry (1992), Fan and Truong (1993), Masry (1991, 1993), and Stefanski and Carroll (1990). See also Horowitz and Markatou (1996) and Horowitz (1998) for an application of these ideas. We establish consistency and asymptotic normality of our methods. The rate of convergence depends on the details of the decay rate of the characteristic function of the data and can be very slow indeed. The motivation for our work was a term paper by a Yale Ph.D. student, Eugene Choo (1998), who estimated a hedonic pricing model for slaves sold in auction in the pre-bellum south. The slaves were sold in job lots sometimes family related, sometimes characteristic related, sometimes more or less randomly composed. The observed price was the price of the lot rather than of the individual. It was of interest to back out the individual price/characteristic relationship from these aggregated data. Our particular interest is to do this without making strong assumptions about the functional form of the latent distribution.

In Section 2 we describe the model and our estimator. In Section 3 we give the asymptotic properties of our estimators in the two leading cases concerning the behavior of the characteristic function. In Section 4 we briefly discuss some practical issues, and in Section 5 we give the results of some simulations. The Appendix contains our proofs. We use $\Rightarrow$ to denote convergence in distribution and $\xrightarrow{p}$ to denote convergence in probability. Let $\|A\| = \mathrm{tr}(A^T A)^{1/2}$ for any matrix A. Also, define the complex-valued quantity $\mathbf{i} = \sqrt{-1}$.

2. MODEL SPECIFICATION AND ESTIMATION

We suppose that the data are organized into family units or batches, i.e., $\{(Y_{i_j}, X_{i_j}) : i = 1,\dots,n;\ j = 1,\dots,r_i\}$. We also suppose that there is a common element to the data series, which we model using the one-factor structure

$$Y_{i_j} = Y_{0i_j} + \eta_i; \qquad X_{i_j} = X_{0i_j} + \varepsilon_i, \tag{1}$$

where (Y_{0i_j}, X_{0i_j}) and (η_i, ε_i) are independent and identically distributed (i.i.d.) across both i and j and (η_i, ε_i) are independent of $\{(Y_{0i_j}, X_{0i_j}),\ j = 1,\dots r_i\}$. Here, r_i is a positive integer perhaps random but independent of all other random variables. The variables (Y_{0i_j}, X_{0i_j}) represent idiosyncratic components, whereas (η_i, ε_i) are common to all members of "the family." The common effect induces dependence across j within the same i, but observations across i are mutually independent. The assumption that the idiosyncratic components are independent is quite strong and implies, e.g., that $E(Y_{0i_j} | X_{0i_1}, \dots, X_{0i_{r_i}}) = E(Y_{0i_j} | X_{0i_j})$, although it should be noted that this still allows for $E(Y_{i_j} | X_{i_1}, \dots, X_{i_{r_i}}) \neq E(Y_{i_j} | X_{i_j})$. We are going to be primarily interested in the marginal effect $E(Y_{i_j} | X_{i_j})$, because under the aggregation rule introduced subsequently the quantity $E(Y_{i_j}|X_{i_1},\ldots,X_{i_{r_i}})$ is unidentified. The common family component can be more or less important depending upon the data. Certainly, when the units are aggregated in a more or less random way, this common effect may be taken as small. This structure is used in many fields of economics and finance. It can easily be extended to allow for multiple factors to the extent that family size permits.

We further suppose that we only observe the grouped or aggregated data

$$\overline{Y}_i = \sum_{j=1}^{r_i} Y_{i_j}; \quad \overline{X}_i = \sum_{j=1}^{r_i} X_{i_j}, \quad i = 1,\ldots,n. \tag{2}$$

This kind of observation rule arises quite often in household surveys where much information is obtained only at the household level; see Chesher (1997) and Choo (1998) for recent examples. Note that this sort of grouping is different from that considered in Amemiya (1985, p. 275) where there are a small number of "families" of large size; we have a large number of families of small size. In many data sets, the "family size" r_i is not the same across units. Nevertheless, the number of different family sizes is small relative to the total number of units. We shall suppose that $r_i \in \{r_1,\ldots,r_R$, some finite integer $R\}$ and that the number of families of each distinct size r_ℓ, denoted n_ℓ, is large, whereas the family sizes themselves are relatively small (we have $\sum_{\ell=1}^{R} n_\ell = n$ with R fixed and $n_\ell \to \infty$ for all ℓ in the asymptotics). We shall further assume that the aggregation is not systematically related to the data distribution itself. To allow for such possibilities requires a model of the relationship between, say, household size and the covariates, which is beyond the scope of this paper.

Subsequently, for notational simplicity, we sometimes denote $(Y_{i_j}, X_{i_j}, Y_{0i_j}, X_{0i_j}, \overline{Y}_i, \overline{X}_i, r_i, n_i)$ as $(Y, X, Y_0, X_0, \overline{Y}, \overline{X}, r, n)$. We shall stratify according to family size and do our calculations on the homogeneous units to obtain consistent estimates. We wish to estimate quantities such as the marginal density $f_X(\cdot)$ and joint density $f_{Y,X}(\cdot)$ of the individual data (Y,X), the regression function

$$E(Y|X = x) = m(x), \tag{3}$$

or various functionals from the conditional distribution of Y given X using the available sample $\{(\overline{Y}_i, \overline{X}_i) : i = 1,\ldots,n\}$ and without imposing functional form restrictions on $f_{Y,X}(\cdot)$. If $m(x) = \alpha + \beta x$, then, $E(\overline{Y}|\overline{X} = x) = r\alpha + \beta x$; i.e., the grouped data regression function is essentially the same as the ungrouped regression. In general, this correspondence is not present, and we must use more sophisticated techniques to extract the ungrouped distribution from the grouped data.

Note that

$$E(Y|X = x) = \frac{g_X(x)}{f_X(x)}, \quad \text{where} \tag{4}$$

$$g_X(x) = \int y f_{Y,X}(y,x)\,dy. \tag{5}$$

Let $\phi_{X_0}(t) = E[\exp(\mathbf{i}tX_0)]$, $\phi_X(t) = E[\exp(\mathbf{i}tX)]$, $\phi_{\overline{X}}(t) = E[\exp(\mathbf{i}t\overline{X})]$, and $\phi_\varepsilon(t) = E[\exp(\mathbf{i}t\varepsilon)]$ denote the characteristic functions. Expressions (1) and (2) imply that

$$\phi_X(t) = \phi_{X_0}(t)\phi_\varepsilon(t), \tag{6}$$

$$\phi_{\overline{X}}(t) = [\phi_{X_0}(t)]^r\phi_\varepsilon(rt) \tag{7}$$

by the convolution theorem. Similarly, letting $\phi_{Y_0,X_0}(s,t) = E[\exp(\mathbf{i}(sY_0 + tX_0))]$, $\phi_{Y,X}(s,t) = E[\exp(\mathbf{i}(sY + tX))]$, $\phi_{\overline{Y},\overline{X}}(s,t) = E[\exp(\mathbf{i}(s\overline{Y} + t\overline{X}))]$, and $\phi_{\eta,\varepsilon}(s,t) = E[\exp(\mathbf{i}(s\eta + t\varepsilon))]$, we have

$$\phi_{Y,X}(s,t) = \phi_{Y_0,X_0}(s,t)\phi_{\eta,\varepsilon}(s,t), \tag{8}$$

$$\phi_{\overline{Y},\overline{X}}(s,t) = [\phi_{Y,X}(s,t)]^r\phi_{\eta,\varepsilon}(rs,rt). \tag{9}$$

If we knew $\phi_\varepsilon(t)$ and $\phi_{\eta,\varepsilon}(s,t)$, then we would obtain the useful relations

$$\phi_X(t) = \left[\frac{\phi_{\overline{X}}(t)}{\phi_\varepsilon(rt)}\right]^{1/r}\phi_\varepsilon(t),$$

$$\phi_{Y,X}(s,t) = \left[\frac{\phi_{\overline{Y},\overline{X}}(t)}{\phi_{\eta,\varepsilon}(rs,rt)}\right]^{1/r}\phi_{\eta,\varepsilon}(s,t),$$

which determine $\phi_X(t)$ and $\phi_{Y,X}(s,t)$. The trick is really how to eliminate the nuisance functions $\phi_\varepsilon(t)$ and $\phi_{\eta,\varepsilon}(s,t)$. We show how to do this in the next section by using two different family size data sets. Suppose for now that we have estimators $\hat{\phi}_\varepsilon(t)$ and $\hat{\phi}_{\eta,\varepsilon}(s,t)$.

We can estimate the characteristic functions of the grouped data by the empirical characteristic functions

$$\hat{\phi}_{\overline{X}}(t) = \frac{1}{n}\sum_{j=1}^{n}\exp(\mathbf{i}t\overline{X}_j), \tag{10}$$

$$\hat{\phi}_{\overline{Y},\overline{X}}(s,t) = \frac{1}{n}\sum_{j=1}^{n}\exp(\mathbf{i}(s\overline{Y}_j + t\overline{X}_j)), \tag{11}$$

and hence

$$\hat{\phi}_X(t) = \left[\frac{\hat{\phi}_{\overline{X}}(t)}{\hat{\phi}_\varepsilon(rt)}\right]^{1/r}\hat{\phi}_\varepsilon(t), \tag{12}$$

$$\hat{\phi}_{Y,X}(s,t) = \left[\frac{\hat{\phi}_{\overline{Y},\overline{X}}(t)}{\hat{\phi}_{\eta,\varepsilon}(rs,rt)}\right]^{1/r}\hat{\phi}_{\eta,\varepsilon}(s,t). \tag{13}$$

We then apply deconvolution to these to obtain the density estimators

$$\hat{f}_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathbf{i}tx)\phi_K(th)\hat{\phi}_X(t)dt, \tag{14}$$

$$\hat{f}_{Y,X}(y,x) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy+tx))\tilde{\phi}_K(sh,th)\hat{\phi}_{Y,X}(s,t)dsdt, \tag{15}$$

where $\phi_K(\cdot)$ and $\tilde{\phi}_K(\cdot,\cdot)$ are the Fourier transforms of the kernels $K(\cdot)$ and $\tilde{K}(\cdot,\cdot)$, respectively, and h is a bandwidth sequence tending to zero with sample size n. Finally, we estimate $m(x) = E(Y|X=x)$ by

$$\hat{m}(x) = \frac{\hat{g}_X(x)}{\hat{f}_X(x)}, \quad \text{where} \tag{16}$$

$$\hat{g}_X(x) = \int y\hat{f}_{Y,X}(y,x)dy. \tag{17}$$

In practice, equations (14)–(17) can be complex, so we shall take the real part only (the imaginary parts are typically small and converge to zero in probability).

Remarks.

1. For each different family size r we have estimates of the desired quantities. One can then aggregate the estimates to improve efficiency, e.g., by minimum distance. Let $\hat{m}_r(x)$ be the estimate of $m(x)$ based on families of size r, where r takes R different values. Then let $\tilde{m}(x)$ be the value of θ that minimizes the quadratic form $(\underline{\hat{m}} - \theta e)^T V(\underline{\hat{m}} - \theta e)$, where $\underline{\hat{m}} = (\hat{m}_{r_1}(x),\dots,\hat{m}_{r_R}(x))^T$ and $e = (1,\dots,1)^T$, whereas V is some positive definite weighting matrix. The explicit representation of $\tilde{m}(x)$ is

   $$\tilde{m}(x) = (e^T V e)^{-1} e^T V \underline{\hat{m}}.$$

   By choosing V to be the inverse of the asymptotic variance of the unrestricted estimator the resulting estimator has minimal variance within this class of estimators. However, the effect on bias is uncertain, and this estimator may even do worse according to mean squared error for some data distributions.
2. In some data sets, some of the variables are observed ungrouped. The ungrouped regression model of interest is $Y_{i_j} = m(X_{i_j}) + u_{i_j}$ for error term u_{i_j} that satisfies $E(u_{i_j}|X_{i_j}) = 0$. Suppose that X_{i_j}, $j = 1,\dots,r$ are observed but only the grouped $\bar{Y}_i$ data are observed. Then we have

   $$\bar{Y}_i = \sum_{j=1}^{r} m(X_{i_j}) + \bar{u}_i, \tag{18}$$

   where $\bar{u}_i = \sum_{j=1}^{r} u_{i_j}$. If also $E(u_{i_j}|X_{i_l}) = 0$ for $l \neq j$, then this is a standard additive nonparametric regression model with the additional constraint that the function m is the same across j. One could estimate the regression function by backfitting or marginal integration as described in Linton and Nielsen (1995) and Mammen, Linton, and Nielsen (1999) or by series estimation (see Andrews and Whang, 1990), which has the important feature that it involves no Fourier inversion. It can be expected that the rate of convergence of these estimators would be the same as that of one-dimensional nonparametric regression, which would be faster than we are able to obtain in our setting. Even when r varies substantially with i, one can still do better than the Fourier inversion method by using the recently developed methods of Linton, Nielsen, Tanggaard, and Mammen (1998) for estimating yield curves.

When $Y_{i_j}, j = 1, \ldots, r$ are observed, but only the grouped $\overline{X}_i$ data are observed, it does not seem possible to obtain a method that bypasses the Fourier inversion, and we seem stuck with the slow rate of convergence in this case too. This is likely to be the case also where some of the covariates are grouped and some are not.

3. Given estimates of $\phi_{\eta,\varepsilon}(s,t)$ one can obtain estimates of $\phi_{Y_0,X_0}(s,t)$ from (8) and hence of the regression function $E(Y_{0i_j}|X_{0i_j})$. We do not present results for this estimation, but no doubt they can be arrived at by minor modification of our theorems.

2.1. Estimation of ϕ_ε and $\phi_{\eta,\varepsilon}$

We give two alternative methods for estimating the error characteristic functions. The first method is suggested by work of Horowitz and Markatou (1996) and does not require functional form restrictions. The second method is based on a semiparametric restriction on the distribution of X, namely, that the distribution of the errors ε, η is parametric. For simplicity we just describe the methods for the problem of estimating ϕ_ε, but similar comments apply to the estimation of $\phi_{\eta,\varepsilon}$. A necessary condition for nonparametric identification of these distributions is that there are at least two distinct family sizes.

Suppose that there are at least two distinct family sizes; call them r_1 and r_2. Then, we have

$$P(t; r_1, r_2) = \frac{[\phi_{\overline{X}, r_1}(t)]^{1/r_1}}{[\phi_{\overline{X}, r_2}(t)]^{1/r_2}} = \frac{[\phi_\varepsilon(r_1 t)]^{1/r_1}}{[\phi_\varepsilon(r_2 t)]^{1/r_2}},$$

where $\phi_{\overline{X}, r_1}(t)$ denotes the characteristic function of $\overline{X}$ from families of size r_1 and likewise $\phi_{\overline{X}, r_2}(t)$. The left-hand side can be consistently estimated at rate root-n, at least for some range of t, by the empirical version of P, which we call P_n. Now suppose that ε is symmetrically distributed about zero, in which case ϕ_ε is real-valued. Then we can write

$$\ln P_n(t; r_1, r_2) \simeq \frac{1}{r_1} \kappa_\varepsilon(r_1 t) - \frac{1}{r_2} \kappa_\varepsilon(r_2 t) + u_n(t; r_1, r_2),$$

where

$$u_n(t; r_1, r_2) = \frac{P_n(t; r_1, r_2) - P(t; r_1, r_2)}{P(t; r_1, r_2)},$$

whereas $\kappa_\varepsilon(t) = \ln \phi_\varepsilon(t)$ is the cumulant generating function of ε. Now let

$$\hat{\phi}_\varepsilon(t) = \exp(\hat{\kappa}_\varepsilon(t)), \quad \hat{\kappa}_\varepsilon(t) = \sum_{j=2}^{J_n} \hat{a}_j t^j,$$

where J_n is some truncation sequence and the "parameters" $a_j, j = 1,\dots,J_n$ minimize the least squares criterion function

$$\sum_{\ell=1}^{L_n} \left\{ \ln P_n(t_\ell; r_1, r_2) - \sum_{j=2}^{J_n} a_j (r_1^{j-1} - r_2^{j-1}) t_\ell^j \right\}^2,$$

where t_ℓ, $\ell = 1,\dots,L_n$ are a grid of points. We have imposed the restriction that $\kappa_\varepsilon(0) = \kappa_\varepsilon'(0) = 0$, the second of which follows from the symmetry assumption. This above procedure is similar to one proposed in Horowitz and Markatou (1996, pp. 162–163) and can be expected to be consistent at the usual rate of convergence of nonparametric smoothing methods (which is faster than the rate of convergence of our deconvolution estimators), provided J_n goes to infinity at a certain rate. The restriction to symmetric errors can also perhaps be relaxed as in Horowitz and Markatou (1996).

Instead suppose that the characteristic function of ε is known except for finite-dimensional vector θ_0, i.e., $\phi_\varepsilon(\cdot) = \phi_\varepsilon(\cdot, \theta_0)$, where the function $\phi_\varepsilon(\cdot, \theta_0)$ is smooth. In this case, one can compute $\hat{\theta}$ to minimize the criterion function

$$\sum_{\ell=1}^{L_n} [\ln P_n(t_\ell) - \pi(t_\ell, \theta)]^2,$$

where $\pi(t_k, \theta) = \kappa_\varepsilon(r_1 t; \theta)/r_1 - (1/r_2)\kappa_\varepsilon(r_2 t; \theta)/r_2$. See Beran and Millar (1994) and Knight and Satchell (1997) for discussion of similar methods. Under some regularity conditions, we can expect $\hat{\theta}$ to be root-n consistent and asymptotically normal.

3. ASYMPTOTIC PROPERTIES

In this section, we analyze the asymptotic properties of the nonparametric density estimator (14) of $f_X(x)$ and regression estimator (16) of $m(x)$. The properties depend crucially on the smoothness of the densities $f_X(x)$ and $f_{Y,X}(y,x)$. The smoothness of a density is related to the tail behavior of the characteristic function. That is, the faster the decay of the characteristic function, the smoother its corresponding density. Subsequently, we consider two types of characteristic functions: characteristic functions with *algebraic decay* and characteristic functions with *exponential decay*. In the literature, the former type is often referred to as the case of *ordinary smooth* distributions and includes gamma and Laplace distributions, whereas the latter type is referred to as that of *super smooth* distributions and includes normal and Cauchy distributions and their mixtures, among others. Our theoretical development is similar to that in Fan and Masry

(1992). The main technical difficulty we have is the nonlinear way in which $\phi_{\bar{X}}(t)$, e.g., enters into (14).

