Higher Order Asymptotic Theory for Semiparametric Averaged Derivatives

by

Yoshihiko Nishiyama

Thesis Submitted for Ph.D. Degree

2000

The London School of Economics

and Political Science

Abstract

This thesis investigates higher order asymptotic properties of a semiparametric averaged derivative estimator. Classical parametric models assume that we know the distribution function of random variables of interest up to finite dimensional parameters, while nonparametric models do not assume this knowledge. Parametric estimators typically enjoy $\sqrt{n}$ - consistency and asymptotic normality under certain conditions, while nonparametric estimators converge to the true functionals of interest slower than parametric ones. Semiparametric estimators, a compromise between the two, have been intensively studied since the 1970s. Some of them have been shown to have the same convergence rate as parametric estimators despite involving nonparametric functional estimates. Semiparametric methods often suit econometrics because economic theory typically does not provide the whole information on economic variables which parametric methods require, and a sample of very large size is rarely available in econometrics. This thesis treats a semiparametric averaged derivative estimator of single index models. Its first order asymptotic theory has been studied since late 1980s. It has been shown to be $\sqrt{n}$ - consistent and asymptotically normally distributed under certain regularity conditions despite involving a nonparametric density estimate. However its higher order properties could be affected by the property of nonparametric estimates. We obtain valid Edgeworth expansions for both normalized and studentized estimators, and moreover show the bootstrap distribution approximates the exact distribution of the estimator asymptotically as well as the Edgeworth expansion for the normalized statistics. We propose optimal bandwidth choices which minimize the normal approximation error using the expansion. We also examine the finite sample performance of the Edgeworth expansions by a Monte Carlo study.

Contents

List of Tables and Figures

Tables

Figures

Preface

This thesis aims to study higher order asymptotic properties of certain semiparametric statistics. Semiparametric methods have recently been paid much attention and are currently one of the main topics in econometrics. The results in this thesis are quite limited in the sense that it deals only one estimator for single index models, however it seems possible to examine other semiparametric statistics with U statistic form in a similar manner.

In preparing this thesis, I was given helpful comments and suggestions by many people at different stages. Among them I would like to acknowledge the following individuals: Kimio Morimune, Peter M. Robinson, and Yoshihiro Yajima. Especially, I am most grateful to Peter M. Robinson who was my supervisor at LSE and suggested me to try this topic when I started research for Ph.D. His stimulating discussions and critical comments have led significant improvement of this thesis. I am very proud of having been one of his students. Javier Hidalgo gave me some useful comments in my preparing Chapter 2. I also would like to acknowledge Professor Yoshihiro Yajima for his comments at the meeting of the Japan Association of Economics and Econometrics. I would like to give my special thanks to Kimio Morimune who first taught me econometrics and was my supervisor at Kyoto University. Without him I might not have chosen econometrics as my subject and might not have had this great opportunity to make research at LSE. Needless to say, all errors in this thesis are exclusively my own. This thesis is dedicated to my parents and my wife Shuko.

December, 2000

YOSHIHIKO NISHIYAMA

Chapter 1.

Introduction

Single index models have been developed to analyze mainly limited dependent variables (LDV) models and some transformation models semiparametrically. The purpose of this thesis is to study higher order asymptotic properties of certain semiparametric estimator of a single index model. In the following section we review how parametric estimation fails in case of model misspecification, then introduce a semiparametric single index model. Section 1.1 reviews an estimator of its parameters called density-weighted averaged derivatives. This estimator is shown to be $\sqrt{n}$-consistent and asymptotically normally distributed even though it involves a nonparametric estimate with slower convergence rate than $n^{-1/2}$. This thesis investigates the higher order properties of the estimator focusing on the point if parametric rate of Edgeworth expansion is possible. Sections 1.2 and 1.3 respectively explain other semiparametric estimation methods and established higher order asymptotic theory related to the estimator of interest. Chapter 2 validates Edgeworth expansions for the density-weighted averaged derivatives suitably normalized, while Chapter 3 derives valid Edgeworth expansions for the studentized statistic. We further propose optimal bandwidth choices minimizing the normal approximation error based on the validated Edgeworth expansion. Chapter 4 compares the bootstrap distribution to the Edgeworth expansion derived in Chapter 2. Chapter 5 provides a Monte Carlo study based on a Tobit model.

1.1 Parametric estimation and averaged derivative estimation for single index model

1.1.1 Parametric regression model and single index model

One of the main interest in econometrics is the mean response of certain economic random variables to others. Given two vectors of random variables X and Y, if we can consider that the value of Y is determined depending on the value of X, we would like to make inference on the regression function,

$$E(Y|X) = g(X) \text{ almost surely (a.s.)}$$

where X and Y are called the independent (or explanatory) variable and dependent (or explained) variable respectively. Let Y be a scalar and X be a $d \times 1$ vector. Suppose

(i) we have independently and identically distributed (iid) observations (X_i^τ, Y_i), $i=1,2,...$ of (X^τ, Y) , τ denoting transpose. Note that it implies

(i)' (X_i^τ, ϵ_i) are also iid where $\epsilon_i = Y_i - g(X_i)$, $i=1,2,$ .....

We have by the construction of ϵ,

(ii) $E(\epsilon_1|X_1)=0$ a.s.

Assume also

(iii) $Var(\epsilon_1|X_1) = \sigma^2 < \infty$ a.s.

and

(iv) $Var(X_1) = V_X$ is finite and positive definite. This implies $\mu_X \equiv E(X_1)$ exists.