We shall assume a uniform rate of convergence of our estimators of $\phi_\varepsilon(t)$ and $\phi_{\eta,\varepsilon}(s,t)$, which can be expected to pertain under some regularity conditions as already discussed. We shall suppose that $n \to \infty$.

Assumption E1. There exists an estimator $\hat{\phi}_\varepsilon(t)$ such that for $j = 0,1,2,3$ we have

$$\sup_{t\in\mathbb{R}}\left|\frac{\partial^j}{\partial t^j}\,\hat{\phi}_\varepsilon(t) - \frac{\partial^j}{\partial t^j}\,\phi_\varepsilon(t)\right| = O_p(n^{-\alpha/2})$$

for some α with $0 < \alpha \leq 1$.

Assumption E2. There exists an estimator $\hat{\phi}_{\eta,\varepsilon}(s,t)$ such that for $j + k = 0,1,2,3$ we have

$$\sup_{(s,t)\in\mathbb{R}^2}\left|\frac{\partial^{j+k}}{\partial s^k t^j}\,\hat{\phi}_{\eta,\varepsilon}(s,t) - \frac{\partial^{j+k}}{\partial s^k t^j}\,\phi_{\eta,\varepsilon}(s,t)\right| = O_p(n^{-\alpha/2})$$

for some α with $0 < \alpha \leq 1$.

3.1. Case I: Characteristic Functions with Algebraic Decay

3.1.1. Density estimation

Assumption A.

(i) $\phi_{X_0}(t)t^{\beta_1} \to A_1$, $\phi_\varepsilon(t)t^{\beta_2} \to A_2$, $|\phi'_{X_0}(t)t^{\beta_1+1}| = O(1)$, and $|\phi'_\varepsilon(t)t^{\beta_2+1}| = O(1)$ as $t \to \infty$ for some constants $A_1 \neq 0$, $A_2 \neq 0$, $\beta_1 \geq 1$, and $\beta_2 \geq 1$ with $(r-1)\beta_1 > \frac{1}{2}$.
(ii) $\phi_{X_0}(t) \neq 0$ and $\phi_\varepsilon(t) \neq 0$ for all $t \in \mathbb{R}$.
(iii) $\phi_K(\cdot)$ is a symmetric function with $k+2$ bounded integrable derivatives, $\phi_K(0) = 1$, and $\phi_K(t) = 1 + O(|t|^k)$ as $t \to 0$ for some $k \geq 0$.
(iv) $\int_{-\infty}^{\infty}|\phi_K(t)||t|^{(2r-1)\beta_1}dt < \infty$, $\int_{-\infty}^{\infty}|\phi'_K(t)||t|^{(r-1)\beta_1}dt < \infty$, and $\int_{-\infty}^{\infty}|\phi_K(t)|^2 \times |t|^{2(r-1)\beta_1}dt < \infty$.
(v) $f_X(\cdot)$ is k-times continuously differentiable with bounded derivatives.

Remark. Assumption A(iii) implies that the kernel function

$$K(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tu)\phi_K(t)dt \tag{19}$$

is a real-valued function integrating to unity and kth order, i.e.,

$$\int_{-\infty}^{\infty}u^jK(u)du = 0 \quad \text{for } j = 1,\ldots,k-1, \qquad \int_{-\infty}^{\infty}|u^kK(u)|\,du < \infty.$$

Define

$$\sigma_{n1}^2(x) = n^{-1}h^{-2(r-1)\beta_1-1}\sigma_1^2(x), \quad \text{where} \tag{20}$$

$$\sigma_1^2(x) = \frac{f_{\overline{X}}(x)r^{2(\beta_2-1)}}{2\pi A_1^{2(r-1)}}\int_{-\infty}^{\infty}|\phi_K(t)|^2|t|^{2(r-1)\beta_1}dt. \tag{21}$$

Let

$$f_X^*(x) = \int_{-\infty}^{\infty} K(u)f_X(x-hu)du \tag{22}$$

be the convolution of K and f_X. The asymptotic normality of the density estimator is established in the following theorem.

THEOREM 1. *Under Assumptions A and E1,*

(a) if $nh^{\max\{2r\beta_1/\alpha,(2\beta_2+1)/\alpha,(2r\beta_1+2\beta_2+1)\}} \to \infty$ *and* $n^{1-\alpha}h^{2(r-1)\beta_1-1} \to 0$, *then*

$$\frac{\hat{f}_X(x) - f_X^*(x)}{\sigma_{n1}(x)} \Rightarrow N(0,1),$$

and

(b) if moreover $nh^{2(r-1)\beta_1+2k+1} \to 0$, *then*

$$\frac{\hat{f}_X(x) - f_X(x)}{\sigma_{n1}(x)} \Rightarrow N(0,1).$$

Remark. The term $f_X^*(x)$ can be expanded in a Taylor series expansion to give $f_X^*(x) = f_X(x) + O(h^k)$. The mean squared error of $\hat{f}_X(x)$ is thus $O(h^{2k}) + O(n^{-1}h^{-2(r-1)\beta_1-1})$; when $h \propto n^{-1/(2(r-1)\beta_1+2k+1)}$ this is $O(n^{-2k/(2(r-1)\beta_1+2k+1)})$.

Let

$$Z_{nj} = \frac{1}{h}G_n\left(\frac{x-\overline{X}_j}{h}\right) \quad \text{for } j = 1,\ldots,n, \tag{23}$$

where

$$G_n(x) = \frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\frac{\phi_K(t)\phi_\varepsilon(t/h)}{[\phi_{\overline{X}}(t/h)]^{(r-1)/r}[\phi_\varepsilon(rt/h)]^{1/r}}dt. \tag{24}$$

Because we can show that $\sigma_{n1}^2(x) = n^{-1}\text{var}(Z_{n1}) + o(1)$, we can estimate the asymptotic variance $\sigma_{n1}^2(x)$ consistently (in a relative sense) by

$$\hat{\sigma}_{n1}^2(x) = \frac{1}{n^2}\sum_{j=1}^{n}\{\hat{Z}_{nj} - \overline{\hat{Z}}_n\}^2, \tag{25}$$

where

$$\hat{Z}_{nj} = \frac{1}{h}\hat{G}_n\left(\frac{x-\bar{X}_j}{h}\right), \quad (26)$$

$$\bar{\hat{Z}}_n = \frac{1}{n}\sum_{j=1}^{n}\hat{Z}_{nj}, \quad \text{and} \quad (27)$$

$$\hat{G}_n(x) = \frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\,\frac{\phi_K(t)\hat{\phi}_\varepsilon(t/h)}{[\hat{\phi}_{\bar{X}}(t/h)]^{(r-1)/r}[\hat{\phi}_\varepsilon(rt/h)]^{1/r}}\,dt. \quad (28)$$

Consistency of $\hat{\sigma}^2_{n1}(x)$ is established in the following lemma.

LEMMA 2. *Under the assumptions of Theorem 1(a), if* $nh^{[(3r-2)\beta_1+\beta_2+2]/\alpha}\to\infty$, *then*

$$\frac{\hat{\sigma}^2_{n1}(x)}{\sigma^2_{n1}(x)}\xrightarrow{p}1.$$

Theorem 1 and Lemma 2 now combine to give the following corollary.

COROLLARY 3. *Under the assumptions of Theorem 1(b), if* $nh^{[(3r-2)\beta_1+\beta_2+2]/\alpha}\to\infty$, *then*

$$\frac{\hat{f}_X(x)-f_X(x)}{\hat{\sigma}_{n1}(x)}\Rightarrow N(0,1).$$

3.1.2. Regression estimation. For simplicity of presentation, we take the kernel function $\tilde{K}(u,v)$ to be the product kernel $K(u)K(v)$, which implies

$$\tilde{\phi}_K(s,t) = \phi_K(s)\phi_K(t). \quad (29)$$

(In treating the case of characteristic functions with exponential decay, however, we find the expression of the general kernel $\tilde{K}(u,v)$ is more convenient to deal with.)

Let $f_{\bar{X}}(\cdot)$ and $f_{\bar{Y},\bar{X}}(y,x)$ be the marginal and joint densities of $\bar{X}$ and $(\bar{Y},\bar{X})$, respectively, and let $\|(s,t)\| = \sqrt{s^2+t^2}$. Define also

$$v_{\bar{X}}(x) = E(\bar{Y}^2|\bar{X}=x). \quad (30)$$

Assumption B.

(i) $\phi_{Y_0,X_0}(s,t)\|(s,t)\|^{\rho_1}\to B_1$, $\phi_{\eta,\varepsilon}(s,t)\|(s,t)\|^{\rho_2}\to B_2$, $|\partial^j\phi_{Y_0,X_0}(s,t)/\partial s^j|\times\|(s,t)\|^{\rho_1+1} = O(1)$ and $|\partial^j\phi_{\eta,\varepsilon}(s,t)/\partial s^j|\|(s,t)\|^{\rho_2+1} = O(1)$ for $j=1,2$, and 3 as $\|(s,t)\|\to\infty$ for some constants $B_1\neq 0$, $B_2\neq 0$, $\rho_1\geq 1$, and $\rho_2\geq 1$ with $(r-1)\rho_1>\frac{3}{2}$.

(ii) $\phi_{Y_0,X_0}(s,t)\neq 0$ and $\phi_{\eta,\varepsilon}(s,t)\neq 0$ for all $(s,t)\in\mathbb{R}^2$.

(iii) $\phi_K(\cdot)$ is a symmetric function with $k+2$ bounded integrable derivatives, $\phi_K(0)=1$, and $\phi_K(t) = 1+O(|t|^k)$ as $t\to 0$ for some $k\geq 0$.

(iv) $\int_{-\infty}^{\infty} |\partial^j \phi_K(t)/\partial t^j||t|^{(2r-1)\rho_1+\rho_2} dt < \infty$ for $j = 0,1,2,$ and 3.
(v) $v_{\overline{X}}(\cdot)$ is continuous at x.
(vi) $g_X(\cdot)$ is integrable and $g_X(\cdot)$ and $f_X(\cdot)$ are both k-times differentiable with bounded continuous kth derivatives.
(vii) $EY_0^6 < \infty$ and $E\eta^6 < \infty$.

Define

$$\sigma_{n2}^2(x) = \frac{\sigma_2^2(x)}{nh^{2(r-1)\rho_1+1} f_X^2(x)}, \tag{31}$$

where

$$\sigma_2^2(x) = \frac{v_{\overline{X}}(x) f_{\overline{X}}(x) r^{2(\rho_2-1)}}{(2\pi)^4 B_1^{2(r-1)}} \times \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy+tx)) \phi_K(s) \phi_K(t) \|(s,t)\|^{(r-1)\rho_1} ds dt dy \right]^2 dx. \tag{32}$$

Let

$$R_n(x) = \frac{R_{n1}^*(x) - R_{n2}^*(x)}{\hat{f}_X(x)}, \tag{33}$$

where

$$R_{n1}^*(x) = m^*(x) - m(x), \tag{34}$$

$$R_{n2}^*(x) = [f_X^*(x) - f_X(x)] m(x), \tag{35}$$

$$m^*(x) = \int_{-\infty}^{\infty} g_X(x-hu) f_X(x-hu) K(u) du, \tag{36}$$

and $f_X^*(x)$ is as defined in (22).

The asymptotic normality of the regression estimator is established in the following theorem.

THEOREM 4. *Under Assumptions E1, E2, A(i) and (ii), and B with* $\rho_1 > \beta_1$,

(a) if $nh^{\max\{2r\rho_1/\alpha, (2\rho_2+3)/\alpha, 2r\rho_1+2\rho_2+3\}} \to \infty$ *and* $n^{1-\alpha} h^{2(r-1)\rho_1-3} \to 0$, *then*

$$\frac{\hat{m}(x) - m(x) - R_n(x)}{\sigma_{n2}(x)} \Rightarrow N(0,1),$$

and

(b) if moreover $nh^{2(r-1)\rho_1+2k+1} \to 0$, *then*

$$\frac{\hat{m}(x) - m(x)}{\sigma_{n2}(x)} \Rightarrow N(0,1).$$

Remark. The convergence rate is similar to that in the density estimation case.

For $j = 1, \ldots, n$, let

$$Z_{nj} = \frac{1}{h^2}\int_{-\infty}^{\infty} y G_n\left(\frac{y-\overline{Y}_j}{h}, \frac{x-\overline{X}_j}{h}\right) dy$$
$$= \overline{Y}_j \frac{1}{h} K_{n1}\left(\frac{x-\overline{X}_j}{h}\right) + K_{n2}\left(\frac{x-\overline{X}_j}{h}\right), \quad \textbf{(37)}$$

where

$$K_{n1}(x) = \int_{-\infty}^{\infty} G_n(y,x)dy, \quad \textbf{(38)}$$

$$K_{n2}(x) = \int_{-\infty}^{\infty} yG_n(y,x)dy, \quad \text{and} \quad \textbf{(39)}$$

$$G_n(y,x) = \frac{1}{(2\pi)^2 r}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy+tx)) \times \frac{\tilde{\phi}_K(s,t)\phi_{\eta,\varepsilon}\left(\frac{s}{h},\frac{t}{h}\right)}{\left[\phi_{\overline{Y},\overline{X}}\left(\frac{s}{h},\frac{t}{h}\right)\right]^{(r-1)/r}\left[\phi_{\eta,\varepsilon}\left(\frac{rs}{h},\frac{rt}{h}\right)\right]^{1/r}} dsdt. \quad \textbf{(40)}$$

Because $\sigma_{n2}^2(x) = n^{-1}\text{var}(Z_{n1}) + o(1)$, we can estimate $\sigma_{n2}^2(x)$ consistently by

$$\hat{\sigma}_{n2}^2(x) = \frac{1}{n^2}\sum_{j=1}^{n}\{\hat{Z}_{nj} - \bar{\hat{Z}}_n\}^2, \quad \textbf{(41)}$$

where

$$\hat{Z}_{nj} = \frac{1}{h^2}\int_{-\infty}^{\infty} y\hat{G}_n\left(\frac{y-\overline{Y}_j}{h}, \frac{x-\overline{X}_j}{h}\right) dy \quad \text{with} \quad \textbf{(42)}$$

$$\hat{G}_n(y,x) = \frac{1}{(2\pi)^2 r}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy+tx)) \times \frac{\tilde{\phi}_K(s,t)\hat{\phi}_{\eta,\varepsilon}\left(\frac{s}{h},\frac{t}{h}\right)}{\left[\hat{\phi}_{\overline{Y},\overline{X}}\left(\frac{s}{h},\frac{t}{h}\right)\right]^{(r-1)/r}\left[\hat{\phi}_{\eta,\varepsilon}\left(\frac{rs}{h},\frac{rt}{h}\right)\right]^{1/r}} dsdt. \quad \textbf{(43)}$$

LEMMA 5. *Under the assumptions of Theorem 4(a), if* $nh^{[(3r-2)\rho_1+\rho_2+4]/\alpha} \to \infty$, *then*

$$\frac{\hat{\sigma}_{n2}^2(x)}{\sigma_{n2}^2(x)} \xrightarrow{p} 1.$$

Combining Theorem 4 and Lemma 5, we have the following corollary.

COROLLARY 6. *Under the assumptions of Theorem 4(b), if* $nh^{[(3r-2)\rho_1+\rho_2+4]/\alpha} \to \infty$, *then*

$$\frac{\hat{m}(x) - m(x)}{\hat{\sigma}_{n2}(x)} \Rightarrow N(0,1).$$

3.2. Case II: Characteristic Functions with Exponential Decay

We next consider the case in which the tail of the characteristic function decays exponentially fast.

3.2.1. Density estimation.

Assumption C.

(i) $A_0|t|^{\beta_0}\exp(-a_0|t|^\beta) \leq |\phi_{X_0}(t)| \leq B_0|t|^{\beta_0}\exp(-a_0|t|^\beta)$ and $A_1|t|^{\beta_1} \times \exp(-a_1|t|^\beta) \leq |\phi_\varepsilon(t)| \leq B_1|t|^{\beta_1}\exp(-a_1|t|^\beta)$ as $|t| \to \infty$ for some positive constants $a_0, a_1, \beta, A_0, B_0, A_1$, and B_1 and constants β_0 and β_1.
(ii) $\phi_{X_0}(t) \neq 0$ and $\phi_\varepsilon(t) \neq 0$ for all $t \in \mathbb{R}$.
(iii) $\phi_K(t)$ has a finite support $(-d, d)$.
(iv) There exist positive constants δ, B_2, and l such that $|\phi_K(t)| \leq B_2(d-t)^l$ for $t \in (d-\delta, d)$.
(v) $\phi_K(t) \geq B_3(d-t)^l$ for $t \in (d-\delta, d)$, where B_3 is a positive constant.
(vi) Either $\tilde{I}(t) = o(\tilde{R}(t))$ or $\tilde{R}(t) = o(\tilde{I}(t))$ as $t \to \infty$, where $\tilde{R}(t)$ and $\tilde{I}(t)$ are real and imaginary parts of $[\phi_{X_0}(t)]^{r-1}\phi_\varepsilon(rt)/\phi_\varepsilon(t)$, respectively.

Remark. Assumption C(i) assumes that the density functions of X_0 and ε are super smooth. It implies that the density functions are bounded and have bounded derivatives of all orders. Assumption C(iv) describes the behavior of $\phi_K(t)$ in the neighborhood of $t = d$. Assumptions C(v) and (vi) are used to develop lower bounds. Assumption C(vi) indicates that, at the tail, the characteristic function $[\phi_X(t)]^{r-1}\ \phi_\varepsilon(rt)/\phi_\varepsilon(t)$ is either purely real or purely imaginary.

Define

$$\sigma_{n3}^2(x) = n^{-1}\mathrm{var}(Z_{n1}), \tag{44}$$

where Z_{n1} is as defined in (23).

THEOREM 7. *Suppose Assumptions E1 and C hold and* $[a_0 r - a_1 r^{\beta}]\gamma + \alpha > \frac{1}{2}$. *If* $h = d(\gamma \ln n)^{-1/\beta}$ *for some* $0 < \gamma < \min\{\alpha/2a_1, (1-\alpha)/2a_0 r\}$, *then*

$$\frac{\hat{f}_X(x) - f_X^*(x)}{\sigma_{n3}(x)} \Rightarrow N(0,1).$$

Remarks.

1. As in the case of ordinary smooth distributions, the term $f_X^*(x)$ can be expanded in a Taylor series expansion to give $f_X^*(x) = f_X(x) + O(h^k)$. Using the result of Lemma 15(a) in the Appendix, the mean squared error of $\hat{f}_X(x)$ is thus

$$O(h^{2k}) + O(n^{-1} h^{2[\beta(l+1)+(r-1)\beta_0 - 1]} (\ln(1/h))^{2l} \times \exp[2\{a_0(r-1) + a_1(r^{\beta} - 1)\}(d/h)^{\beta}]).$$

   When $h = d(\gamma \ln n)^{-1/\beta}$, the rate of convergence is very sensitive to the value of γ; when γ is large, the bias is a negligible term compared to its variance; and, when γ is sufficiently small, the variance will be a small-order term in comparison to the bias. As in Fan (1991), we expect that the optimal rate of convergence in our case is also $O((\ln n)^{-c})$ for some $c > 0$, which is very slow for moderate sample sizes.
2. Contrary to Theorem 1(b), the asymptotic bias in Theorem 7 does not vanish even if h is sufficiently small as long as $\gamma < 1/(2a_0 r)$. The latter condition is needed to make the remainder term of the Taylor expansion asymptotically negligible; see equation (A.78) in the proof of Theorem 7 in the Appendix. For the desired result $(f_X^*(x) - f_X(x))/\sigma_{n3}(x) \xrightarrow{p} 0$; however, we need $\gamma > 1/(2a_0(r-1))$.

As an estimator of $\sigma_{n3}^2(x)$, we consider

$$\hat{\sigma}_{n3}^2(x) = \frac{1}{n^2} \sum_{j=1}^{n} \{\hat{Z}_{nj} - \bar{\hat{Z}}_n\}^2, \quad \textbf{(45)}$$

where $\hat{Z}_{nj}$ and $\bar{\hat{Z}}_n$ are as defined in (26) and (27), respectively. Consistency of $\hat{\sigma}_{n3}^2(x)$ is established in the following lemma.

LEMMA 8. *Under Assumptions E1 and C, if* $h = d(\gamma \ln n)^{-1/\beta}$ *for some* $0 < \gamma < (\alpha/2)[2a_0(r-1) + a_1\{(2r-1)r^{\beta-1} - 1 + r^{-1}\}]^{-1}$, *then*

$$\frac{\hat{\sigma}_{n3}^2(x)}{\sigma_{n3}^2(x)} \xrightarrow{p} 1.$$

Theorem 7 and Lemma 8 now combine to give the following corollary.

COROLLARY 9. *Under Assumptions E1 and C, if* $h = d(\gamma \ln n)^{-1/\beta}$ *for some* $0 < \gamma < (\alpha/2)[2a_0(r-1) + a_1\{(2r-1)r^{\beta-1} - 1 + r^{-1}\}]^{-1}$, *then*

$$\frac{\hat{f}_X(x) - f_X^*(x)}{\hat{\sigma}_{n3}(x)} \Rightarrow N(0,1).$$

3.2.2. Regression estimation.

Assumption D.

(i) $D_0\|(s,t)\|^{\rho_0}\exp(-b_0\|(s,t)\|^{\rho}) \le |\phi_{Y_0,X_0}(s,t)| \le E_0\|(s,t)\|^{\rho_0}\exp(-b_0\|(s,t)\|^{\rho})$ and $D_1\|(s,t)\|^{\rho_1}\exp(-b_1\|(s,t)\|^{\rho}) \le |\phi_{\eta,\varepsilon}(s,t)| \le E_1\|(s,t)\|^{\rho_1}\times \exp(-b_1\|(s,t)\|^{\rho})$ as $\|(s,t)\| \to \infty$ for some positive constants $b_0, b_1, \rho, D_0, D_1, E_0$, and E_1 and constants ρ_0 and ρ_1.
(ii) $\phi_{Y_0,X_0}(s,t) \ne 0$ and $\phi_{\eta,\varepsilon}(s,t) \ne 0$ for all $(s,t) \in \mathbb{R}^2$.
(iii) $\tilde{\phi}_K(s,t)$ has a finite support $\{(s,t) \in \mathbb{R}^2 : \|(s,t)\| < d\}$.
(iv) There exist positive constants δ, D_2, and m such that $|\tilde{\phi}_K(s,t)| \le D_2(d - \|(s,t)\|)^m$ for $\|(s,t)\| \in (d-\delta, d)$.
(v) $\tilde{\phi}_K(s,t) \ge D_3(d - \|(s,t)\|)^m$ for $\|(s,t)\| \in (d-\delta, d)$, where D_3 is a positive constant.
(vi) $\tilde{\phi}_K(s,t)$ is symmetric in (s,t); i.e., $\tilde{\phi}_K(s,t) = \tilde{\phi}_K(-s,t) = \tilde{\phi}_K(s,-t) = \tilde{\phi}_K(-s,-t)$.
(vii) Either $I^*(s,t) = o(R^*(s,t))$ or $R^*(s,t) = o(I^*(s,t))$ as $\|(s,t)\| \to \infty$, where $R^*(s,t)$ and $I^*(s,t)$ are real and imaginary parts of $[\phi_{Y_0,X_0}(s,t)]^{r-1}\phi_{\eta,\varepsilon}(rs,rt)/\phi_{\eta,\varepsilon}(s,t)$, respectively.
(viii) The support of $\overline{Y}$ (i.e., $\mathcal{Y}$) is bounded.

Remark. The boundedness of the support of $\overline{Y}$ can be restrictive in some cases. This assumption, however, simplifies the proof of Theorem 10, which follows; see the proof of Lemma 16(c) in the Appendix.

Let

$$Z_{nj} = \frac{1}{h^2}\int_{\mathcal{Y}} yG_n\left(\frac{y-\overline{Y}_j}{h}, \frac{x-\overline{X}_j}{h}\right)dy$$

$$= \overline{Y}_j\frac{1}{h}K_{n1}\left(\frac{x-\overline{X}_j}{h}\right) + K_{n2}\left(\frac{x-\overline{X}_j}{h}\right), \tag{46}$$

where

$$K_{n1}(x) = \int_{\mathcal{Y}} G_n(y,x)dy, \tag{47}$$

$$K_{n2}(x) = \int_{\mathcal{Y}} yG_n(y,x)dy, \tag{48}$$

and $G_n(\cdot,\cdot)$ are as defined in (38)–(40). Define

$$\sigma_{n4}^2(x) = n^{-1}\mathrm{var}(Z_{n1}), \tag{49}$$

where Z_{n1} is as defined in (46).

Let

$a^* = a_0(r-1) + a_1(r^\beta - 1)$ and

$b^* = b_0(r-1) + b_1(r^\rho - 1)$.

THEOREM 10. *Suppose Assumptions E1, E2, C, and D hold and* $\rho \geq \beta$, $b^* > a^*$, $[b_1 r^\rho - b_0 r]\gamma < \alpha - \frac{1}{2}$, $(a^* - b^* + a_1)\gamma < \alpha/2$, $(a^* - b^* + a_0 r)\gamma < \frac{1}{2}$, $(a^* - b^* + a_1 r^\beta - a_0 r)\gamma < \alpha - \frac{1}{2}$, $(a^* - b^* - a_0 r)\gamma < (\alpha - 1)/2$, *for some* $(1-\alpha)/2(b^* + b_0) < \gamma < 1/2b_0 r$. *If* $h = d(\gamma \log n)^{-1/\rho}$, *then*

$$\frac{\hat{m}(x) - m(x) - R_n(x)}{\sigma_{n4}(x)} \Rightarrow N(0,1).$$

The asymptotic variance $\sigma_{n4}^2(x)$ can be consistently estimated by

$$\hat{\sigma}_{n4}^2(x) = \frac{1}{n^2}\sum_{j=1}^{n}\{\hat{Z}_{nj} - \bar{\hat{Z}}_n\}^2, \tag{50}$$

where

$$\hat{Z}_{nj} = \frac{1}{h^2}\int_{\mathcal{Y}} y\hat{G}_n\left(\frac{y - \overline{Y}_j}{h}, \frac{x - \overline{X}_j}{h}\right) dy$$

with $\hat{G}_n(\cdot,\cdot)$ as defined in (43).

LEMMA 11. *Under Assumptions E1, E2, and D, if* $h = d(\gamma \log n)^{-1/\rho}$ *for some* $0 < \gamma < (\alpha/2)[2b_0(r-1) + b_1\{(2r-1)r^{\rho-1} - 1 + r^{-1}\}]^{-1}$, *then*

$$\frac{\hat{\sigma}_{n4}^2(x)}{\sigma_{n4}^2(x)} \xrightarrow{p} 1.$$

Combining Theorem 10 and Lemma 11, we have the following corollary.

COROLLARY 12. *Under the conditions of Theorem 10 and Lemma 11,*

$$\frac{\hat{m}(x) - m(x) - R_n(x)}{\hat{\sigma}_{n4}(x)} \Rightarrow N(0,1).$$

4. BANDWIDTH SELECTION

We have developed the theory necessary to conduct inference on the functions f_X and m in both ordinary smooth and super smooth cases. For practical application it is important to have some method for choosing the bandwidth parameter h, because this quantity determines the finite sample properties of our estimators. One method is based on estimating the integrated mean squared error; this requires consistent estimation of the derivatives of f_X and m, unless some parametric specification is adopted as in Silverman (1986). The alternative method of cross-validation, based on minimizing the sum of squared residuals from the leave-one-out version of $\hat{m}$, is very time consuming here. If one could find the equivalent penalty function to apply to the sum of squared residuals from the original $\hat{m}$, then this method might be feasible (for an exposition of the penalty function method in standard nonparametric regression; see Härdle, 1990). However, because our estimators are all nonlinear this situation is not covered by existing theory to our knowledge. In our simulations we have reported results for a range of bandwidth values; this is a popular approach in applied work. Nevertheless, the development of automatic bandwidth selection methods remains an important and interesting line of research to be pursued in the future.

5. MONTE CARLO

5.1. Design

We suppose that $X_{i_j} = X_{0i_j} + \varepsilon_i$, $Y_{0i_j} = \mu(X_{0i_j})$ for some function μ specified subsequently, and $Y_{i_j} = Y_{0i_j} + \eta_i$, where X_{0i_j}, ε_i, and η_i are mutually independent. Then, e.g.,

$$f_X(x) = \int p_\varepsilon(x-z)p_{X_0}(z)dz,$$

$$m(x) = E(Y_{i_j}|X_{i_j}=x) = E(\mu(X_{0i_j})|X_{i_j}=x)$$

$$= E(\mu(X_{0i_j})|X_{0i_j}+\varepsilon_i = x) = \frac{\int \mu(z)p_\varepsilon(x-z)p_{X_0}(z)dz}{\int p_\varepsilon(x-z)p_{X_0}(z)dz},$$

where $p_{X_0}(\cdot)$ and $p_\varepsilon(\cdot)$ are the densities of X_{0i_j} and ε_i, respectively. We use normal, uniform, and double exponential distributions for p_ε and for p_{X_0}, which combined with specifications for μ (we choose linear and quadratic functions, i.e., $\mu(x) = c_1 + c_2x$ and $\mu(x) = c_1 + c_2x + c_3x^2$ for some parameter values c_j) give the functions f and m, which are our focus. The calculations to obtain f, m are quite complicated to do by hand but have been obtained using a symbolic algebra package. In fact, with our parameter values, the resulting functions m are not far from the original function μ. More details are available upon request. In the normal case, X_{0i_j}, Y_{0i_j} are generated from $N(0,1)$ and ε_i, η_i are generated from $N(0,0.1)$. In the double exponential case, we generate X_{0i_j}, Y_{0i_j} with variance 0.5 and ε_i, η_i with variance 0.05. In the linear case we use $c_1 = 0, c_2 = 1$, whereas in the nonlinear case we use the same c_1, c_2, and take $c_3 = -0.1$. We have considered two different family sizes $r = 2,3$.

We use the product kernel $\tilde{K}(u,v) = K(u)K(v)$, which implies that $\tilde{\phi}_K(s,t) = \phi_K(s)\phi_K(t)$. We use two different kernel: the biquadratic and the normal. For bandwidth we have taken

$$h = c_h \times s_X n^{-1/12} \quad \text{and} \quad h = c_h \times s_X (\log n)^{-1/2}$$

in the case of ordinary smooth and super smooth densities, respectively, where s_X is the sample standard deviation of the variable X and c_h is a constant. We examine the performance of our method for a range of values for c_h.

We tried three different sample sizes $n = 100$, 250, 500 with 100 replications. We took 30 evaluation points in the interval $(-3,3)$. We calculated the truncated integrated mean squared error (IMSE) on this restricted range $(-3,3)$. Tables 1 and 2 show the IMSE of density estimates and regression function estimates in normal and double exponential cases. Figures 1–4 show 10 simulated density and regression function estimates.

5.2. Results

The graphs confirm that the estimated densities and regression functions are not far from the truth, but exhibit some variation in shape, especially in the end regions. We now turn to the IMSE results reported in our tables.

Density estimation works very well for any kind of distribution; we just show the normal and double exponential case, but the same is true also for the gamma, chi-square, exponential, and uniform cases, which are not shown here. IMSE decreases with sample size and is relatively insensitive to bandwidth in the range $0.3 \leq c_h \leq 0.4$. Decomposition of the IMSE into bias and variance (not shown) reveals that as expected squared bias increases with c_h, whereas variance decreases.

The regression function estimation appears to be somewhat more difficult, and performance depends more dramatically on bandwidth. Indeed for small bandwidths, the IMSE actually increases with sample size (this effect is more pronounced in the super smooth case). This is mostly a bias phenomenon—in fact very small bandwidths lead to big biases, which is contrary to our usual intuition. However, for larger bandwidths (e.g., when $c_h \geq 0.36$ in the super smooth case) the usual pattern reasserts itself. This is most likely a small sample phenomenon. The practical implications of this are that one should err on the side of larger bandwidths.

6. CONCLUSIONS AND EXTENSIONS

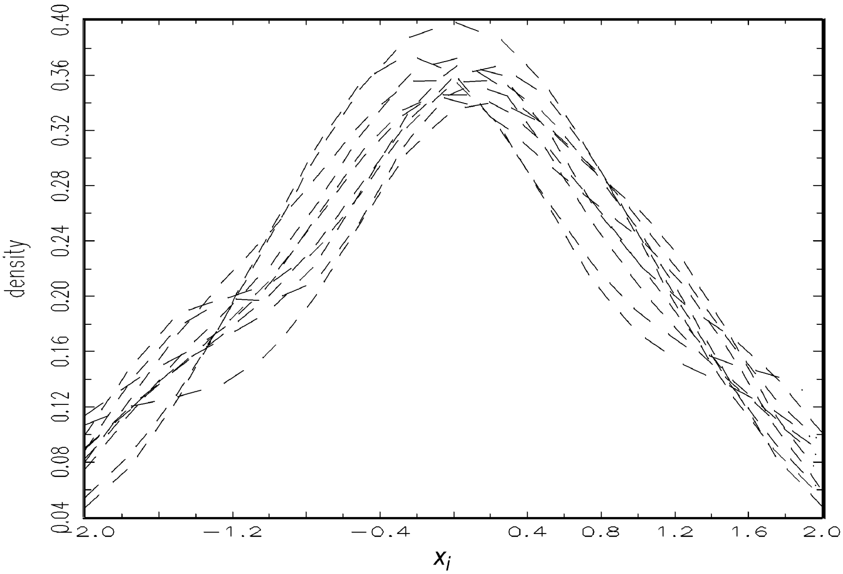
We have shown how to estimate the density and regression functions of individuals from aggregated data. Extensions to multiple covariates and to estimation of derivatives are straightforward. As Horowitz and Markatou (1996) point out, these methods are best applied to very large data sets. However, our sim-

TABLE 1. Normal (super smooth class)

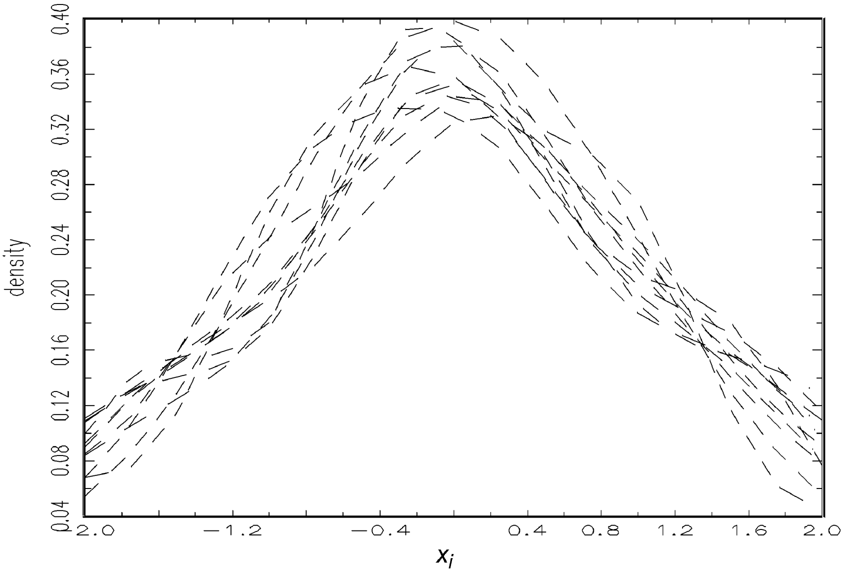
	Bandwidth(c_h)										
	0.30	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
	Truncated integrated mean squared error of $\hat{f}(x)$										
$n = 100$	0.0026	0.0027	0.0027	0.0029	0.0028	0.0028	0.0028	0.0030	0.0034	0.0034	0.0034
$n = 250$	0.0019	0.0021	0.0020	0.0021	0.0022	0.0023	0.0023	0.0026	0.0026	0.0027	0.0028
$n = 500$	0.0018	0.0017	0.0018	0.0018	0.0019	0.0021	0.0021	0.0021	0.0023	0.0023	0.0024
	Truncated integrated mean squared error of $\hat{m}(x)$; $\mu(x) = 1 + x$										
$n = 100$	0.7602	0.4844	0.2709	0.1763	0.1730	0.1978	0.1995	0.2268	0.2546	0.2617	0.2727
$n = 250$	2.9354	1.6461	0.6465	0.3121	0.1829	0.1312	0.1352	0.1500	0.1735	0.1918	0.2034
$n = 500$	4.7715	3.4107	2.0658	1.2680	0.4694	0.1598	0.0804	0.0793	0.1090	0.1379	0.1657
	Truncated integrated mean squared error of $\hat{m}(x)$; $\mu(x) = 1 + x + cx^2$										
$n = 100$	0.9998	0.5321	0.2635	0.2114	0.2313	0.2152	0.2275	0.2281	0.2740	0.2797	0.3100
$n = 250$	2.7636	1.4345	0.8243	0.4107	0.1434	0.1269	0.1445	0.1678	0.1891	0.2033	0.2186
$n = 500$	4.7713	3.9076	2.3675	1.1192	0.5128	0.2336	0.0871	0.0978	0.1324	0.1524	0.1804