Supposing

(v) $g(x)$ is linear in x,

namely $g(x) = \beta^\tau x$, classical statistical theory has considered inference on β. Ordinary least squares (OLS) method $\min_\beta \sum_{i=1}^n (Y_i - \beta^\tau X_i)^2$ provides a very satisfactory point estimate of β. If $rank(X) = d$, this problem has a nice closed form solution

$$\beta_{OLS} = (X^\tau X)^{-1} X^\tau Y$$

where $X = (X_1, X_2, \cdots, X_n)^\tau$ and $Y = (Y_1, Y_2, \cdots, Y_n)^\tau$, and it is known to be the best linear unbiased, consistent and asymptotically normally distributed under (i)-(v). The proof is based on the equality

$$\beta_{OLS} = (X^\tau X)^{-1} X^\tau Y = (X^\tau X)^{-1} X^\tau (X\beta + \epsilon) = \beta + (X^\tau X)^{-1} X^\tau \epsilon \tag{1.1}$$

where $\epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^\tau$. When $E\|\beta_{OLS}\| < \infty$ [1], $\|\cdot\|$ denoting Euclidian norm, unbiasedness is straightforward by (i)' and (ii) because

$$E(\beta_{OLS}) = \beta + E[E\{(X^\tau X)^{-1} X^\tau \epsilon | X\}] = \beta + E[(X^\tau X)^{-1} X^\tau E(\epsilon | X)] = \beta \ .$$

(i), (iv) and Khintchine's weak law of large numbers (WLLN) yield

$$plim \frac{1}{n} X^\tau X = Q \tag{1.2}$$

where $Q = V_X + \mu_X \mu_X^\tau$ is finite and positive definite due to (iv). $plim \frac{1}{n} X^\tau \epsilon = 0$ by (i)', (ii)-(iv) and Khintchine's WLLN. Therefore we have

$$plim \beta_{OLS} = \beta + (plim \frac{1}{n} X^\tau X)^{-1} plim \frac{1}{n} X^\tau \epsilon = \beta \tag{1.3}$$

due to Slutsky's theorem. Because of (1.2) and $\frac{1}{\sqrt{n}} X^\tau \epsilon \xrightarrow{d} N(0, \sigma^2 Q)$ by (i)', (ii)-(iv), Lindeberg-Levy's central limit theorem and Cramèr-Wold device, we have

$$\sqrt{n}(\beta_{OLS} - \beta) = (\frac{1}{n} X^\tau X)^{-1} \frac{1}{\sqrt{n}} X^\tau \epsilon \xrightarrow{d} N(0, \sigma^2 Q^{-1}) \ . \tag{1.4}$$

If furthermore ϵ_i are normally distributed, β_{OLS} coincides with the maximum likelihood (ML) estimator, so that it is efficient in the sense its variance attains Cramèr-Rao lower bound.

As far as (i)-(v) are satisfied, the OLS estimate enjoys the above desirable properties. Some modified methods have been proposed when they are not satisfied. In case (ii) is violated, we can apply the instrumental variable (IV) method. If (iii)

---

[1] A sufficient condition is $E[\lambda_{\min}^{-1}(X^\tau X)] < \infty$ where $\lambda_{\min}(A)$ is the minimum eigenvalue of A.

is violated, OLS still provides a consistent estimate, but it is no longer efficient, when generalized least squares (GLS) method can be used to derive an efficient estimate. The asymptotic distribution in case of heteroscedasticity is typically normal or some Gaussian functional depending on the data generating process (DGP). We also have enormous research on regression analysis of correlated variables.

When (v) is violated, however, we face difficult problems to resolve. Suppose $g(x)$ is nonlinear in x in fact but we are ignorant of it, then OLS gives

$$\beta_{OLS} = (X^\tau X)^{-1}X^\tau Y = (X^\tau X)^{-1}X^\tau G + (X^\tau X)^{-1}X^\tau \epsilon$$

where $G = (g(X_1),...,g(X_n))^\tau$, so that roughly speaking we can think that β_{OLS} estimates quantities such as $E[(X^\tau X)^{-1}X^\tau G]$ or $(plim\frac{1}{n}X^\tau X)^{-1}plim\frac{1}{n}X^\tau G$ but these would not be what we want to estimate in regression analysis.[2] Thus OLS estimation will collapse under misspecification in regression function. In econometrics, we easily face this situation. Important examples include limited dependent variable (LDV) models such as:

1) censored regression model (type 1 Tobit model)

$$: \quad Y_i = (\beta^\tau X_i + u_i)\, I(\beta^\tau X_i + u_i \geq 0) \ . \tag{1.5}$$

2) Truncated regression model

$$: \quad Y_i = \beta^\tau X_i + u_i \ , \ \text{ but } Y_i \text{ are observable only when } \beta^\tau X_i + u_i \geq 0 \ . \tag{1.6}$$

3) Probit model : $Y_i = I(\beta^\tau X_i + u_i \geq 0)$, $u_i|X_i \sim iidN(0,\sigma^2)$. (1.7)

4) Logit model : $Y_i = I(\beta^\tau X_i + u_i \geq 0)$, $u_i|X_i \sim iid\ logistic$. (1.8)

5) Transformation models such as

$$\text{Box-Cox} \quad : \quad h_\lambda(Y_i) = \beta^\tau X_i + u_i, \ h_\lambda(Y_i) = \begin{cases} \log Y_i, & \lambda = 0 \\ \dfrac{Y_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases} \ . \tag{1.9}$$

---

[2] We can always calculate β_{OLS} formally as long as $|X^\tau X| \neq 0$ even if it does not make much sense theoretically. We might think it gives a rough relation between Y and X as a kind of descriptive statistics.

Here $I(.)$ is the indicator function and u_i are iid disturbance terms. Letting $F(.)$ be the distribution function of u_i conditionally on X_i , the regression functions for the above models are

censored regression model

$$: \; E(Y_i|X_i) = \beta^\tau X_i\{1 - F(-\beta^\tau X_i)\} + \int_{-\beta^\tau X_i}^{\infty} u\, dF(u) \;\text{ a.s.,} \qquad (1.10)$$

truncated regression model

$$: \; E(Y_i|X_i) = \beta^\tau X_i + \frac{1}{1 - F(-\beta^\tau X_i)} \int_{-\beta^\tau X_i}^{\infty} u\, dF(u) \;\text{ a.s.,} \qquad (1.11)$$

Probit/Logit model : $E(Y_i|X_i) = 1 - F(-\beta^\tau X_i)$ a.s. (1.12)

where $F(.)$ is normal and logistic for Probit and Logit respectively,

$$\text{Box-Cox model : } \; E(Y_i|X_i) = \begin{cases} e^{\beta^\tau X_i} \int_{-\infty}^{\infty} e^{u}\, dF(u), & \lambda = 0 \\ \int_{-\infty}^{\infty} \{1 + \lambda(\beta^\tau X_i + u)^{1/\lambda}\} dF(u), & \lambda \neq 0 \end{cases} \;\text{ a.s.} \qquad (1.13)$$

The functional form of the regression function is nonlinear in X_i in each of the above cases so that the OLS does not work to estimate parameters in these models. Classical theory has developed two ways that may be able to handle nonlinear models. One is nonlinear least squares (NLS) estimation and the other is ML estimation. We explain them in terms of the Probit model (1.7). Exactly the same thinking is possible for the other models.

Since the disturbances u_i are assumed to be independently and identically normally distributed with mean zero and variance σ^2 in (1.7), (1.12) gives

$$E(Y_i|X_i) = 1 - \int_{-\infty}^{-\beta^\tau X_i} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{u^2}{2\sigma^2}) du = 1 - \Phi(-\frac{\beta^\tau X_i}{\sigma}) \;\text{ a.s.} \qquad (1.14)$$

where $\Phi(.)$ is the distribution function of a standard normal variate. Only $\alpha = \beta/\sigma$ is identifiable here. The NLS estimator is the solution to

$$\min_{\alpha} \sum_{i=1}^{n} \{ Y_i - [1 - \Phi(-\alpha^\tau X_i)]\}^2 \; .$$

It is known that the estimator is consistent and asymptotically normally distributed. However if u_i are not normally distributed in fact, it means corresponding regression function (1.14) is incorrect. Thus the NLS estimator becomes inconsistent similarly to the OLS estimation with incorrect regression form.

We implement the ML estimation as follows. Under the assumption of normal disturbances conditionally on regressors, we construct the conditional probability function of Y_i given X_i

$$P(Y_i = y | X_i; \beta, \sigma^2) = F_i^y(1 - F_i)^{1-y},\ y = 0, 1$$

where $F_i = P(Y_i = 1 | X_i; \beta, \sigma^2) = \Phi(-\frac{\beta^T X_i}{\sigma}) = \Phi(-\alpha^T X_i)$. α is identifiable again. The ML estimator for α is defined by

$$\alpha_{ML} = \underset{\alpha}{argmax}\, LL(\alpha)$$

where

$$LL(\alpha) = \sum_{i=1}^{n} \{Y_i \log F_i + (1 - Y_i)\log(1 - F_i)\} \quad , \tag{1.15}$$

is the log-likelihood function. The first order condition for the maximization is

$$s(\alpha_{ML}) = \sum_{i=1}^{n} s_i(\alpha_{ML}) = 0 \tag{1.16}$$

where $s(\alpha) \equiv \dfrac{\partial LL(\alpha)}{\partial \alpha}$ is the score function of the whole sample and $s_i(\alpha) \equiv \dfrac{\partial}{\partial \alpha}\{Y_i \log F_i + (1-Y_i) \log (1-F_i)\}$ is the score associated with observation i. ML method is known in general to provide consistent, asymptotically normal and efficient estimates under certain regularity conditions (see e.g. Amemiya (1985, p.115-125)), so that it dominates the NLS in the current cases in terms of efficiency. We refer to Amemiya (1985, p.270-273) for the rigorous proof and conditions of consistency and asymptotic normality for the current Probit model. Here we would like to give only a heuristic explanation of the asymptotic properties. Letting α_0 be the true parameter value, mean value theorem gives

$$s(\alpha_{ML}) = s(\alpha_0) + \frac{\partial s(\alpha_0)}{\partial \alpha^\tau}(\alpha_{ML} - \alpha_0) + R$$

where R is a $d\times 1$ vector with kth element

$$R_k = (\alpha_{ML} - \alpha_0)^\tau \frac{\partial^2 s(\alpha^+)}{\partial \alpha^\tau \partial \alpha_{(k)}}(\alpha_{ML} - \alpha_0) \ , \ \alpha^+ = \theta\alpha_0 + (1-\theta)\alpha_{ML} ,\ 0<\theta<1 \ ,$$

$\alpha_{(k)}$ denoting the kth element of α. Supposing α_0 is bounded, α_{ML} is $O_p(1)$ and $\frac{1}{n}\frac{\partial^2 s(\alpha)}{\partial \alpha^\tau \partial \alpha_{(k)}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 s_i(\alpha)}{\partial \alpha^\tau \partial \alpha_{(k)}}$ is $O_p(1)$ uniformly in α, we have

$$s(\alpha_{ML}) = s(\alpha_0) + \frac{\partial s(\alpha_0)}{\partial \alpha^\tau}(\alpha_{ML} - \alpha_0) + O_p(n) \ . \tag{1.17}$$