TABLE 2. Double exponential (ordinary smooth class)

	Bandwidth(c_h)										
	0.30	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39	0.40
	Truncated integrated mean squared error of $\hat{f}(x)\cdot 10^2$										
$n = 100$	0.20	0.18	0.19	0.18	0.18	0.19	0.17	0.19	0.20	0.21	0.21
$n = 250$	0.18	0.17	0.16	0.16	0.15	0.16	0.16	0.16	0.16	0.17	0.17
$n = 500$	0.18	0.18	0.16	0.16	0.15	0.14	0.14	0.14	0.15	0.15	0.15
	Truncated integrated mean squared error of $\hat{m}(x)$; $\mu(x) = 1 + x$										
$n = 100$	0.4476	0.3142	0.3786	0.4084	0.4128	0.4469	0.4716	0.5084	0.5226	0.5369	0.5501
$n = 250$	0.4850	0.2417	0.2120	0.2886	0.3533	0.3768	0.4118	0.4509	0.4765	0.4730	0.4917
$n = 500$	1.0678	0.5247	0.2184	0.1247	0.1848	0.2882	0.3443	0.3958	0.4190	0.4490	0.4606
	Truncated integrated mean squared error of $\hat{m}(x)$; $\mu(x) = 1 + x + cx^2$										
$n = 100$	0.4463	0.3390	0.3948	0.4411	0.4303	0.4804	0.5023	0.5377	0.5484	0.5676	0.5795
$n = 250$	0.5046	0.2420	0.2232	0.3104	0.3714	0.4001	0.4327	0.4696	0.5067	0.5001	0.5201
$n = 500$	1.2494	0.4271	0.1937	0.1744	0.2379	0.2903	0.3694	0.4184	0.4477	0.4643	0.4890

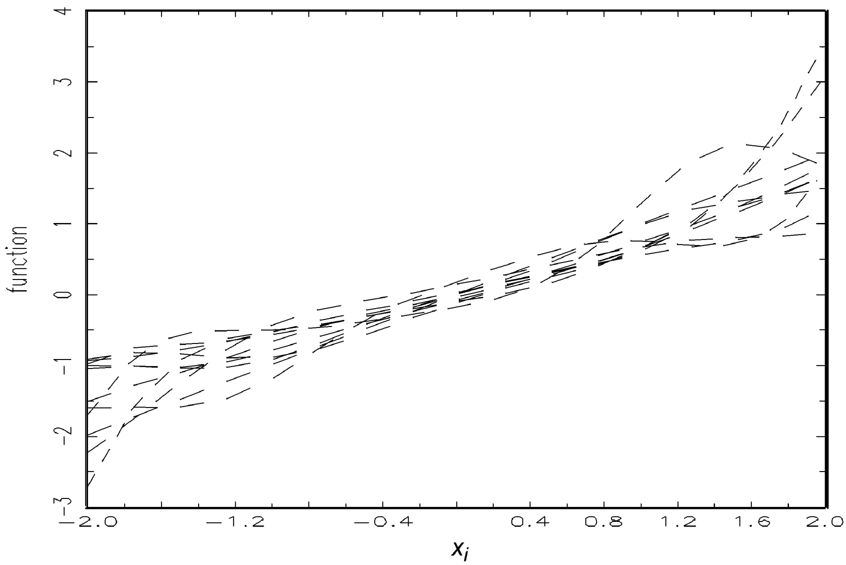


(a) n = 250

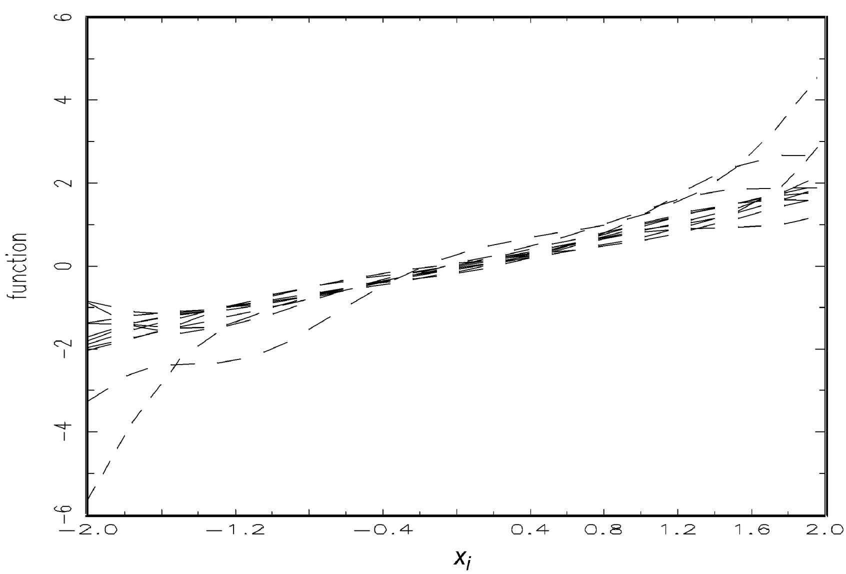


(b) n = 500

FIGURE 1. Density estimates, normal distribution.

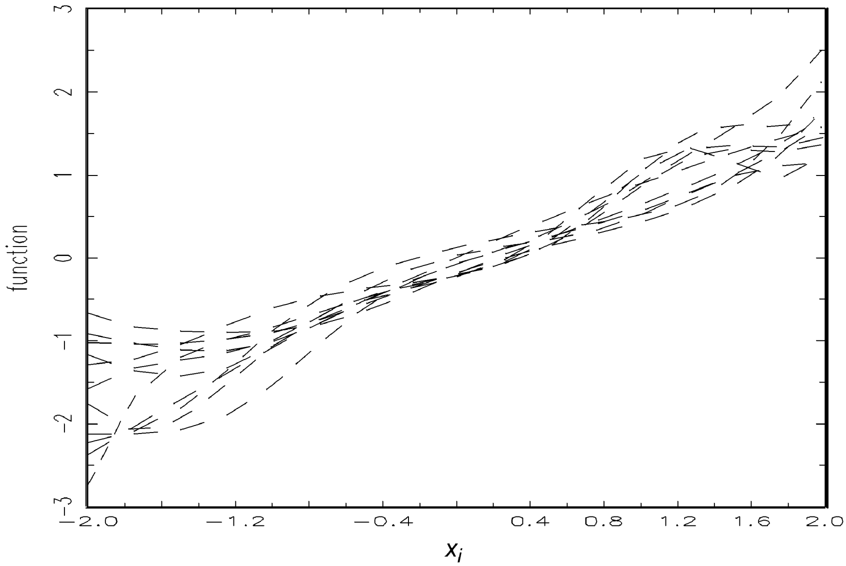


(a) n = 250

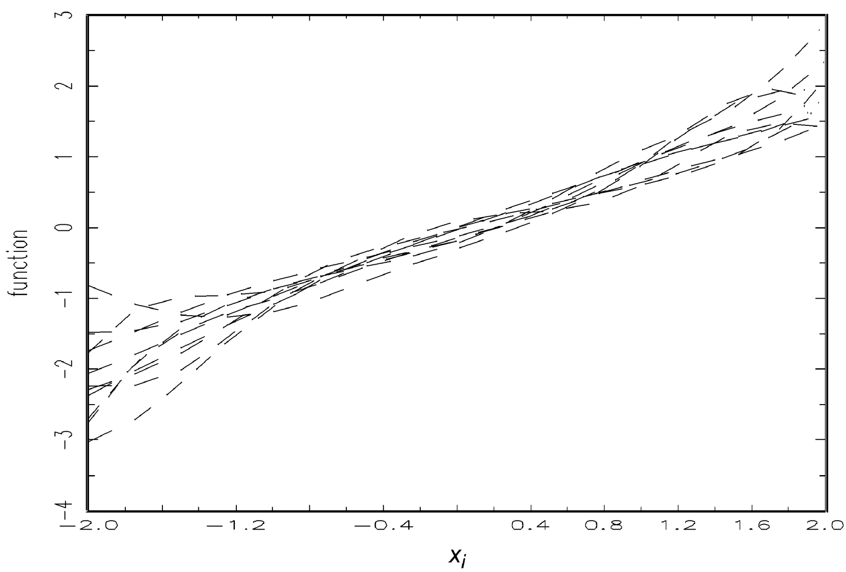


(b) n = 500

Figure 2. (a) and (b) Linear function, normal distribution.

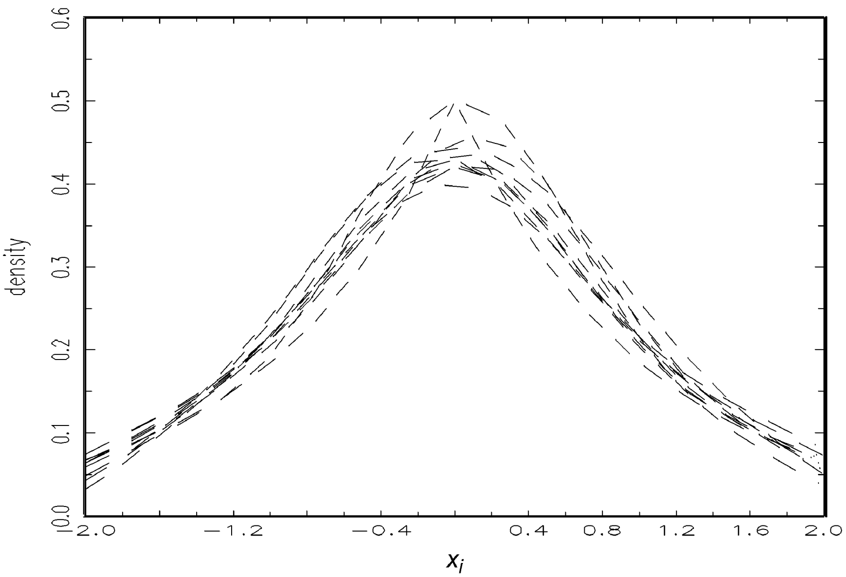


(c) n = 250

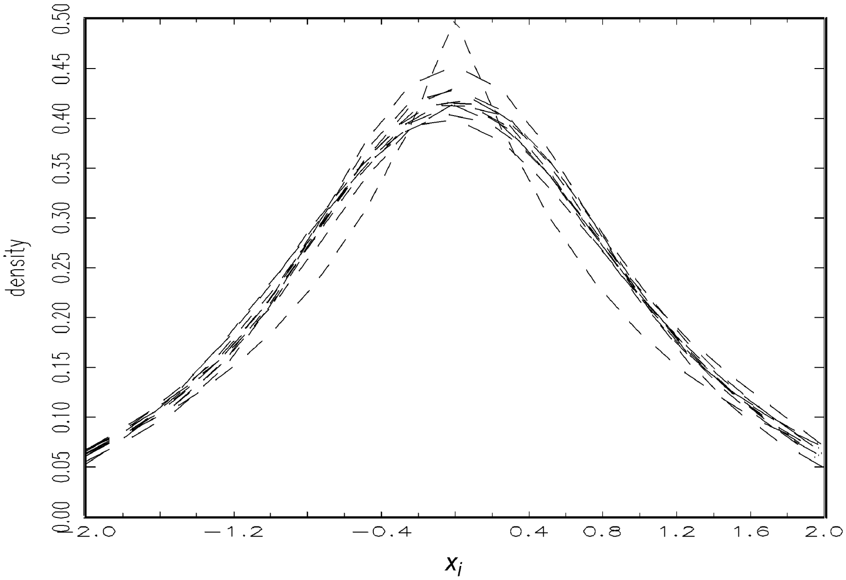


(d) n = 500

FIGURE 2. (c) and (d) nonlinear function, normal distribution.

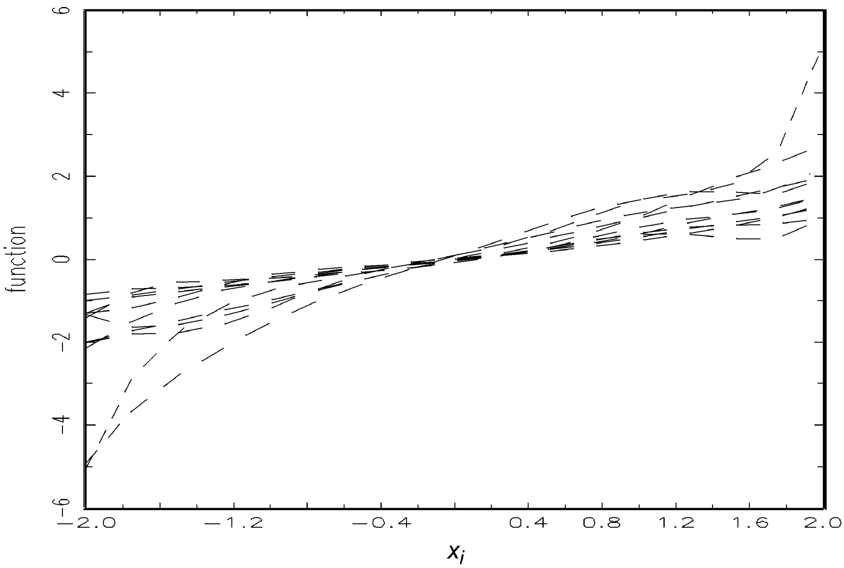


(a) n = 250

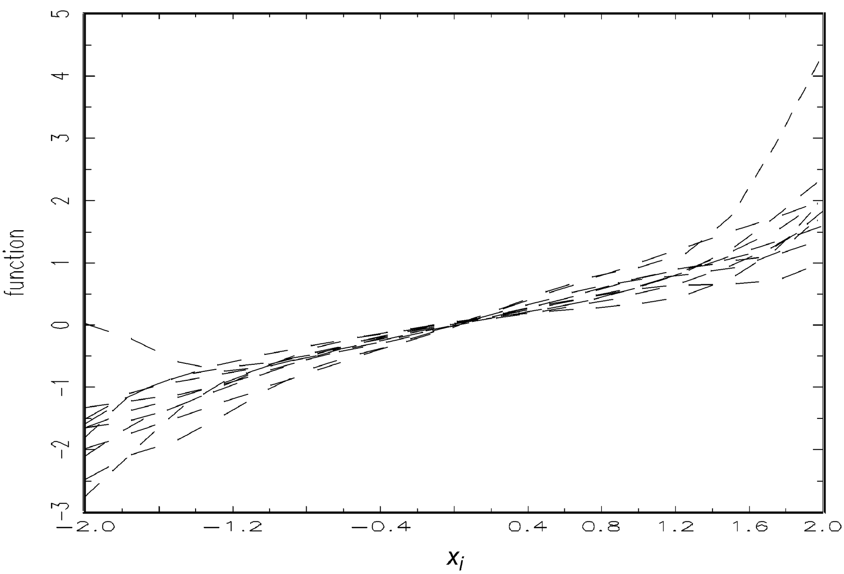


(b) n = 500

FIGURE 3. Density estimates, double exponential.

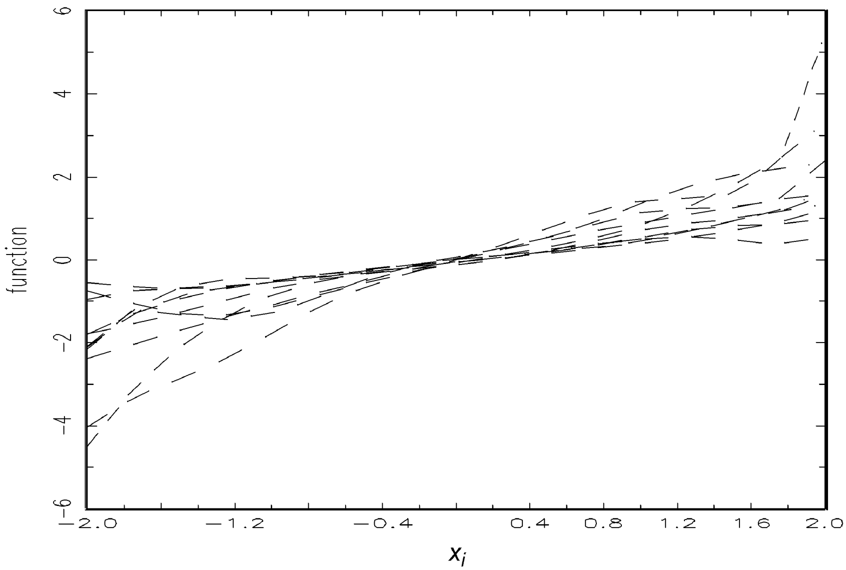


(a) n = 250

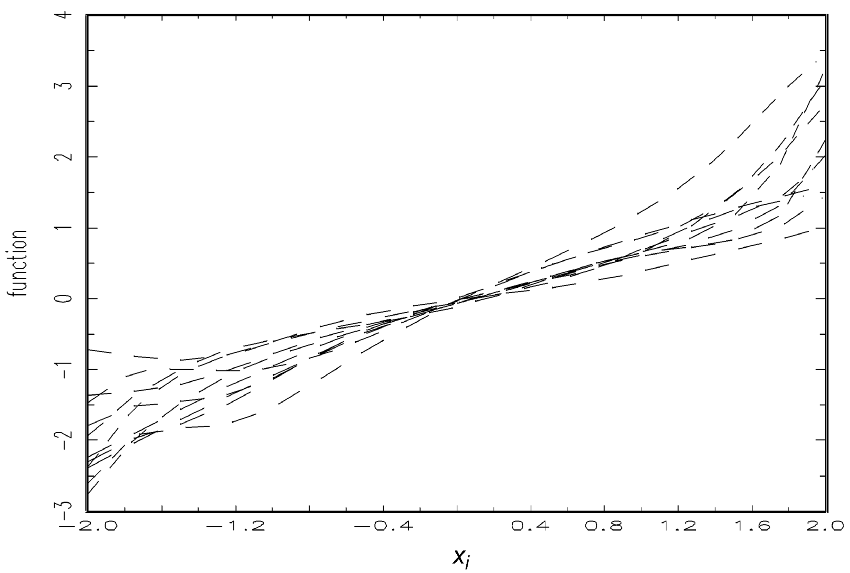


(b) n = 500

FIGURE 4. (a) and (b) Linear function, double exponential.