(1.16) and (1.17) give

$$\alpha_{ML} - \alpha_0 = -\left\{\frac{1}{n}\frac{\partial s(\alpha_0)}{\partial \alpha^\tau} + O_p(1)\right\}^{-1}\frac{1}{n}s(\alpha_0) \tag{1.18}$$

if the inverse exists. Khintchine's WLLN gives

$$\frac{1}{n}s(\alpha_0) = \frac{1}{n}\sum_{i=1}^{n}s_i(\alpha_0) \xrightarrow{p} E\{s_1(\alpha_0)\} = 0 \tag{1.19}$$

and

$$-\frac{1}{n}\frac{\partial s(\alpha_0)}{\partial \alpha^\tau} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial s_i(\alpha_0)}{\partial \alpha^\tau} \xrightarrow{p} I(\alpha_0) \tag{1.20}$$

where $I(\alpha_0) = -E\{\frac{\partial^2 LL(\alpha_0)}{\partial \alpha \partial \alpha^\tau}\}$ is the information matrix. The last equality in (1.19) is because

$$\begin{aligned} E[s_1(\alpha_0)] &= E[E\{s_1(\alpha_0)|X_1\}] \\ &= E\left[\Phi(-\alpha_0^\tau X_1)\frac{\phi(-\alpha_0^\tau X_1)}{\Phi(-\alpha_0^\tau X_1)}X_1 - \{1-\Phi(-\alpha_0^\tau X_1)\}\frac{\phi(-\alpha_0^\tau X_1)}{1-\Phi(-\alpha_0^\tau X_1)}X_1\right] = 0 \end{aligned}$$

using (1.12) where $\phi(z)=d\Phi(z)/dz$. Thus we have $\alpha_{ML} \xrightarrow{p} \alpha_0$ by (1.18)-(1.20).[3] We next see the asymptotic normality. We now have $R_k = o_p(n)$ due to the consistency, so that we can modify (1.18) times $\sqrt{n}$ to

$$\sqrt{n}(\alpha_{ML} - \alpha_0) = -\left\{\frac{1}{n}\frac{\partial s(\alpha_0)}{\partial \alpha^\tau} + o_p(1)\right\}^{-1}\frac{1}{\sqrt{n}}s(\alpha_0) \ . \tag{1.21}$$

---

[3] For the rigorous proof of consistency, we examine if $n^{-1}LL(\alpha)$ converges uniformly to a nonstochastic function of α taking maximum at $\alpha=\alpha_0$, and other technical conditions on the objective function such as differentiability and measurability.

Since $s_i(\alpha_0)$ are iid with mean zero and variance $I(\alpha_0)$,

$$\frac{1}{\sqrt{n}}s(\alpha_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i(\alpha_0) \overset{d}{\to} N(0,I(\alpha_0)) \tag{1.22}$$

by Lindeberg-Levi's central limit theorem and Cramèr-Wold device. Slutsky's theorem and (1.20)-(1.22) give the asymptotic normality

$$\sqrt{n}(\alpha_{ML}-\alpha_0) \overset{d}{\to} N(0,I^{-1}(\alpha_0)) \ .$$

The above heuristic explanation suggests that (1.19) is vital for the consistency of ML estimators.

Similar calculation to (1.16)-(1.22) applies to the following log-likelihood functions corresponding to (1.5), (1.6), (1.8) and (1.9). The log-likelihood function of the transformation model under normal disturbances is

$$\log L(\beta,\lambda,\sigma^2) = -\frac{n}{2}(\log 2\pi + \log\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}h_\lambda'(Y_i)\{h_\lambda(Y_i) - \beta^\tau X_i\}^2 \ . \tag{1.23}$$

For the truncated regression model (1.6), assuming normality of u, the log-likelihood function is

$$\log L(\beta,\sigma^2) = -\frac{n}{2}(\log 2\pi + \log\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\beta^\tau X_i)^2 - \sum_{i=1}^{n}\log\{1-\Phi(-\beta^\tau X_i)\} \ . \tag{1.24}$$

Tobin (1958) first proposes an estimation method of (1.5) to analyze household purchases of durable goods so that it is called a Tobit model. Assuming normal disturbances, the log-likelihood function is

$$\log L(\beta,\sigma^2) = \sum_{Y_i>0}\left\{-\frac{1}{2}(\log 2\pi + \log\sigma^2) - \frac{1}{2\sigma^2}(Y_i-\beta^\tau X_i)^2\right\} + \sum_{Y_i=0}\log\{1-\Phi(\frac{\beta^\tau X_i}{\sigma})\} \ . \tag{1.25}$$

The first term on the right corresponds to the likelihood of the observations with positive dependent variables and the second term is for the observations with $Y=0$. Amemiya (1973) proves that this estimator is strongly consistent, asymptotically normally distributed and efficient. There is no closed form solution for the maximization of (1.15), (1.23), (1.24) and (1.25) so that we will need to maximize them numerically.

We briefly review related research to these. Heckman (1976) proposed a least squares type estimation for type 3 Tobit model mentioned later. The idea can be applied to estimate (1.5). Since

$$E(Y_i|Y_i>0) = \beta^T X_i + E(u_i|u_i>-\beta^T X_i) = \beta^T X_i + \sigma\frac{\phi_i}{\Phi_i} \text{ a.s.} \quad (1.26)$$

where σ^2 is the variance of the disturbance term, $\phi_i = \phi(\frac{\beta^T X_i}{\sigma})$ and $\Phi_i = \Phi(\frac{\beta^T X_i}{\sigma})$, we have

$$E(Y_i|X_i) = \Phi_i \beta^T X_i + \sigma\phi_i \text{ a.s.}$$

Thus after getting some estimates of ϕ_i and Φ_i , we can perform OLS estimation. Robinson (1982) considers ML estimation of the Tobit model when independence assumption is violated. He proves it is consistent and asymptotically normally distributed. Maddala and Nelson (1974) show how consistency of ML estimator does not hold any more in the existence of heteroscedasticity. A modification to adjust for the heteroscedasticity is in Maddala (1983). Other types of Tobit model are also considered:

Type 2 Tobit model : $Y = (\beta_0^T X_0 + \epsilon_0) I(\beta_1^T X_1 + \epsilon_1 \geq 0)$,

Type 3 Tobit model : $Z = (\beta_1^T X_1 + \epsilon_1) I(\beta_1^T X_1 + \epsilon_1 \geq 0), Y = (\beta_0^T X_0 + \epsilon_0) I(\beta_1^T X_1 + \epsilon_1 \geq 0)$.

Gronau (1973) and Heckman (1974) estimate wage function of female workers based on type 2 Tobit model by ML method, where the latter takes into account the determination of hours worked, while the former does not. LS type estimator is considered in Heckman (1979). Regarding type 3 Tobit model, Heckman (1976) proposes a two-step least squares estimation based on the similar transformation as

(1.26), while Amemiya (1978, 1979) extends the method to estimate all types of Tobit models based on their reduced forms by least squares or generalized least squares method. Comprehensive survey on LDV models can be found in Amemiya (1981, 1984, 1985) and Maddala (1983) among others.

We now discuss when and how ML estimation can fail. As seen in the above, (1.19) is crucial for the consistency of ML estimators. There are two sources of violation of (1.19) in the above models. One is when the functional relationship is misspecified in (1.5)-(1.9) and the other is when the conditional distribution of disturbances is misspecified. If, for example, the model specification (1.7) is correct, we have $E[s_1(\alpha_0)]=0$ as shown in the above. However, suppose it is incorrect but the correct relation between Y and X is in fact

$$Y_i = I(\,h(\beta^\tau X_i) + u_i \geq 0)\ \ ,\ \ h(x) \neq x\ . \tag{1.27}$$

Then $\frac{1}{n}s(\alpha_0)$ does not converge to zero in probability since the expected score is non-zero because $E(Y_i|X_i) = 1 - \Phi(-h(\beta^\tau X_i)/\sigma) \neq 1 - \Phi(-\alpha^\tau X_i)$, which means (1.18) does not converge to zero in probability. Therefore ML method does not give a consistent estimate in the case of misspecified relation between variables. The same thing happens when the functional specification in (1.5), (1.6), (1.8) or (1.9) is incorrect. Similarly misspecification of disturbance distribution results in inconsistent estimate in general due to the violation of (1.19). For example, Robinson (1982) illustrates how misspecified underlying distribution leads the expectation of score to non-zero and yields inconsistent estimates for a Tobit model.[4]

Taking into account that economic theory typically does not provide us the

---

[4] There are some exceptional cases like Gaussian pseudo maximum likelihood estimation for linear regression model, when misspecification of the disturbance distribution does not cause inconsistency under some conditions.

whole information necessary to construct likelihood functions[5], it is unlikely that we are sure of the specification like (1.5)-(1.9) with normal or some other specified disturbance distribution. Nonparametric and semiparametric methods have been developed to overcome these misspecification problems. There are many kinds of semiparametric regression models and their estimation procedures studied until now. Among them, semiparametric single index model is a useful alternative to some parametric models. It assumes Y depends on a $d \times 1$ vector X only through its linear combination without specifying the disturbance distribution and functional relationship between Y and $\beta^{\tau}X$, namely

$$E(Y|X) = g(X) = G(\beta^{\tau}X) \text{ a.s.} \tag{1.28}$$

where G: $R \rightarrow R$ is assumed unknown. This is a rather general model including linear regression model, (1.10)-(1.13) as well as (1.27) as special cases. Since ML estimation is not robust to model misspecification, the estimation based only on (1.28) without specification of $G(.)$ and the joint distribution of (Y,X) or conditional distribution of the error $Y - E(Y|X)$ given X may be more favourable than ML method. We note that models like type 2 and 3 Tobit are not nested in (1.28), however these are special cases of a multiple index model, a generalized version of the single index model, reviewed in Section 1.2.

1.1.2 Density-weighted averaged derivative estimation

Several semiparametric estimation methods for (1.28) have been proposed since late 1980's. One of them is the density-weighted averaged derivative (AD)

[5] It is especially unlikely that economic theory provides the true disturbance distribution.

estimator by Powell, Stock and Stoker (1989), Robinson (1989) and Cheng and Robinson (1994),

$$U \equiv \frac{2}{n}\sum_{i=1}^{n} \tilde{f}'(X_i)Y_i \tag{1.29}$$

where

$$\tilde{f}'(X_i) = \frac{1}{(n-1)h^{d+1}}\sum_{j\neq i}^{n} K'(\frac{X_i - X_j}{h}) \ , \tag{1.30}$$

h is a positive constant converging to zero as $n \to \infty$ and $K'(x) = \partial K(x)/\partial x$ for a differentiable function $K: R^d \to R$ satisfying $K(u) = -K(-u)$ and $\int K(u)du = 1$. This estimator has the following intuition. It is easily seen that β in (1.28) is identifiable only up to scale since $G(.)$ can be any function. For example $\beta = \beta_0$, $G(u) = u^2$ and $\beta = 2\beta_0$, $G(u) = u^2/4$ are equivalent with respect to (1.28). We could impose a normalization to make β identifiable, such as $\|\beta\| = 1$ (Härdle, Hall and Ichimura (1993)) or $\beta_1 = 1$ (Horowitz and Härdle(1996)) where β_1 is the leading element of β . AD estimation does not employ these sorts of explicit normalization. Assuming $g(x)$ differentiable, the AD is a nonparametric quantity defined by $E\{g'(X)\}$ which measures mean response of Y to marginal change in X. This is proportional to β since

$$\mu_1 \equiv E\{g'(X)\} = E\{G'(\beta^T X)\}\beta = c_0\beta \tag{1.31}$$

where $G'(u) = dG(u)/du$ and $c_0 = E[G'(\beta^T X)]$ is an unknown constant, so that estimation of AD means estimation of β up to a constant. Weighted averaged derivative