(c) $n = 250$



(d) $n = 500$

Figure 4. (c) and (d) Nonlinear function, normal distribution.

ulation experiments show reasonable behavior for sample sizes of 500 provided the bandwidth is chosen appropriately.

REFERENCES

Amemiya, T. (1985) *Advanced Econometrics*. Boston: Harvard University Press.

Andrews, D.W.K. & Y.J. Whang (1990) Additive interactive regression models: Circumvention of the curse of dimensionality. *Econometric Theory* 6, 466–479.

Beran, R. & P.W. Millar (1994) Minimum distance estimation in random coefficient regression models. *Annals of Statistics* 22, 1976–1992.

Chesher, A. (1997) Diet revealed: Semiparametric estimation of nutrient intake-age relationships. *Journal of the Royal Statistical Society, Series A* 160, 389–428.

Choo, E. (1998) Adverse Selection in the Interregional Market for Slaves. Manuscript, Yale University.

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

Fan, J. (1992) Asymptotic normality for deconvoluting kernel density estimators. *Sankhyā Series A* 53, 97–110.

Fan, J. & E. Masry (1992) Multivariate regression estimation with errors-in-variables: Asymptotic normality for mixing processes. *Journal of Multivariate Analysis* 43, 237–271.

Fan, J. & Y.K. Troung (1993) Nonparametric regression with errors in variables. *Annals of Statistics* 21, 1900–1925.

Härdle, W. (1990) *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

Horowitz, J.L. (1998) *Semiparametric Methods in Econometrics.* Lecture Notes in Statistics 131. New York: Springer-Verlag.

Horowitz, J.L. & M. Markatou (1996) Semiparametric estimation of regression models for panel data. *Review of Economic Studies* 63, 145–168.

Knight, J.L. & S.E. Satchell (1997) The cumulant generating function estimation method: Implementation and asymptotic efficiency. *Econometric Theory* 13, 170–184.

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

Linton, O.B., E. Mammen, J. Nielsen, & C. Tanggaard (1998) Estimating the yield curve by kernel smoothing methods. *Journal of Econometrics* 105, 185–223.

Mammen, E., O. Linton, & J.P. Nielsen (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics*.

Masry, E. (1991) Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory* 37, 1105–1115.

Masry, E. (1993) Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of Multivariate Analysis* 44, 47–68.

Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Stefanski, L. & R.J. Carroll (1990) Deconvoluting kernel density estimators. *Statistics* 2, 169–184.

APPENDIX

In the discussion that follows, we let C_j for some integer $j \geq 1$ denote a generic constant. (It is not meant to be equal in any two places it appears.) To simplify notation, we let $\int\int$ and $\int\int\int$ denote $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}$ and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}$, respectively, and we drop the subscripts on φ and $\hat{\varphi}$, so that we write $\varphi(t)$ for $\varphi_\varepsilon(t)$. The proof of the main results in the text uses the following lemma, which slightly extends Lemma 1 of Fan (1991) to the case where $v(\cdot)$ is any integrable function.

LEMMA 13. *Suppose that* $Q_n(\cdot) : \mathbb{R} \to \mathbb{R}$ *is a sequence of functions satisfying*

$$Q_n(u) \to Q(u) \quad \text{and} \quad \sup_n |Q_n(u)| \le Q^*(u),$$

where $Q^*(u)$ *satisfies*

$$\int_{-\infty}^{\infty} Q^*(u)du < \infty \quad \text{and} \quad \lim_{u\to\infty} |uQ^*(u)| = 0.$$

Suppose $v(\cdot) : \mathbb{R} \to \mathbb{R}$ *is an integrable function continuous at* x. *Then for any sequence* $h_n \to 0$, *we have*

$$\lim_{n\to\infty} \frac{1}{h_n} \int_{-\infty}^{\infty} Q_n\left(\frac{x-u}{h_n}\right) v(u)du = v(x) \int_{-\infty}^{\infty} Q(u)du.$$

Proof of Lemma 13. Let $\delta > 0$ be a constant. We have

$$\begin{aligned}
&\left| \frac{1}{h_n} \int_{-\infty}^{\infty} Q_n\left(\frac{x-u}{h_n}\right) v(u)du - v(x) \int_{-\infty}^{\infty} Q(u)du \right| \\
&\quad \le \int_{-\infty}^{\infty} [v(x-u) - v(x)] \frac{1}{h_n} Q_n\left(\frac{u}{h_n}\right) du + |v(x)| \left| \int_{-\infty}^{\infty} [Q_n(u) - Q(u)]\, dy \right| \\
&\quad \le \max_{|u|\le\delta} |v(x-u) - v(x)| \int_{-\infty}^{\infty} Q^*(u)du + \frac{1}{\delta} \sup_{|u|>\delta/h_n} |uQ^*(u)| \int_{-\infty}^{\infty} |v(u)|\, du \\
&\qquad + |v(x)| \int_{|u|>\delta/h_n} |Q^*(u)|\, du + |v(x)| \left| \int_{-\infty}^{\infty} [Q_n(u) - Q(u)]\, dy \right|. \qquad \textbf{(A.1)}
\end{aligned}$$

By the dominated convergence theorem and the assumptions, the last three terms in (A.1) tend to zero as $n \to \infty$. Then, let $\delta \to 0$ have the desired result. ■

Proof of Theorem 1. By a two-term Taylor expansion, we have

$$\begin{aligned}
\hat{f}_X(x) - f_X(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathbf{i}tx) \phi_{X_0}(t) [\hat{\varphi}(t) \phi_K(th) - \varphi(t)]\, dt \\
&\quad + \frac{1}{2\pi r} \int_{-\infty}^{\infty} \exp(-\mathbf{i}tx) \frac{\phi_K(th)\hat{\varphi}(t)}{[\phi_{X_0}(t)]^{r-1}} \left\{ \frac{\hat{\phi}_{\overline{X}}(t)}{\hat{\varphi}(rt)} - \frac{\phi_{\overline{X}}(t)}{\varphi(rt)} \right\} dt \\
&\quad + \frac{1-r}{2\pi r^2} \int_{-\infty}^{\infty} \int_0^1 (1-w) \exp(-\mathbf{i}tx) \frac{\phi_K(th)\hat{\varphi}(t)}{[\hat{\phi}^w(t)]^{2-1/r}} \\
&\quad \times \left\{ \frac{\hat{\phi}_{\overline{X}}(t)}{\hat{\varphi}(rt)} - \frac{\phi_{\overline{X}}(t)}{\varphi(rt)} \right\}^2 dwdt \\
&\equiv A_{1n} + A_{2n} + A_{3n}, \quad \text{say,} \qquad \textbf{(A.2)}
\end{aligned}$$

where

$$\hat{\phi}^{w}(t) = \frac{\phi_{\overline{X}}(t)}{\varphi(rt)} + w\left\{\frac{\hat{\phi}_{\overline{X}}(t)}{\hat{\varphi}(rt)} - \frac{\phi_{\overline{X}}(t)}{\varphi(rt)}\right\}. \tag{A.3}$$

Consider A_{1n}. By rearranging terms, we have

$$\begin{aligned} A_{1n} &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathbf{i}tx)\phi_{X_0}(t)\varphi(t)[\phi_K(th)-1]\,dt \\ &\quad + \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathbf{i}tx)\phi_{X_0}(t)\phi_K(th)[\hat{\varphi}(t)-\varphi(t)]\,dt \\ &\equiv A_{1n}^{*} + A_{1n}^{**}, \quad \text{say.} \end{aligned} \tag{A.4}$$

The convolution theorem implies

$$\begin{aligned} A_{1n}^{*} &= \int_{-\infty}^{\infty} K(u) f_X(x-hu)\,du - f_X(x) \\ &= f_X^{*}(x) - f_X(x). \end{aligned}$$

Therefore, for part (a) of Theorem 1, it suffices to establish the following results:

$$\frac{A_{1n}^{**}}{\sigma_{n1}(x)} \xrightarrow{p} 0, \tag{A.5}$$

$$\frac{A_{2n}}{\sigma_{n1}(x)} \Rightarrow N(0,1), \tag{A.6}$$

and

$$\frac{A_{3n}}{\sigma_{n1}(x)} \xrightarrow{p} 0. \tag{A.7}$$

The result (A.5) holds straightforwardly because we have

$$\begin{aligned} |A_{1n}^{**}| &\le \frac{1}{2\pi h}\int_{-\infty}^{\infty} |\phi_K(t)|\,dt \cdot \sup_{t\in\mathbb{R}} |\hat{\varphi}(t)-\varphi(t)| \\ &= O_p(n^{-\alpha/2}h^{-1}) \end{aligned}$$

using Assumptions E1 and A(iv) and hence $A_{1n}^{**}/\sigma_{n1}(x) = O_p(n^{(1-\alpha)/2}h^{(r-1)\beta_1-0.5}) = o_p(1)$.

Next, we verify (A.6). We first note that

$$\sup_{t\in\mathbb{R}}|\hat{\phi}_{\overline{X}}(t)-\phi_{\overline{X}}(t)|=O_p\left(\frac{1}{\sqrt{n}}\right) \tag{A.8}$$

by Chebyshev's inequality. We have

$$\begin{aligned}
A_{2n} &= \frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\,\frac{\phi_K(th)\varphi(t)}{[\phi_{X_0}(t)]^{r-1}\varphi(rt)}\,\{\hat{\phi}_{\overline{X}}(t)-\phi_{\overline{X}}(t)\}dt\\
&\quad+\frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\,\frac{\phi_K(th)}{[\phi_{X_0}(t)]^{r-1}\varphi(rt)}\,\{\hat{\varphi}(rt)-\varphi(rt)\}\{\hat{\phi}_{\overline{X}}(t)-\phi_{\overline{X}}(t)\}dt\\
&\quad+\frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\,\frac{\phi_K(th)\hat{\phi}_{\overline{X}}(t)\hat{\varphi}(t)}{[\phi_{X_0}(t)]^{r-1}\hat{\varphi}(rt)\varphi(rt)}\,\{\hat{\varphi}(rt)-\varphi(rt)\}dt\\
&= A_{2n}^{*}+A_{2n}^{**}+A_{2n}^{***},\quad\text{say.}
\end{aligned} \tag{A.9}$$

We first show that A_{2n}^{**} and A_{2n}^{***} are asymptotically negligible in the sense that both $A_{2n}^{**}/\sigma_{n1}(x)$ and $A_{2n}^{***}/\sigma_{n1}(x)$ are $o_p(1)$. Note that

$$A_{2n}^{*}=\frac{1}{n}\sum_{i=1}^{n}(Z_{nj}-EZ_{nj}),$$

where Z_{nj} is as defined in (23). By Assumption A(i), there exists a large (but fixed) constant $M>0$ such that for $|t|>M$,

$$|\phi_{X_0}(t)t^{\beta_1}|>\frac{|A_1|}{2};\qquad |\varphi(t)t^{\beta_2}|>\frac{|A_2|}{2}.$$

Therefore,

$$\begin{aligned}
&\int_{-\infty}^{\infty}\frac{|\phi_K(t)|}{|\phi_{X_0}(t/h)|^{r-1}|\varphi(rt/h)|}\,dt\\
&\quad\le 2\int_0^{Mh}\frac{|\phi_K(t)|}{|\phi_{X_0}(t/h)|^{r-1}|\varphi(rt/h)|}\,dt+2^{r+1}r^{\beta_2}\int_{Mh}^{\infty}\frac{|\phi_K(t)|}{A_1^{r-1}A_2}\left|\frac{t}{h}\right|^{(r-1)\beta_1+\beta_2}dt\\
&\quad\le 2Mh\,\frac{\max|\phi_K(t)|}{\min\limits_{|t|\le M}|\phi_{X_0}(t)|^{r-1}\min\limits_{|t|\le rM}|\varphi(t)|}+h^{-(r-1)\beta_1-\beta_2}\,\frac{2^{r+1}r^{\beta_2}}{A_1^{r-1}A_2}\\
&\qquad\times\int_0^{\infty}|\phi_K(t)||t|^{(r-1)\beta_1+\beta_2}dt\\
&\quad=O(h^{-(r-1)\beta_1-\beta_2}).
\end{aligned} \tag{A.10}$$

This result implies

$$|A_{2n}^{**}| \le \frac{1}{2\pi rh}\int_{-\infty}^{\infty}\frac{|\phi_K(t)|}{|\phi_{X_0}(t/h)|^{r-1}|\varphi(rt/h)|}\,dt\cdot\sup_{t\in\mathbb{R}}|\hat{\varphi}(t)-\varphi(t)|\cdot\sup_{t\in\mathbb{R}}|\hat{\phi}_{\overline{X}}(t)-\phi_{\overline{X}}(t)|$$

$$= O_p(n^{-1/2}n^{-\alpha/2}h^{-(r-1)\beta_1-\beta_2-1}) \quad \textbf{(A.11)}$$

using Assumptions E1 and (A.8). Therefore, $A_{2n}^{**}/\sigma_{n1}(x) = O_p(n^{-\alpha/2}h^{-\beta_2-1/2}) = o_p(1)$. Similarly, we have

$$|A_{2n}^{***}| \le C_1\,\frac{1}{h}\int_{-\infty}^{\infty}\frac{|\phi_K(t)||\varphi(t/h)|}{|\varphi(rt/h)|}\,dt\cdot\sup_{t\in\mathbb{R}}|\hat{\varphi}(t)-\varphi(t)|$$

$$= O_p(n^{-\alpha/2}h^{-1}), \quad \textbf{(A.12)}$$

where the first inequality holds with probability tending to one using (A.8) and Assumption E1 and the equality holds by Assumptions E1 and A(iv). Therefore, we also have $A_{2n}^{***}/\sigma_{n1}(x) = O_p(n^{(1-\alpha)/2}h^{(r-1)\beta_1-0.5}) = o_p(1)$. To establish the asymptotic normality (A.6), it now suffices to verify the following Lyapunov's condition: i.e., for some $\delta > 0$,

$$\frac{E|Z_{n1}-EZ_{n1}|^{2+\delta}}{n^{\delta/2}[\mathrm{var}(Z_{n1})]^{1+\delta/2}} \to 0 \quad \text{as } n\to\infty. \quad \textbf{(A.13)}$$

Let

$$\Psi_n(t) = \frac{\phi_K(t)\varphi(t/h)}{[\phi_{X_0}(t/h)]^{r-1}\varphi(rt/h)}. \quad \textbf{(A.14)}$$

By Fubini's theorem and the convolution theorem, we have

$$EZ_{n1} = \frac{1}{h}\int_{-\infty}^{\infty}G_n\left(\frac{x-u}{h}\right)f_{\overline{X}}(u)du$$

$$= \frac{1}{2\pi rh}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-\mathbf{i}t\left(\frac{x-u}{h}\right)\right)\Psi_n(t)f_{\overline{X}}(u)dtdu$$

$$= \frac{1}{2\pi rh}\int_{-\infty}^{\infty}\exp\left(-\mathbf{i}t\,\frac{x}{h}\right)\left[\int_{-\infty}^{\infty}\exp\left(\mathbf{i}t\,\frac{u}{h}\right)f_{\overline{X}}(u)du\right]\Psi_n(t)dt$$

$$= \frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\phi_X(t)\phi_K(th)dt$$

$$= \frac{1}{r}\int_{-\infty}^{\infty}K(u)f_X(x-hu)du \to \frac{1}{r}f_X(x), \quad \textbf{(A.15)}$$

where the last convergence holds by Lemma 13. By Assumption A1(i), we have

$$h^{(r-1)\beta_1}\Psi_n(t) \to \frac{r^{\beta_2}}{A_1^{r-1}}\,\phi_K(t)t^{(r-1)\beta_1}. \tag{A.16}$$

Furthermore, by Assumption A(i), there exists a large (but fixed) constant $M > 0$ such that for $|t| > M$, we have

$$|\phi_{X_0}(t)t^{\beta_1}| > \frac{|A_1|}{2}; \qquad |\varphi(t)t^{\beta_2}| > \frac{|A_2|}{2}; \qquad |\varphi(t)t^{\beta_2}| < 2|A_2|.$$

Therefore,

$$\begin{aligned}|h^{(r-1)\beta_1}\Psi_n(t)| \le {}& \frac{h^{(r-1)\beta_1}}{\min\limits_{|t|\le M}|\phi_{X_0}(t)|^{r-1}\min\limits_{|t|\le rM}|\varphi(t)|} \\ & \times 1(|t|\le hM) + \frac{2^{r+1}r^{\beta_2}}{|A_1|^{r-1}}\,|\phi_K(t)||t|^{(r-1)\beta_1}1(|t|>hM).\end{aligned} \tag{A.17}$$

For any $\varepsilon > 0$ and for all $h < \varepsilon/M$, we have

$$\begin{aligned}|h^{(r-1)\beta_1}\Psi_n(t)| &\le C_1\left(\frac{\varepsilon}{M}\right)^{(r-1)\beta_1}1(|t|\le\varepsilon) + \frac{2^{r+1}r^{\beta_2}}{|A_1|^{r-1}}\,|\phi_X(t)||t|^{(r-1)\beta_1} \\ &\equiv \Delta(t).\end{aligned} \tag{A.18}$$

Because $\Delta(t)$ is integrable by Assumption A(iv), we have

$$\begin{aligned}h^{(r-1)\beta_1}G_n(x) &= \frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)h^{(r-1)\beta_1}\Psi_n(t)dt \\ &\to \frac{r^{\beta_2-1}}{2\pi A_1^{r-1}}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\phi_K(t)t^{(r-1)\beta_1}dt\end{aligned} \tag{A.19}$$

by (A.16) and dominated convergence theorem (A.18). Integrability of $\Delta(t)$ also implies that

$$|h^{(r-1)\beta_1}G_n(x)| \le \frac{1}{2\pi r}\int_{-\infty}^{\infty}\Delta(t)dt \equiv C_2 < \infty. \tag{A.20}$$

By integration by parts,

$$(\mathbf{i}x)G_n(x) = \frac{1}{2\pi r}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\left(\frac{\partial}{\partial t}\,\Psi_n(t)\right)dt. \tag{A.21}$$

Using arguments similar to those in (A.17) and (A.18) and Assumptions A(i) and (iv), we have

$$|xG_n(x)| \le O(h^{-(r-1)\beta_1}). \tag{A.22}$$

Expressions (A.20) and (A.22) combine to give

$$|h^{(r-1)\beta_1}G_n(x)| \leq \frac{C_3}{1+|x|}. \tag{A.23}$$

Now, we have

$$\begin{aligned}
EZ_{n1}^2 &= \frac{1}{h^2}\int_{-\infty}^{\infty}\left[G_n\left(\frac{x-u}{h}\right)\right]^2 f_{\overline{X}}(u)du \\
&= \frac{f_{\overline{X}}(x)}{h^{2(r-1)\beta_1+1}}\int_{-\infty}^{\infty}\left[\frac{r^{\beta_2-1}}{2\pi A_1^{r-1}}\int_{-\infty}^{\infty}\exp(-\mathbf{i}tx)\phi_K(t)t^{(r-1)\beta_1}dt\right]^2 dy(1+o(1)) \\
&= \frac{1}{h^{2(r-1)\beta_1+1}}\cdot\frac{f_{\overline{X}}(x)r^{2(\beta_2-1)}}{2\pi A_1^{2(r-1)}}\int_{\infty}^{\infty}|\phi_K(t)|^2|t|^{2(r-1)\beta_1}dt(1+o(1)) \\
&= h^{-2(r-1)\beta_1-1}\sigma_1^2(x)(1+o(1)),
\end{aligned} \tag{A.24}$$

where the second equality holds by (A.19), (A.23), and Lemma 13 and the third equality holds by Parseval's identity.