$$\mu_w \equiv E\{w(X)g'(X)\} \tag{1.32}$$

is also proportional to β for any weight function $w(.)$ since

$$E\{w(X)g'(X)\} = E\{w(X)G'(\beta^T X)\}\beta = c_1\beta \ . \tag{1.33}$$

Therefore estimates of weighted AD also estimate β up to scale. Suppose X is a continuous random vector with unknown density $f(x)$. When the density is used for the weight, μ_f is called the density-weighted average derivatives. Under the assumptions that $f(x)^2 g(x)$ diminishes on the boundary or more precisely $\lim_{|x|\to\infty} |f(x)^2 g(x)| = 0$ and the integrals in the following equation exist, we have

$$\mu_f \equiv E\{f(X)g'(X)\} = \int f(x)^2 g'(x)dx = -\int 2f'(x)f(x)g(x)dx$$
$$= -E\{2f'(X)g(X)\} = -2E\{f'(X)Y\} \qquad (1.34)$$

using iterated expectation in the last equality. The third equality is because

$$\int f(x)^2 \frac{\partial g(x)}{\partial x_i}dx = \int [f(x)^2 g(x)]_{x_i=-\infty}^{x_i=\infty} dx_{(-i)} - 2\int g(x)\frac{\partial f(x)}{\partial x_i}f(x)dx$$
$$= -2\int g(x)\frac{\partial f(x)}{\partial x_i}f(x)dx$$

due to Fubini's theorem, integration by parts and the above assumption on $f(x)^2 g(x)$, where $dx_{(-i)} = dx_1 \cdots dx_{i-1}dx_{i+1}\cdots dx_d$. We are concerned with an estimator of $\bar{\mu} \equiv -\mu_f = 2E\{f'(X)Y\}$ in this thesis which is a weighted AD with $w(x) = -f(x)$ and is also proportional to β. Supposing we have iid observations (X_i^T, Y_i), $i = 1,2,\ldots$, we may estimate $\bar{\mu}$ by its sample analogue $\frac{2}{n}\sum_{i=1}^{n} f'(X_i)Y_i$. It however involves unknown $f'(x)$ so that it is replaced by a consistent estimate $\tilde{f}'(x)$, then we obtain (1.29). We discuss the properties of $\tilde{f}'(x)$ in 1.1.4. Plugging (1.30) into (1.29) and some algebra using $K'(u) = -K'(-u)$ yield

$$U = \binom{n}{2}^{-1} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} U_{ij} \ , \quad U_{ij} = h^{-d-1}K'\left(\frac{X_j - X_i}{h}\right)(Y_j - Y_i) \ . \qquad (1.35)$$

This has a U-statistic form so that it is computationally less expensive than many other estimators involving nonlinear optimization such as ML estimators for LDV models in the above or semiparametric estimators seen in Section 1.2. There are some variants of AD type estimation depending on the choice of weight function in (1.32) and what estimate of $f'(x)$ is plugged in instead of (1.30).

The above authors have studied the asymptotic properties of U under various

DGPs. Small sample theory is virtually impossible because of unspecified underlying distribution. Powell, Stock and Stoker (1989) and Robinson (1989) prove the $\sqrt{n}$ - consistency and asymptotic normality for iid and weakly dependent observations respectively under certain assumptions on $g(.)$, $f(.)$ and its derivatives and function $K(.)$ including

$$\int_{R^d} u_1^{l_1} \cdots u_d^{l_d} K(u) du \begin{cases} = 1, & if \quad l_1+\cdots+l_d=0 \\ = 0, & if \ 0<l_1+\cdots+l_d<L \\ \neq 0, & if \quad l_1+\cdots+l_d=L \end{cases} \tag{1.36}$$

for some positive integer L. $K(.)$ used in (1.30) is called a kernel function and those satisfying (1.36) are called higher order kernels. L there is called the kernel order. Higher order kernels are originally developed to reduce asymptotic bias of kernel estimators of nonparametric functionals and first introduced to semiparametric framework by Robinson (1988b) for estimation of partially linear regression model. In the current estimation of $\bar{\mu}$, it is easy to show $EU - \bar{\mu} = O(h^L)$ (see (3.21) of Powell, Stock and Stoker), so that larger L reduces the asymptotic bias of U more since $h \rightarrow 0$ as $n \rightarrow \infty$. We see this also in a Monte Carlo study in Chapter 5. Some methods to construct higher order kernel function are found in e.g. Robinson (1988b) or Wand and Jones (1995, p.32). Applying the standard decomposition of U-statistics, we have

$$U - EU = \frac{2}{n}\sum_{i=1}^{n} (U_i - EU) + \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (U_{ij} - U_i - U_j + EU) \tag{1.37}$$

wher $U_i \equiv E(U_{ij}|X_i,Y_i) = E\left\{h^{-d-1}K'(\frac{X_j-X_i}{h})(Y_j - Y_i)|X_i,Y_i\right\}$

$$= h^{-d-1}\int K'(\frac{X_i-x}{h})\{Y_i - g(x)\}f(x)dx$$

$$= h^{-d-1}\int K'(\frac{X_i-x}{h})\{Y_i - g(x)\}f(x)dx$$

$$= h^{-1}\int K'(u)\{Y_i - g(X_i-hu)\}f(X_i-hu)du$$, by variable transformation

$$= h^{-1}Y_i\int K'(u)f(X_i - hu)du - h^{-1}\int K'(u)e(X_i - hu)du$$

$$= Y_i\int K(u)f'(X_i - hu)du - \int K(u)e'(X_i - hu)du \quad \text{, by integration by parts} \quad (1.38)$$

where $e(x)=g(x)f(x)$. Powell, Stock and Stoker (1989) and Robinson (1989) prove both terms on the right of (1.37) converge to zero in probability, and the first term times $\sqrt{n}$ converges to a normal variate while the second term times $\sqrt{n}$ converges to zero in probability under certain conditions for iid and weakly dependent observations respectively, hence U is a $\sqrt{n}$ - consistent and asymptotically normally distributed estimator of $\bar{\mu}$. For iid case, Theorem 3.1 of Powell, Stock and Stoker (1989) shows the asymptotic covariance matrix of U is

$$\Sigma = 4Var[f(X)g'(X)] + 4E[\{Y - g(X)\}^2 f'(X)f'(X)^T] \quad (1.39)$$

and they provide a consistent estimator of V, while Robinson (1989) also calculates the asymptotic covariance matrix and proposes its consistent estimator for dependent case (see equations (2.14)-(2.21) of Robinson (1989)). We compare (1.39) with the semiparametric efficiency bound in the following subsection. Cheng and Robinson (1994) study the properties of (1.35) when the observations are long range dependent. They show the convergence rate is not necessarily of $n^{-1/2}$ and the asymptotic distribution is not normal in general but it can be some Gaussian functional depending on the DGP.