Similarly, by (A.23) and Lemma 13, we have

$$E|Z_{n1}|^{2+\delta} = O(h^{-(2+\delta)[(r-1)\beta_1+1]+1}). \tag{A.25}$$

Therefore, by (A.15), (A.24), and (A.25), the Lyapunov condition holds using the fact that $nh \to \infty$.

Next, we verify (A.7). We have

$$\begin{aligned}
\hat{\phi}^w\left(\frac{t}{h}\right)\left(\frac{t}{h}\right)^{r\beta_1} &= \left[\frac{\phi_{\overline{X}}(t/h)}{\varphi(rt/h)} + w\left\{\frac{\hat{\phi}_{\overline{X}}(t/h)}{\hat{\varphi}(rt/h)} - \frac{\phi_{\overline{X}}(t/h)}{\varphi(rt/h)}\right\}\right]\left(\frac{t}{h}\right)^{r\beta_1} \\
&= \left[\phi_{X_0}\left(\frac{t}{h}\right)\left(\frac{t}{h}\right)^{\beta_1}\right]^r + o_p(1)
\end{aligned}$$

uniformly in $w \in (0,1)$ using Assumption E1 and (A.8) because $n^{\alpha}h^{2r\beta_1} \to \infty$. Therefore, (A.7) holds because we then have

$$\begin{aligned}
|A_{3n}| &\leq \frac{r-1}{2\pi r^2}\int_{-\infty}^{\infty}\frac{|\phi_K(th)||\hat{\varphi}(t)|}{|\hat{\phi}^w(t)|^{2-1/r}}\left|\frac{\hat{\phi}_{\overline{X}}(t)}{\hat{\varphi}(rt)} - \frac{\phi_{\overline{X}}(t)}{\varphi(rt)}\right|^2 dt \\
&\leq \frac{r-1}{\pi r^2}\int_{-\infty}^{\infty}\frac{|\phi_K(th)||\hat{\varphi}(t)|}{|\hat{\phi}^w(t)|^{2-1/r}|\varphi(rt)|^2}dt\cdot\sup_{t\in\mathbb{R}}|\hat{\phi}_{\overline{X}}(t) - \phi_{\overline{X}}(t)|^2 \\
&\quad + \frac{r-1}{\pi r^2}\int_{-\infty}^{\infty}\frac{|\phi_K(th)||\hat{\varphi}(t)||\hat{\phi}_{\overline{X}}(t)|^2}{|\hat{\phi}^w(t)|^{2-1/r}|\hat{\varphi}(rt)|^2|\varphi(rt)|^2}dt\cdot\sup_{t\in\mathbb{R}}|\hat{\varphi}(t) - \varphi(t)|^2 \\
&\leq O_p(n^{-1}h^{-(2r-1)\beta_1-\beta_2-1})
\end{aligned} \tag{A.26}$$

uniformly in $w \in (0,1)$. Now the proof of part (a) is complete because $A_{3n}/\sigma_{n1}(x) = O(n^{-\alpha/2}h^{-r\beta_1-\beta_2-0.5}) = o_p(1)$.

Finally, part (b) follows by dominated convergence theorem using the continuity and boundedness of the kth derivative of $f_X(\cdot)$ (see Assumption A(v)). ■

Proof of Lemma 2. It suffices to establish

$$\frac{1}{n}\sum_{j=1}^{n}(\hat{Z}_{nj}^2 - Z_{nj}^2) \xrightarrow{p} 0; \tag{A.27}$$

$$\frac{1}{n}\sum_{j=1}^{n}(\hat{Z}_{nj} - Z_{nj}) \xrightarrow{p} 0; \tag{A.28}$$

$$\frac{\sum_{j=1}^{n} Z_{nj}^2}{nEZ_{n1}^2} \xrightarrow{p} 1; \tag{A.29}$$

$$\frac{1}{n}\sum_{j=1}^{n} Z_{nj} - EZ_{n1} \xrightarrow{p} 0. \tag{A.30}$$

First, consider (A.28). We have

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{j=1}^{n}(\hat{Z}_{nj} - Z_{nj})\right| \\
&\quad\le \sup_{1\le j\le n}|\hat{Z}_{nj} - Z_{nj}| \\
&\quad\le \frac{1}{2\pi rh}\int_{-\infty}^{\infty}\left|\frac{\hat{\varphi}(t/h)}{[\hat{\phi}_{\overline{X}}(t/h)]^{(r-1)/r}[\hat{\varphi}(t/h)]^{1/r}} - \frac{\varphi(t/h)}{[\phi_{\overline{X}}(t/h)]^{(r-1)/r}[\varphi(t/h)]^{1/r}}\right| dt \\
&\quad\le C_1\frac{1}{h}\int_{-\infty}^{\infty}\frac{|\hat{\varphi}(t/h)|}{[\phi_{\overline{X}}^{*}(t/h)]^{(2r-1)/r}[\varphi^{*}(t/h)]^{1/r}}\,dt\cdot\sup_{t\in\mathbb{R}}|\hat{\phi}_{\overline{X}}(t) - \phi_{\overline{X}}(t)| \\
&\qquad + C_2\frac{1}{h}\int_{-\infty}^{\infty}\frac{|\hat{\varphi}(t/h)|}{[\phi_{\overline{X}}^{*}(t/h)]^{(r-1)/r}[\varphi^{*}(t/h)]^{(r+1)/r}}\,dt\cdot\sup_{t\in\mathbb{R}}|\hat{\varphi}(t) - \varphi(t)| \\
&\quad\le O_p(n^{-1/2}h^{-(2r-1)\beta_1-\beta_2-1}) + O_p(n^{-\alpha/2}h^{-\beta_2}) \xrightarrow{p} 0,
\end{aligned} \tag{A.31}$$

where the third inequality follows from a one-term Taylor expansion and the last inequality holds using arguments analogous to (A.26). Expression (A.27) can be similarly verified:

$$\left|\frac{1}{n}\sum_{j=1}^{n}(\hat{Z}_{nj}^2 - Z_{nj}^2)\right| \le O_p(n^{-1/2}h^{-(3r-2)\beta_1-\beta_2-2}) + O_p(n^{-\alpha/2}h^{-(r-1)\beta_1-\beta_2-1}) \xrightarrow{p} 0. \tag{A.32}$$

Next, (A.29) holds by the weak law of large numbers because

$$\frac{1}{EZ_{n1}^2} E[Z_{n1}^2 1(|Z_{n1}|^2 \geq \varepsilon n E Z_{n1}^2)]$$

$$\leq \frac{E|Z_{n1}|^{2(1+\delta)}}{(\varepsilon n)^{\delta}[EZ_{n1}^2]^{1+\delta}} = \frac{O(h^{-2(1+\delta)[(r-1)\beta_1+1]+1})}{(\varepsilon n)^{\delta}[h^{-2(r-1)\beta_1-1}\sigma_1^2(x)(1+o(1))]^{1+\delta}}$$

$$= O((nh)^{-\delta}) \to 0 \qquad \textbf{(A.33)}$$

for each $\varepsilon > 0$ and $\delta > 0$ using the fact that $nh \to \infty$. Finally, (A.30) holds because

$$\frac{1}{n} \operatorname{var}(Z_{n1}) = O(n^{-1}h^{-2(r-1)\beta_1-1}) \to 0 \qquad \textbf{(A.34)}$$

using Chebyshev's inequality. Now the proof of Lemma 2 is complete. ■

Proof of Theorem 4. By a two-term Taylor expansion and rearranging terms, we have

$$\begin{aligned}
&\hat{g}_X(x) - g_X(x) \\
&= \frac{1}{(2\pi)^2} \iiint y \exp(-\mathbf{i}(sy+tx))\phi_{Y,X}(s,t)[\tilde{\phi}_K(sh,th) - 1]\, dsdtdy \\
&+ \frac{1}{(2\pi)^2} \iiint y \exp(-\mathbf{i}(sy+tx))\phi_{Y_0,X_0}(s,t)\tilde{\phi}_K(sh,th)[\hat{\varphi}(s,t) - \varphi(s,t)]\, dsdtdy \\
&+ \frac{1}{(2\pi)^2 r} \iiint y \exp(-\mathbf{i}(sy+tx)) \frac{\tilde{\phi}_K(sh,th)\hat{\varphi}(s,t)}{[\phi_{Y_0,X_0}(s,t)]^{r-1}} \\
&\times \left\{ \frac{\hat{\phi}_{\overline{Y},\overline{X}}(s,t)}{\hat{\varphi}(rs,rt)} - \frac{\phi_{\overline{Y},\overline{X}}(s,t)}{\varphi(rs,rt)} \right\} dsdtdy \\
&+ \frac{1-r}{(2\pi)^2 r^2} \iiiint_0^1 \frac{(1-w)y\exp(-\mathbf{i}(sy+tx))\tilde{\phi}_K(sh,th))\hat{\varphi}(s,t)}{[\hat{\phi}^w(s,t)]^{2-1/r}} \\
&\times \left\{ \frac{\hat{\phi}_{\overline{Y},\overline{X}}(s,t)}{\hat{\varphi}(rs,rt)} - \frac{\phi_{\overline{Y},\overline{X}}(s,t)}{\varphi(rs,rt)} \right\}^2 dwdsdtdy \\
&\equiv B_{1n} + B_{1n}^* + B_{2n} + B_{3n}, \quad \text{say,}
\end{aligned} \qquad \textbf{(A.35)}$$

where

$$\hat{\phi}^w(s,t) = \frac{\phi_{\overline{Y},\overline{X}}(s,t)}{\varphi(rs,rt)} + w\left(\frac{\hat{\phi}_{\overline{Y},\overline{X}}(s,t)}{\hat{\varphi}(rs,rt)} - \frac{\phi_{\overline{Y},\overline{X}}(s,t)}{\varphi(rs,rt)} \right). \qquad \textbf{(A.36)}$$

By a straightforward argument, we have

$$B_{1n} = \iiint y[f_{Y,X}(y-hu, x-hv) - f_{Y,X}(y,x)]K(u)K(v)dudvdy$$

$$= \int_{-\infty}^{\infty} [g_X(x-hu)f_X(x-hu) - g_X(x)f_X(x)]K(u)du$$

$$= R_{n1}^*, \tag{A.37}$$

where R_{n1}^* is as defined in (34).

Subsequently we establish the following results:

$$\sqrt{nh^{2(r-1)\rho_1+1}}B_{1n}^* \xrightarrow{p} 0, \tag{A.38}$$

$$\frac{\sqrt{nh^{2(r-1)\rho_1+1}}B_{2n}}{\sigma_2(x)} \Rightarrow N(0,1), \tag{A.39}$$

and

$$\sqrt{nh^{2(r-1)\rho_1+1}}B_{3n} \xrightarrow{p} 0. \tag{A.40}$$

Then, part (a) of Theorem 4 follows by noting

$$\hat{m}(x) - m(x) - R_n(x)$$

$$= \hat{f}_X^{-1}(x)\{[\hat{g}(x) - g(x)] - [\hat{f}_X(x) - f_X(x)]m(x) - [R_{n1}^* - R_{n2}^*]\}$$

$$= \hat{f}_X^{-1}(x)\{[\hat{g}(x) - g(x) - R_{n1}^*] - [\hat{f}_X(x) - f_X^*(x)]m(x)\}$$

$$= \hat{f}_X^{-1}(x)\{[B_{1n}^* + B_{2n} + B_{3n}] - [A_{2n} + A_{3n}]m(x)\}$$

$$= (f_X^{-1}(x) + o_p(1))\{[B_{1n}^* + B_{2n} + B_{3n}] + O_p(n^{-1/2}h^{-(r-1)\beta_1-1/2})\}, \tag{A.41}$$

and hence

$$\sqrt{nh^{2(r-1)\rho_1+1}}(\hat{m}(x) - m(x) - R_n(x)) = \sqrt{nh^{2(r-1)\rho_1+1}}f_X^{-1}(x)B_{2n} + o_p(1), \tag{A.42}$$

where A_{2n} and A_{3n} are as defined in (A.2) and the last equality in (A.41) follows by the proof of Theorem 1.

First, we verify (A.38). We first write

$$B_{1n}^* = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} y\Psi_n(x,y)dy, \tag{A.43}$$

where

$$\Psi_n(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy + tx))H_n(s, t)Q_n(s, t)dsdt, \tag{A.44}$$

$$H_n(s, t) = \phi_{Y_0, X_0}(s, t)\phi_K(sh)\phi_K(th), \quad \text{and} \tag{A.45}$$

$$Q_n(s, t) = \hat{\varphi}(s, t) - \varphi(s, t). \tag{A.46}$$

By integration by parts, we have

$$(\mathbf{i}y)\Psi_n(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy + tx))\left\{\frac{\partial}{\partial s}[H_n(s, t)Q_n(s, t)]\right\} dsdt, \tag{A.47}$$

$$(\mathbf{i}y)^3\Psi_n(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy + tx))\left\{\frac{\partial^3}{\partial s^3}[H_n(s, t)Q_n(s, t)]\right\} dsdt. \tag{A.48}$$

By Assumption E2, we have

$$\sup_{(s,t)\in\mathbb{R}^2}\left|\frac{\partial^j}{\partial s^j} Q_n(s, t)\right| = O_p(n^{-\alpha/2}). \tag{A.49}$$

Therefore, we have

$$\begin{aligned}
|y\Psi_n(x, y)| &\leq \iint |H_n(s, t)|\, dsdt \cdot \sup_{(s,t)\in\mathbb{R}^2}\left|\frac{\partial}{\partial s} Q_n(s, t)\right| \\
&\quad + \iint \left|\frac{\partial}{\partial s} H_n(s, t)\right| dsdt \cdot \sup_{(s,t)\in\mathbb{R}^2} |Q_n(s, t)| \\
&\leq h^{-2}\left[\int_{-\infty}^{\infty} |\phi_K(t)|\, dt\right]^2 \cdot O_p(n^{-\alpha/2}) \\
&\quad + \left[h^{-2}E|\overline{Y}|\left\{\int_{-\infty}^{\infty} |\phi_K(t)|\, dt\right\}^2\right. \\
&\qquad \left. + h^{-1}\left\{\int_{-\infty}^{\infty} |\phi_K(t)|\, dt\right\}\left\{\int_{-\infty}^{\infty} |\phi'_K(t)|\, dt\right\}\right] \cdot O_p(n^{-\alpha/2}) \\
&= O_p(n^{-\alpha/2}h^{-2}).
\end{aligned} \tag{A.50}$$

Similarly, we can also show that

$$|y^3\Psi_n(x, y)| \leq O_p(n^{-\alpha/2}h^{-2}). \tag{A.51}$$

Now (A.50) and (A.51) imply

$$|B^*_{1n}| \leq \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} |y\Psi_n(x, y)|\, dy \leq O_p(n^{-\alpha/2}h^{-2}). \tag{A.52}$$

Thus, because $\sqrt{nh^{2(r-1)\rho_1+1}}B^*_{1n} = O_p(n^{(1-\alpha)/2}h^{(r-1)\rho_1-3/2}) = o_p(1)$, the desired result (A.38) follows.

We next verify (A.39). Rewrite

$$
\begin{aligned}
B_{2n} &= \frac{1}{(2\pi)^2 r} \iiint y \exp(-\mathbf{i}(sy+tx)) \frac{\tilde{\phi}_K(sh,th)\varphi(s,t)}{[\phi_{Y_0,X_0}(s,t)]^{r-1}\varphi(rs,rt)} \\
&\quad \times \{\hat{\phi}_{\overline{Y},\overline{X}}(s,t) - \phi_{\overline{Y},\overline{X}}(s,t)\} dsdtdy \\
&\quad + \frac{1}{(2\pi)^2 r} \iiint y \exp(-\mathbf{i}(sy+tx)) \frac{\tilde{\phi}_K(sh,th)}{[\phi_{Y_0,X_0}(s,t)]^{r-1}\varphi(rs,rt)} \\
&\quad \times \{\hat{\varphi}(rs,rt) - \varphi(rs,rt)\}\{\hat{\phi}_{\overline{Y},\overline{X}}(s,t) - \phi_{\overline{Y},\overline{X}}(s,t)\} dsdtdy \\
&\quad - \frac{1}{(2\pi)^2 r} \iiint y \exp(-\mathbf{i}(sy+tx)) \frac{\tilde{\phi}_K(sh,th)\hat{\phi}_{\overline{Y},\overline{X}}(s,t)\hat{\varphi}(s,t)}{[\phi_{Y_0,X_0}(s,t)]^{r-1}\hat{\varphi}(rs,rt)\varphi(rs,rt)} \\
&\quad \times \{\hat{\varphi}(rs,rt) - \varphi(rs,rt)\} dsdtdy \\
&= B_{2n}^* + B_{2n}^{**} - B_{2n}^{***}, \quad \text{say.}
\end{aligned}
$$

Observe that

$$
\begin{aligned}
B_{2n}^* &= \frac{1}{(2\pi)^2 r} \iiint y \exp(-\mathbf{i}(sy+tx)) \frac{\tilde{\phi}_K(sh,th)\varphi(s,t)}{[\phi_{Y_0,X_0}(s,t)]^{r-1}\varphi(rs,rt)} \\
&\quad \times \left\{ \frac{1}{n} \sum_{\mathbf{i}=1}^{n} (\exp(\mathbf{i}(s\overline{Y}_j + t\overline{X}_j)) - E\exp(\mathbf{i}(s\overline{Y}_j + t\overline{X}_j))) \right\} dsdtdy \\
&= \frac{1}{n} \sum_{\mathbf{i}=1}^{n} (Z_{nj} - EZ_{nj}),
\end{aligned}
\tag{A.53}
$$

where Z_{nj} is as defined in (37). Using arguments similar to (A.52), we have

$$
\begin{aligned}
|B_{2n}^{**}| &= O_p(n^{-1/2} n^{-\alpha/2} h^{-(r-1)\rho_1 - \rho_2 - 2}), \\
|B_{2n}^{***}| &= O_p(n^{-\alpha/2} h^{-2}),
\end{aligned}
$$

so that $\sqrt{nh^{2(r-1)\rho_1+1}}(B_{2n}^{**} + B_{2n}^{***}) = o_p(1)$. Therefore, to establish (A.39), it suffices to verify

$$
\frac{\sqrt{nh^{2(r-1)\rho_1+1}} B_{2n}^*}{\sigma_2(x)} \Rightarrow N(0,1).
\tag{A.54}
$$

For (A.54), we verify the Lyapunov condition (A.13). We have

$$
\begin{aligned}
EZ_{n1} &= \frac{1}{r} \int_{-\infty}^{\infty} y \left[\frac{1}{(2\pi)^2} \iiiint \exp(-\mathbf{i}s(y-y^*) - \mathbf{i}t(x-x^*)) \right. \\
&\qquad \left. \times \frac{\tilde{\phi}_K(sh,th)\varphi(s,t)}{[\phi_{Y_0,X_0}(s,t)]^{r-1}\varphi(rs,rt)} f_{\overline{Y},\overline{X}}(y^*,x^*) dsdtdy^*dx^* \right] dy \\
&= r^{-1} \int_{-\infty}^{\infty} y \left[\iint f_{Y,X}(y-hu, x-hv) \tilde{K}(u,v) dudv \right] dy \\
&= r^{-1} \int_{-\infty}^{\infty} K(u) g_X(x-hu) f_X(x-hu) du \to r^{-1} m(x),
\end{aligned}
\tag{A.55}
$$

where the last convergence holds by Lemma 13. We also have

$$EZ_{n1}^2 = E\left[\overline{Y}_1 \frac{1}{h} K_{n1}\left(\frac{x-\overline{X}_1}{h}\right) + K_{n2}\left(\frac{x-\overline{X}_1}{h}\right)\right]^2$$

$$= h^{-1}\int_{-\infty}^{\infty} [K_{n1}(u)]^2 v_{\overline{X}}(x-hu) f_{\overline{X}}(x-hu)du$$

$$+ h\int_{-\infty}^{\infty} [K_{n2}(u)]^2 f_{\overline{X}}(x-hu)du$$

$$+ 2\int_{-\infty}^{\infty} K_{n1}(u)K_{n2}(u) m_{\overline{X}}(x-hu) f_{\overline{X}}(x-hu)du$$

$$\equiv C_{1n} + C_{2n} + C_{3n}, \quad \text{say.} \tag{A.56}$$

Subsequently we show that C_{1n} is the dominating term. Using the arguments similar to those to establish (A.20) and (A.22), we have

$$|h^{(r-1)\rho_1} y^j G_n(y,x)| \leq C_j \quad \text{and} \tag{A.57}$$

$$|h^{(r-1)\rho_1} x y^l G_n(y,x)| \leq D_l \tag{A.58}$$

for some constant C_j ($j = 0,1,2,3$) and D_l ($l = 0,2$). Note that, similarly to (A.19), we have

$$h^{(r-1)\rho_1} G_n(y,x) \to \frac{r^{\rho_2-1}}{(2\pi)^2 B_1^{r-1}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\mathbf{i}(sy+tx))\phi_K(s)\phi_K(t)\|(s,t)\|^{(r-1)\rho_1} ds dt$$

$$\equiv G^*(y,x). \tag{A.59}$$

Therefore, (A.57) (with $j = 0$ and 2) together with (A.59) implies

$$h^{(r-1)\rho_1} K_{n1}(x) \to \int_{-\infty}^{\infty} G^*(y,x)dy \tag{A.60}$$

by the dominated convergence theorem. Note also that (A.57) (with $j = 0$) together with (A.58) (with $l = 0$ and 2) implies

$$|h^{(r-1)\rho_1} K_{n1}(x)| \leq \frac{C_4}{1+|x|}. \tag{A.61}$$

Therefore,

$$h^{2(r-1)\rho_1+1} C_{1n} = \int_{-\infty}^{\infty} [h^{(r-1)\rho_1} K_{n1}(u)]^2 v_{\overline{X}}(x-hu) f_{\overline{X}}(x-hu)du$$

$$\to v_X(x) f_X(x) \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} G^*(y,x)dy\right]^2 dx = \sigma_2^2(x) \tag{A.62}$$

by Lemma 13. Similarly, we have

$$\begin{aligned} h^{2(r-1)\rho_1+1}C_{2n} &= h\int_{-\infty}^{\infty}[h^{(r-1)\rho_1}K_{n2}(u)]^2 f_{\overline{X}}(x-hu)du \\ &= h\left(f_X(x)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} yG^*(y,x)dy\right]^2 dx + o(1)\right) \\ &= o(1). \end{aligned} \quad \textbf{(A.63)}$$

By Cauchy–Schwarz inequality, (A.62) and (A.63) imply that $h^{2(r-1)\rho_1+1}C_{3n}$ is also $o(1)$. Therefore, this establishes that C_{1n} in (A.56) is the dominating term. Because $EZ_{n1} = O(1)$, we now have

$$\begin{aligned} h^{2(r-1)\rho_1+1}\mathrm{var}(Z_{n1}) &= h^{2(r-1)\rho_1+1}E(Z_{n1}^2) + o(1) \\ &\to \sigma_2^2(x). \end{aligned} \quad \textbf{(A.64)}$$

We also have

$$E|Z_{n1}|^{2+\delta} = O(h^{-(2+\delta)[(r-1)\rho_1+1]+1}). \quad \textbf{(A.65)}$$

Therefore, the Lyapunov condition holds because $nh \to \infty$ as is required.

Next, we verify (A.40). It can be verified using an argument similar to that of (A.38) after we rewrite

$$\begin{aligned} &\frac{1-r}{(2\pi)^2r^2}\iiiint_0^1 \frac{(1-w)y\exp(-\mathbf{i}(sy+tx))\widetilde{\phi}_K(sh,th)\hat{\varphi}(s,t)}{[\hat{\phi}^w(s,t)]^{2-1/r}} \\ &\quad\times\left\{\frac{\hat{\phi}_{\overline{Y},\overline{X}}(s,t)}{\hat{\varphi}(rs,rt)} - \frac{\phi_{\overline{Y},\overline{X}}(s,t)}{\varphi(rs,rt)}\right\}^2 dwdsdtdy \\ B_{3n} &= \frac{1-r}{(2\pi)^2r^2}\int_{-\infty}^{\infty}\int_0^1 (1-w)y\Psi_n(w,x,y)dwdy, \end{aligned} \quad \textbf{(A.66)}$$

where

$$\Psi_n(w,x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-\mathbf{i}(sy+tx))H_n(w,s,t)Q_n(s,t)dsdt, \quad \textbf{(A.67)}$$

$$H_n(w,s,t) = \frac{\widetilde{\phi}_K(sh,th)\hat{\varphi}(s,t)}{[\hat{\phi}^w(s,t)]^{2-1/r}}, \quad \text{and} \quad \textbf{(A.68)}$$

$$Q_n(s,t) = \left\{\frac{\hat{\phi}_{\overline{Y},\overline{X}}(s,t)}{\hat{\varphi}(rs,rt)} - \frac{\phi_{\overline{Y},\overline{X}}(s,t)}{\varphi(rs,rt)}\right\}^2. \quad \textbf{(A.69)}$$

Some tedious calculation yields

$$|y\Psi_n(w,x,y)| \le O_p(n^{-1}h^{-(2r-1)\rho_1-\rho_2-2})$$

and

$$|y^3\Psi_n(w,x,y)| \le O_p(n^{-1}h^{-(2r-1)\rho_1-\rho_2-2})$$

uniformly in $w \in (0,1)$. Therefore, we have

$$nh^{(2r-1)\rho_1+\rho_2+2}|B_{3n}| \le C_4 nh^{(2r-1)\rho_1+\rho_2+2} \int_{-\infty}^{\infty}\int_0^1 |y\Psi_n(w,x,y)|\, dy = O_p(1).$$

Thus, because $nh^{2r\rho_1+2\rho_2+3} \to \infty$, the desired result (A.40) follows.

Finally, part (b) of Theorem 4 follows by the dominated convergence theorem using the continuity and boundedness of the kth derivative of $f_X(\cdot)$ and $g_X(\cdot)$. ∎

Proof of Lemma 5. This is similar to the proof of Lemma 2. ∎

The proof of Theorem 7 uses the following lemmas. (The proofs of Lemma 14 and 15 are similar to [but simpler than] those of Lemmas 16 and 17 given subsequently and hence are omitted.)

LEMMA 14. *Under Assumptions C(i)–(iv),*

(a) we have as $h \to 0$

$$\sup_{x\in\mathbb{R}} |G_n(x)| = O\left(h^{\beta(l+1)+(r-1)\beta_0}\left(\ln\frac{1}{h}\right)^l \exp\left[\{a_0(r-1)+a_1(r^\beta-1)\}\left(\frac{d}{h}\right)^\beta\right]\right)$$

and

(b) if moreover Assumptions C(v) and (vi) hold, then we have

$$|G_n(x)| \ge B_5 H(x) h^{\beta(l+1)+(r-1)\beta_0} \exp\left[\{a_0(r-1)+a_1(r^\beta-1)\}\left(\frac{d}{h}\right)^\beta\right]$$

for some B_5 *uniformly in x on a bounded interval, where*

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

LEMMA 15. *Under Assumption C, we have for large n*

(a)

$$\operatorname{var}(Z_{n1}) \le B_6 h^{2[\beta(l+1)+(r-1)\beta_0-1]}\left(\ln\frac{1}{h}\right)^{2l} \times \exp\left[2\{a_0(r-1)+a_1(r^\beta-1)\}\left(\frac{d}{h}\right)^\beta\right]$$

and

(b)

$$\operatorname{var}(Z_{n1}) \ge B_7 h^{2[\beta(l+1)+(r-1)\beta_0]-1} \exp\left[2\{a_0(r-1)+a_1(r^\beta-1)\}\left(\frac{d}{h}\right)^\beta\right].$$

Proof of Theorem 7. Consider the Taylor expansion (A.2). To prove Theorem 7, it suffices to verify the conditions (A.5)–(A.7) with $\sigma_{n1}(x)$ replaced by $\sigma_{n3}(x)$.

We first verify (A.5). Using arguments similar to the proof of Lemma 14, we can show

$$\int_{-\infty}^{\infty} |\phi_K(t)| \left| \phi_{X_0}\left(\frac{t}{h}\right) \right| dt = O\left(h^{\beta(l+1)-\beta_0-1} \left(\ln \frac{1}{h} \right)^l \exp\left[-a_0 \left(\frac{d}{h} \right)^\beta \right] \right). \qquad \textbf{(A.70)}$$

Therefore, we have

$$\begin{aligned} \left| \frac{A_{1n}^{**}}{\sigma_{n3}(x)} \right| &\le \frac{1}{\sigma_{n3}(x)} \cdot \frac{1}{2\pi h} \int_{-\infty}^{\infty} |\phi_K(t)| \left| \phi_{X_0}\left(\frac{t}{h}\right) \right| dt \cdot \sup_{t\in\mathbb{R}} |\hat{\varphi}(t) - \varphi(t)| \\ &= O_p\left(n^{1/2} n^{-\alpha/2} h^{-r\beta_0 - 1/2} \left(\ln \frac{1}{h} \right)^l \exp\left[-\{a_0 r + a_1(r^\beta - 1)\} \left(\frac{d}{h} \right)^\beta \right] \right) \\ &= O_p(n^{1/2} n^{-\alpha/2} n^{-\{a_0 r + a_1(r^\beta - 1)\}\gamma}) \xrightarrow{p} 0, \end{aligned} \qquad \textbf{(A.71)}$$

where the last equality holds because $h = d(\gamma \ln n)^{-1/\beta}$.

Next, we consider (A.6). By Lemma 14, we have

$$\begin{aligned} E|Z_{n1}|^{2+\delta} &\le \frac{1}{h^{2+\delta}} E \left| G_n\left(\frac{x - \bar{X}_1}{h} \right) \right|^{2+\delta} \\ &\le \frac{1}{h^{2+\delta}} \sup_{x\in\mathbb{R}} |G_n(x)|^{2+\delta} \\ &= C_1 h^{[\beta(l+1)+(r-1)\beta_0-1](2+\delta)} \left(\ln \frac{1}{h} \right)^{(2+\delta)l} \\ &\quad \times \exp\left[\{a_0(r-1) + a_1(r^\beta - 1)\}(2+\delta) \left(\frac{d}{h} \right)^\beta \right]. \end{aligned} \qquad \textbf{(A.72)}$$

Note that EZ_{n1} is $O(1)$ by (A.15). Thus the Lyapunov condition holds because

$$\frac{E|Z_{n1} - EZ_{n1}|^{2+\delta}}{n^{\delta/2}[\mathrm{var}(Z_{n1})]^{1+\delta/2}} \le C_2 \frac{\left(\ln \frac{1}{h} \right)^{(2+\delta)l}}{n^{\delta/2} h^{1+\delta/2}} \qquad \textbf{(A.73)}$$

by (A.72) and Lemma 14(b) and the right-hand side of (A.73) tends to zero with $h = d(\gamma \ln n)^{-1/\beta}$ and $\delta > 0$. This establishes

$$\frac{A_{2n}^*}{\sigma_{n3}(x)} \Rightarrow N(0,1). \qquad \textbf{(A.74)}$$

On the other hand, by arguments similar to the proof of Lemma 16, we have

$$\left|\frac{A_{2n}^{**}}{\sigma_{n3}(x)}\right| \le \frac{n^{-(1+\alpha)/2}}{\sigma_{n3}(x)} \cdot O_p\left(h^{\beta(l+1)+(r-1)\beta_0+\beta_1-1}\left(\ln\frac{1}{h}\right)^l \exp\left[\{a_0(r-1)+a_1 r^\beta\}\left(\frac{d}{h}\right)^\beta\right]\right)$$
$$= O_p\left(n^{-\alpha/2}h^{\beta_1-1/2}\left(\ln\frac{1}{h}\right)^l \exp\left[a_1\left(\frac{d}{h}\right)^\beta\right]\right)$$
$$= O_p(n^{a_1\gamma-\alpha/2}) \xrightarrow{p} 0 \tag{A.75}$$

and

$$\left|\frac{A_{2n}^{***}}{\sigma_{n3}(x)}\right| \le \frac{n^{-\alpha/2}}{\sigma_{n3}(x)} \cdot O_p\left(h^{\beta(l+1)-\beta_0}\left(\ln\frac{1}{h}\right)^l \exp\left[\{-a_0+a_1(r^\beta-1)\}\left(\frac{d}{h}\right)^\beta\right]\right)$$
$$= O_p\left(n^{1/2}n^{-\alpha/2}h^{-\beta_0-(r-1)\beta+1/2}\left(\ln\frac{1}{h}\right)^l \exp\left[-a_0 r\left(\frac{d}{h}\right)^\beta\right]\right)$$
$$= O_p(n^{1/2}n^{-\alpha/2}n^{-a_0 r\gamma}) \xrightarrow{p} 0. \tag{A.76}$$

Now (A.6) follows from (A.74)–(A.76).

Finally, we verify (A.7). Consider the expression (A.26). We have

$$|A_{3n}| \le C_1\int_{-\infty}^{\infty}\frac{|\phi_K(th)||\hat{\varphi}(t)|}{|\hat{\phi}^w(t)|^{2-1/r}|\varphi(rt)|^2}\,dt\cdot\sup_{t\in\mathbb{R}}|\hat{\phi}_{\overline{X}}(t)-\phi_{\overline{X}}(t)|^2$$
$$+ C_2\int_{-\infty}^{\infty}\frac{|\phi_K(th)||\hat{\varphi}(t)||\hat{\phi}_{\overline{X}}(t)|^2}{|\hat{\phi}^w(t)|^{2-1/r}|\hat{\varphi}(rt)|^2|\varphi(rt)|^2}\,dt\cdot\sup_{t\in\mathbb{R}}|\hat{\varphi}(t)-\varphi(t)|^2$$
$$\le O_p\left(n^{-1}h^{\beta(l+1)+\beta_0(2r-1)-1}\left(\ln\frac{1}{h}\right)^l \exp\left[\{a_0(2r-1)+a_1(r^\beta-1)\}\left(\frac{d}{h}\right)^\beta\right]\right)$$
$$+ O_p\left(n^{-\alpha}h^{\beta(l+1)+\beta_1-\beta_0-1}\left(\ln\frac{1}{h}\right)^l \exp\left[\{-a_0+a_1(2r^\beta-1)\}\left(\frac{d}{h}\right)^\beta\right]\right). \tag{A.77}$$

Therefore, this implies

$$\left|\frac{A_{3n}}{\sigma_{n3}(x)}\right| = O_p(n^{a_0 r\gamma-1/2}) + O_p(n^{[a_1 r^\beta - a_0 r]\gamma - \alpha + 1/2}) \xrightarrow{p} 0. \tag{A.78}$$

Now the proof of Theorem 7 is complete. ■

Proof of Lemma 8. The proof of Lemma 8 is similar to that of Lemma 5 except that we now have for each $\delta > 0$

$$\frac{E|Z_{n1}|^{2(1+\delta)}}{(\varepsilon n)^\delta[EZ_{n1}^2]^{1+\delta}} = O\left(\frac{\left(\ln\frac{1}{h}\right)^{2(1+\delta)l}}{n^\delta h^{1+\delta}}\right) \to 0, \tag{A.79}$$

where the equality follows from Lemmas 14 and 15 and the convergence to zero holds by using the fact that $h = d(\gamma \ln n)^{-1/\beta}$ for some $\gamma > 0$. ■

The proof of Theorem 10 uses the following lemmas.