We would like to give two remarks on the AD estimation. Firstly, (1.29)-(1.34) suggest the following possibility of generalization. Suppose we can consider

$$E(Y|X,Z) = H(\beta^T X, Z)$$

where $H(.,.)$ and β are unknown. This model nests type 2 and 3 Tobit models as well as the partially linear regression model (1.69) explained later in Section 1.2. Letting $f(x,z)$ be the joint density of X, Z and $\tilde{f}(x,z)$ be its estimate, we might be able to estimate β up to scale by $\frac{2}{n}\sum_{i=1}^{n} Y_i\, \partial \tilde{f}(X_i,Z_i)/\partial x$ since

$$E[f(X,Z)\partial g(X,Z)/\partial x] = \beta E[f(X,Z)\partial H(\beta^{\tau}X,Z)/\partial (\beta^{\tau}x)] = -2E[Yf(X,Z)]$$

similarly to (1.34) under some regularity conditions. Secondly, even if the specification (1.28) is incorrect namely Y depends on X not only through its linear combination, weighted AD estimates may still make sense in that it estimates the nonparametric quantity $E\{w(X)g'(X)\}$ which is a kind of average gradient of $g(x)$ with some weight, especially, when $w(x)=1$, this is obviously of direct interest.

1.1.3 Semiparametric efficiency bound for the density-weighted AD estimator

We can consider efficiency among semiparametric estimators in the manner like Cramèr-Rao lower bounds. See Stein (1956), Koshevnik and Levit (1976), Pfanzagl and Wefelmeyer (1982), Begun, Hall, Huang and Wellner (1983) and Bickel, Klaassen, Ritov and Wellner (1993) among others. Semiparametric models formally assume the distribution of random variables Z is $F(z)=F(z;\alpha,J(z))$ where α is a vector of unknown parameters, $J(z)$ is an unknown function and $F(.;.,.)$ may be assumed known or unknown. (1.28) is an example with $Z=(X^{\tau},Y)$, $\alpha=\beta$ and $J(.)=G(.)$. Semiparametric efficiency bound is based on the following consideration. Suppose we know $F(.;.,.)$ and also we can parametrize $J(z)=J(z;\theta)$ where the functional form of $J(.;.)$ is known, then the ML principle gives an efficient estimate of $\gamma=(\alpha,\theta)$. The basic idea of semiparametric efficiency bound comes from the fact that semiparametric estimators should not be more efficient than its parametric ones, because the semiparametric model $F(z;\alpha,J(z))$ is a wider class of models than the parametric $F(z;\alpha,J(z;\theta))$. Let $l(\gamma;Z)=\log F(Z;\alpha,J(Z;\theta))$ be the log-likelihood function and

$$I_{\gamma\gamma} = E[\frac{\partial\, l(\gamma;Z)}{\partial\,\gamma}\frac{\partial\, l(\gamma;Z)}{\partial\,\gamma^{\tau}}] = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\theta} \\ I_{\theta\alpha} & I_{\theta\theta} \end{pmatrix} \tag{1.40}$$

be the information matrix of the parametric model. Cramèr-Rao lower bound for estimating α is the upper left block of the inverse of $I_{\gamma\gamma}$, $I^{\alpha\alpha} = [I_{\alpha\alpha} - I_{\alpha\theta} I_{\theta\theta}^{-1} I_{\theta\alpha}]^{-1}$. Since semiparametric estimators of α cannot be more efficient than parametric ones, we define the semiparametric covariance bound as

$$I_{*}^{\alpha\alpha}(\alpha) = sup\,[I_{\alpha\alpha} - I_{\alpha\theta} I_{\theta\theta}^{-1} I_{\theta\alpha}]^{-1} \tag{1.41}$$

where the supremum is taken over all finite dimensional parametrization of $J(.;.)$. There is no guarantee that there exists a semiparametric estimator which attains this bound for a semiparametric model, but some estimators have been shown to attain this efficiency bound. (1.41) indicates that the nuisance function $J(.;.)$ causes efficiency loss compared with parametric estimation in general. In some cases, however, we can estimate the parameters of interest asymptotically equally well when the nuisance function is unknown to the case when it is known. This situation is studied in e.g. Stein (1956) and Bickel (1982) and referred to as adaptive.

Although the definition of semiparametric efficiency bounds does not directly lead to its calculation methods, Begun et.al. (1983), Bickel et.al. (1993) and Newey (1990b) provide methods to compute the bound. Using the method by Newey (1990b), Newey and Stoker (1993) derive the efficiency bound for the weighted AD $\mu_w = E[w(X) g'(X)]$,

$$Var[w(X)\, g'(X)] + E[u^2 l(X) l(X)^{\tau}] \tag{1.42}$$

where $u = Y - g(X)$ and $l(X) = -w'(X) - w(X) f'(X)/f(X)$. The first term of (1.42) comes from not knowing $g(x)$, the second from not knowing $f(x)$. Substituting the weight $w(x) = -f(x)$, we derive the efficiency bound for density-weighted AD $E[-f(X) g'(X)]$,

$$Var[f(X)\, g'(X)] + 4E[\{Y-g(X)\}^2 f'(X) f'(X)^T] \quad . \tag{1.43}$$

Comparing (1.39) with (1.43), we see the density-weighted AD estimator U does not attain the semiparametric efficiency bound. We will see efficient semiparametric estimators for single index models in Section 1.2.