LEMMA 16. *Under Assumptions D(i)–(iv),*

(a) we have as $h \to 0$

$$\sup_{x\in\mathbb{R}} \left| \int_{\mathcal{Y}} G_n(x,y)\,dy \right| = O\left(h^{\rho(m+1)+(r-1)\rho_0} \left(\ln \frac{1}{h}\right)^m \exp\left[b^* \left(\frac{d}{h}\right)^\rho \right] \right);$$

(b)

$$\sup_{x\in\mathbb{R}} \left| \int_{\mathcal{Y}} y G_n(x,y)\,dy \right| = O\left(h^{\rho(m+1)+(r-1)\rho_0} \left(\ln \frac{1}{h}\right)^m \exp\left[b^* \left(\frac{d}{h}\right)^\rho \right] \right);$$

and

(c) if moreover Assumptions D(v) and (vi) hold, then we have

$$\left| \int_{\mathcal{Y}} G_n(x,y)\,dy \right| \ge D_5 H(x) h^{\rho(m+1)+(r-1)\rho_0} \exp\left[b^* \left(\frac{d}{h}\right)^\rho \right]$$

for some D_5 uniformly in x on a bounded interval, where

$$H(x) = \begin{cases} \left| \int_{\mathcal{Y}} \cos(d(x+y))\,dy \right|, & \text{if } I^*(s,t) = o(R^*(s,t)) \\ \left| \int_{\mathcal{Y}} \sin(d(x+y))\,dy \right|, & \text{if } R^*(s,t) = o(I^*(s,t)). \end{cases}$$

LEMMA 17. *Under Assumption D, we have for large n*

(a)

$$\operatorname{var}(Z_{n1}) \le D_6 h^{2[\rho(m+1)+(r-1)\rho_0-1]} \left(\ln \frac{1}{h}\right)^{2m} \exp\left[2b^* \left(\frac{d}{h}\right)^\rho \right]$$

and

(b)

$$\operatorname{var}(Z_{n1}) \ge D_7 h^{2[\rho(m+1)+(r-1)\rho_0]-1} \exp\left[2b^* \left(\frac{d}{h}\right)^\rho \right]$$

for some positive constants D_6 and D_7.

Proof of Lemma 16. We prove Lemma 16 by adapting the proof of Lemma 3.1 of Fan and Masry (1992). Let

$$\tau = \lambda h^\rho \ln \frac{1}{h}, \tag{A.80}$$

where λ is a positive constant. Let

$S(a,b) = \{(s,t) \in \mathbb{R}^2 : a \le \|(s,t)\| \le b\}$

denote an index set for some $a \ge 0$ and $b \ge 0$.

We first establish part (a). We have

$$\int_{\mathcal{Y}} G_n(x,y)dy$$

$$\le \frac{1}{(2\pi)^2 r}\int_{\mathcal{Y}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\tilde{\phi}_K(s,t)\varphi\left(\frac{s}{h},\frac{t}{h}\right)}{\left[\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right]^{r-1}\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)}dsdtdy$$

$$= \frac{1}{(2\pi)^2 r}\int_{\mathcal{Y}}\left\{\left(\iint_{S(0,d-\tau)}+\iint_{S(d-\tau,d)}\right)\frac{\tilde{\phi}_K(s,t)\varphi\left(\frac{s}{h},\frac{t}{h}\right)}{\left[\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right]^{r-1}\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)}dsdt\right\}dy$$

$$\equiv \frac{1}{(2\pi)^2 r}(I_1+I_2). \qquad \textbf{(A.81)}$$

First, consider I_1. Let M be a large constant. We have

$$I_1 = \int_{\mathcal{Y}}\left\{\left(\iint_{S(0,Mh)}+\iint_{S(Mh,d-\tau)}\right)\frac{\tilde{\phi}_K(s,t)\varphi\left(\frac{s}{h},\frac{t}{h}\right)}{\left[\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right]^{r-1}\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)}dsdt\right\}dy$$

$$\le C_1\frac{h^2}{\min_{S(0,M)}|\phi_{Y_0,X_0}(s,t)|^{r-1}\min_{S(0,rM)}|\varphi(s,t)|^{r-1}}$$

$$+ C_2\iint_{S(Mh,d-\tau)}\left\|\left(\frac{s}{h},\frac{t}{h}\right)\right\|^{-\rho_0(r-1)}\exp\left[b^*\left\|\left(\frac{s}{h},\frac{t}{h}\right)\right\|^{\rho}\right]dsdt$$

$$\le C_3 h^{\rho_0(r-1)}\iint_{S(Mh,d-\tau)}\|(s,t)\|^{-\rho_0(r-1)}\exp[b^* h^{-\rho}\|(s,t)\|^{\rho}]dsdt$$

$$= O\left(h^{\rho_0(r-1)}\exp\left[b^*\left(\frac{d}{h}\right)^{\rho}\left(1-\frac{\tau}{d}\right)^{\rho}\right]\right)$$

$$= O\left(h^{\rho_0(r-1)+b^*\rho\lambda d^{\rho-1}}\exp\left[b^*\left(\frac{d}{h}\right)^{\rho}\right]\right), \qquad \textbf{(A.82)}$$

where the first inequality holds by Assumption D(i); the second inequality holds by Assumption D(ii) and the second equality follows because the integrand in the right-hand side of the second inequality is an increasing function of $\|(s,t)\|$ and is bounded by its value at the point $d-\tau$; and the last equality follows by a Taylor expansion of $(1-\tau/d)^{\rho}$ around 1. Next, we consider I_2. We have

$$
\begin{aligned}
I_2 &\le C_1 \iint_{S(d-\tau,d)} (d - \|(s,t)\|)^m \left\|\left(\frac{s}{h},\frac{t}{h}\right)\right\|^{-\rho_0(r-1)} \exp\left[b^* \left\|\left(\frac{s}{h},\frac{t}{h}\right)\right\|^{\rho}\right] dsdt \\
&\le C_2 \tau^m h^{\rho_0(r-1)} \iint_{S(d-\tau,d)} \|(s,t)\|^{\rho-2} \exp\left[b^* \left\|\left(\frac{s}{h},\frac{t}{h}\right)\right\|^{\rho}\right] dsdt \\
&= O\left(h^{\rho_0(r-1)+\rho(m+1)} \left(\ln\frac{1}{h}\right)^m \exp\left[b^*\left(\frac{d}{h}\right)^{\rho}\right]\right), \qquad \textbf{(A.83)}
\end{aligned}
$$

where the first inequality holds by Assumptions D(i) and (iv) and the second inequality holds because $(d - \|(s,t)\|)^m \le \tau^m$ and $\|(s,t)\|^{-\rho_0(r-1)-(\rho-2)} < C_3$ for $(s,t) \in S(d-\tau,d)$. By choosing a large value of the constant λ, the upper bound of I_2 dominates I_1. Thus part (a) of Lemma 16 is established. The proof of part (b) is similar.

We next establish part (c). We first write

$$
\begin{aligned}
&\int_{\mathcal{Y}} G_n(y,x)dy \\
&= \frac{1}{(2\pi)^2 r}\left\{\int_{\mathcal{Y}}\left(\iint_{S(0,d-\tau)} + \iint_{S(d-\tau,d)}\right)\right. \\
&\qquad \left. \times \exp(\mathbf{i}(sy+tx)) \frac{\tilde{\phi}_K(s,t)\varphi\left(\frac{s}{h},\frac{t}{h}\right)}{\left[\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right]^{r-1}\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)} dsdt\right\} dy \\
&\equiv J_1 + J_2. \qquad \textbf{(A.84)}
\end{aligned}
$$

By (A.82), we have

$$
|J_1| \le I_1 = O\left(h^{\rho_0(r-1)+b^*\rho\lambda d^{\rho-1}} \exp\left[b^*\left(\frac{d}{h}\right)^{\rho}\right]\right). \qquad \textbf{(A.85)}
$$

By symmetry of $\tilde{\phi}_K(s,t)$ (Assumption D(vi)), we have

$$
\begin{aligned}
J_2 = \frac{1}{(2\pi)^2 r}\int_{\mathcal{Y}} &\left\{\left(\iint_{S(d-\tau,d)}\right)\right. \\
&\times \tilde{\phi}_K(s,t)\left[\cos(sy+tx)\frac{R^*\left(\frac{s}{h},\frac{t}{h}\right)\left|\varphi\left(\frac{s}{h},\frac{t}{h}\right)\right|^2}{\left|\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right|^{2(r-1)}\left|\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)\right|^2}\right. \\
&\left.\left. + \sin(sy+tx)\frac{I^*\left(\frac{s}{h},\frac{t}{h}\right)\left|\varphi\left(\frac{s}{h},\frac{t}{h}\right)\right|^2}{\left|\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right|^{2(r-1)}\left|\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)\right|^2}\right] dsdt\right\} dy.
\end{aligned}
\qquad \textbf{(A.86)}
$$

Without loss of generality, we consider only the case $I^*(s/h, t/h) = o(|R^*(s/h, t/h)|)$. In this case, we have

$$J_2 = \frac{1}{(2\pi)^2 r} \int_{\mathcal{Y}} \left\{ \iint_{S(d-\tau,d)} \tilde{\phi}_K(s,t)\cos(sy+tx) \times \frac{R^*\left(\frac{s}{h},\frac{t}{h}\right)\left|\varphi\left(\frac{s}{h},\frac{t}{h}\right)\right|^2}{\left|\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right|^{2(r-1)}\left|\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)\right|^2} dsdt \right\} dy(1+o(1))$$

$$= \frac{1}{(2\pi)^2 r} \int_{\mathcal{Y}} \left\{ \left(\iint_{S(d-\tau,d-h^\rho)} + \iint_{S(d-h^\rho,d)} \right) \times \left[\tilde{\phi}_K(s,t)\cos(sy+tx) \frac{R^*\left(\frac{s}{h},\frac{t}{h}\right)\left|\varphi\left(\frac{s}{h},\frac{t}{h}\right)\right|^2}{\left|\phi_{Y_0,X_0}\left(\frac{s}{h},\frac{t}{h}\right)\right|^{2(r-1)}\left|\varphi\left(\frac{rs}{h},\frac{rt}{h}\right)\right|^2} \right] dsdt \right\} dy$$

$$\equiv J_2^a + J_2^b. \qquad \textbf{(A.87)}$$

Note that $R^*(s/h, t/h)$ cannot change its sign for $\|(s,t)\| \in S(d-\tau, d)$. (Otherwise, $R^*(s/h, t/h)$ would have a root, say, $(s^*/h, t^*/h)$, which implies that $[\phi_{Y_0,X_0}(s^*,t^*)]^{r-1}\varphi(rs^*, rt^*)/\varphi(s^*,t^*) = R^*(s^*/h, t^*/h) + i\ I^*(s^*/h, t^*/h) = 0$ and contradicts Assumption D(ii).) Also, by Assumption D(v), $\tilde{\phi}_K(s,t) > 0$ for $\|(s,t)\| \in (d-\delta, d)$. Note also that $\cos(sy+tx)$ cannot change its sign on $S(d-\tau,d)$, because $\cos(sy+tx) = \cos(d(y+x))(1+o(1))$ uniformly in y and x on $S(d-\tau,d)$. These imply that J_2^a and J_2^b have the same signs, say, positive. Therefore, $|J_2| \geq |J_2^b|$. By Assumptions D(i) and (v), we have

$$|J_2| \geq C_1 \left| \int_{\mathcal{Y}} \cos(d(y+x))dy(1+o(1)) \right| \times \iint_{S(d-h^\rho,d)} \left\{ (d-\|(s,t)\|)^m \left\| \left(\frac{s}{h},\frac{t}{h}\right) \right\|^{-\rho_0(r-1)} \exp\left[b^* \left\| \left(\frac{s}{h},\frac{t}{h}\right) \right\|^\rho \right] \right\} dsdt$$

$$\geq C_2 \left| \int_{\mathcal{Y}} \cos(d(y+x))dy \right| \left(\frac{d-h^\rho}{h} \right)^{-\rho_0(r-1)} \exp\left[b^* \left(\frac{d-h^\rho}{h} \right)^\rho \right] \times \iint_{S(d-h^\rho,d)} (d-\|(s,t)\|)^m dsdt$$

$$\geq C_3 \left| \int_{\mathcal{Y}} \cos(d(y+x))dy \right| h^{\rho_0(r-1)+(m+1)\rho} \exp\left[b^* \left(\frac{d}{h}\right)^\rho \left(1-\frac{h^\rho}{d}\right)^\rho \right], \qquad \textbf{(A.88)}$$

where the second inequality follows from the fact that the function $f(z) = z^{-\rho_0(r-1)}\exp(b^* h^{-\rho} z^\rho)$ is increasing in z when $z \in (d - h^\rho, d)$. Using the fact that $(1-z)^\rho \geq 1 - \rho z/2$ for small z, we have

$$J_2 \geq C_4 \left| \int_{\mathcal{Y}} \cos(d(y+x))dy \right| \cdot h^{\rho_0(r-1)+(m+1)\rho} \exp\left[b^* \left(\frac{d}{h}\right)^\rho \right]. \quad \textbf{(A.89)}$$

This together with (A.84) and (A.85) gives the desired lower bound in part (c) by choosing a large value of λ so that J_2 dominates J_1. ■

Proof of Lemma 17. Consider (A.56). Part (a) holds because we have, by Lemma 16(a),

$$\begin{aligned} \mathrm{var}(Z_{n1}) &\leq C_1 h^{-2} \sup_{x\in\mathbb{R}} |K_{n1}(x)|^2 \int_{-\infty}^{\infty} v_{\overline{X}}(x) f_X(x)dx \\ &= O\left(h^{2[\rho(m+1)+(r-1)\rho_0 - 1]} \left(\ln\frac{1}{h}\right)^{2m} \exp\left[2b^* \left(\frac{d}{h}\right)^\rho \right] \right). \end{aligned} \quad \textbf{(A.90)}$$

Part (b) follows using arguments similar to those in the proof of Lemma 16(c). ■

Proof of Theorem 10. To prove Theorem 10, it suffices to verify the following conditions:

$$\frac{B_{2n}}{\sigma_{n4}(x)} \Rightarrow N(0,1); \quad \textbf{(A.91)}$$

$$\frac{B_{1n}^*}{\sigma_{n4}(x)} \xrightarrow{p} 0; \quad \textbf{(A.92)}$$

$$\frac{B_{3n}}{\sigma_{n4}(x)} \xrightarrow{p} 0; \quad \textbf{(A.93)}$$

$$\frac{A_{2n}}{\sigma_{n4}(x)} \xrightarrow{p} 0; \quad \textbf{(A.94)}$$

$$\frac{A_{3n}}{\sigma_{n4}(x)} \xrightarrow{p} 0. \quad \textbf{(A.95)}$$

By Lemma 16, for n sufficiently large, we have

$$\begin{aligned} E|Z_{n1}|^{2+\delta} &\leq C_1 \frac{1}{h^{2+\delta}} \sup_{x\in\mathbb{R}} |K_{n1}(x)|^{2+\delta} \\ &= O\left(h^{[\rho(m+1)+(r-1)\rho_0 - 1](2+\delta)} \left(\ln\frac{1}{h}\right)^{(2+\delta)m} \exp\left[b^*(2+\delta) \left(\frac{d}{h}\right)^\rho \right] \right). \end{aligned} \quad \textbf{(A.96)}$$

Because EZ_{n1} is $O(1)$ by (A.55), the Lyapunov condition holds because

$$\frac{E|Z_{n1} - EZ_{n1}|^{2+\delta}}{n^{\delta/2}[\mathrm{var}(Z_{n1})]^{1+\delta/2}} \leq C_2 \frac{\left(\ln\frac{1}{h}\right)^{m(2+\delta)}}{n^{\delta/2} h^{1+\delta/2}} \to 0. \quad \textbf{(A.97)}$$

Therefore, (A.91) is established.

Now, (A.92)-(A.95) hold because some calculation yields

$$\left|\frac{B_{1n}^*}{\sigma_{n4}(x)}\right| \le O_p\left(n^{1/2}n^{-\alpha/2}h^{-r\rho_0-3/2}\left(\ln\frac{1}{h}\right)^m \exp\left[-\{b_0 r + b_1(r^\rho - 1)\}\left(\frac{d}{h}\right)^\rho\right]\right)$$

$$= O_p(n^{1/2}n^{-\alpha/2}n^{-\{b_0 r + b_1(r^\rho-1)\}\gamma}) \xrightarrow{p} 0, \tag{A.98}$$

$$\left|\frac{B_{3n}}{\sigma_{n4}(x)}\right| \le O_p\left(n^{-1/2}h^{r\rho_0-3/2}\left(\ln\frac{1}{h}\right)^m \exp\left[b_0 r\left(\frac{d}{h}\right)^\beta\right]\right)$$

$$+ O_p\left(n^{1/2-\alpha}h^{\rho_1-(r-2)\rho_0-3/2}\left(\ln\frac{1}{h}\right)^m \exp\left[(b_1 r^\rho - b_0 r)\left(\frac{d}{h}\right)^\rho\right]\right)$$

$$= O_p(n^{b_0 r\gamma - 1/2}) + O_p(n^{(b_1 r^\rho - b_0 r)\gamma + 1/2 - \alpha}) \xrightarrow{p} 0. \tag{A.99}$$

Similarly,

$$\left|\frac{A_{2n}}{\sigma_{n4}(x)}\right| = O_p(n^{(a^*-b^*)\gamma}) + O_p(n^{(a^*-b^*+a_1)\gamma-\alpha/2})$$

$$+ O_p(n^{(a^*-b^*+a_0 r)\gamma+(1-\alpha)/2}) \xrightarrow{p} 0 \tag{A.100}$$

and

$$\left|\frac{A_{3n}}{\sigma_{n4}(x)}\right| = O_p(n^{(a^*-b^*+a_0 r)\gamma-1/2}) + O_p(n^{(a^*-b^*+a_1 r^\beta - a_0 r)\gamma+1/2-\alpha}) \xrightarrow{p} 0. \tag{A.101}$$

Now the proof of Theorem 10 is complete. ■

Proof of Lemma 11. Similar to the proof of Lemma 8. ■