1.1.4 Estimation of density and its derivatives

(1.30) can be seen as an estimator of density derivatives $f'(x)$ at $x = X_i$. Density and density derivatives estimation has a long history. The most primitive estimation method for density is histogram. More sophisticated methods have been developed based on smoothing techniques. Supposing $X_i, i = 1, 2, \ldots$ is an iid sample from a scalar variate X with absolutely continuous density $f(x)$, Rosenblatt (1956) first applies the kernel smoothing technique to estimate $f(x)$ by

$$\tilde{f}(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) \tag{1.44}$$

where h is a positive sequence satisfying $h \to 0$ and $nh \to \infty$ as $n \to \infty$ and $K(.)$ integrates to unity. This is called kernel density estimator. First order asymptotic properties of this estimator such as asymptotic unbiasedness, L_1 or L_2 convergence, $\sqrt{nh}$ - consistency and asymptotic normality have been studied under various conditions on $K(.)$ and h (see e.g. Parzen (1962), Nadaraya (1965, 1974), Epanechnikov (1969), Sethuraman and Sibuya (1961)). Berry-Esseen bounds (see 1.3.1) are obtained by Prakasa Rao (1977), while an Edgeworth expansion is validated by Hall (1991). A natural multivariate generalization

$$\tilde{f}(x) = \frac{1}{nh^d} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) \tag{1.45}$$

is proposed in Cacoullos (1966), where X is d-dimensional and $K: R^d \to R$. (1.30) is based on its derivatives

$$\tilde{f}^{\prime}(x) = \frac{1}{nh^{d+1}} \sum_{j=1}^{n} K^{\prime}\left(\frac{x - X_j}{h}\right) . \tag{1.46}$$

This way of estimating density derivatives is originally proposed by Bhattacharya (1967) when $d=1$. He suggests to estimate the p-th order derivative of $f(x)$ by

$$\tilde{f}^{(p)}(x) = \frac{1}{nh^{p+1}} \sum_{i=1}^{n} K^{(p)}\left(\frac{x - X_i}{h}\right) , \tag{1.47}$$

the p-th order derivative of the kernel estimator (1.44) when $d=1$. Bhattacharya (1967) and Schuster (1969) show its asymptotic unbiasedness and strong consistency respectively, while Silverman (1978) proves its weak and strong uniform consistency. Singh (1976) generalizes this to multivariate density derivatives and proves the asymptotic unbiasedness and strong consistency. Other estimation method is proposed in Singh (1977, 1979).

Kernel methods include user-determined bandwidth h. The first order theory requires only $h \to 0$ and $nh \to \infty$ for consistency or asymptotic normality so that the practical choice of bandwidth is of interest for empirical use. One possibility of determining h is to minimize the mean integrated squared error (MISE) with respect to h. Asymptotic MISE of (1.45) is $\int_{-\infty}^{\infty} E\{\tilde{f}(x) - f(x)\}^2 dx = O(h^{2L} + n^{-1}h^{-d})$ under (1.36) so that the optimal bandwidth is $O(n^{-1/(2L+d)})$ and hence the asymptotic MISE results in $O(n^{-2L/(2L+d)})$ (see e.g. Silverman (1986), p.85 for the case of $L=2$). This indicates that larger d yields slower convergence of (1.45) which Bellman (1961) calls curse of dimensionality. We could consider other criteria such as $\int_{-\infty}^{\infty} E|\tilde{f}(x) - f(x)| dx$ instead of MISE (see e.g. Devroye and Györfi (1985)). Typically, MISE involves unknown functionals of $f(x)$ (see e.g. Silverman (1986), P.39) and so does the bandwidth minimizing MISE, thus it is infeasible. Plug-in approach has been studied by many authors for feasible bandwidth choices, where the unknown functionals in MISE are replaced by their consistent estimates (see e.g.

Woodroofe (1970), Nadaraya (1974), Sheather (1983, 1986), Park and Marron (1990), Sheather and Jones (1991)). Cross validation methods have also been developed for feasible bandwidth choices. Suppose we consider the bandwidth choice in (1.44). Since minimization of MISE is equivalent to minimization of $\int \tilde{f}^2(x)dx - 2\int \tilde{f}(x)f(x)dx$ and the second term can be estimated by $-\frac{2}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K(\frac{X_i-X_j}{h})$, h minimizing

$$\int \tilde{f}^2(x)dx - \frac{2}{n(n-1)h}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K(\frac{X_i-X_j}{h})$$

is a reasonable choice. This is called the least squares cross validation. The ML principle also can be used for cross validation. ML cross validation treats the bandwidth like parameters to be estimated and maximizes nonparametrically estimated likelihood function over h, namely,

$$\max_h \sum_{i=1}^{n} \log \tilde{f}_{-i}(X_i),\ \ \tilde{f}_{-i}(X_i) = \frac{1}{(n-1)h}\sum_{j\neq i}^{n} K(\frac{X_i-X_j}{h}) \ .$$

Marron (1985) proves asymptotic optimality of this choice. Marron (1987) compares cross validation techniques, while Marron (1988) gives a brief introduction of these methods. Faraway and Jhun (1990) introduce bootstrap method of bandwidth selection where estimated MISE by bootstrapping is minimized with respect to h.

Optimal bandwidth selection $h = O(n^{-1/(2L+d)})$ for density estimate (1.45) mentioned in the previous paragraph is different from that for density derivative estimation. Härdle, Marron and Wand (1990) provides an optimal bandwidth choice for (1.47) by least squares cross validation. The MISE of (1.46) is

$$\int_{-\infty}^{\infty} E\{\tilde{f}'(x) - f'(x)\}^2 dx = \int_{-\infty}^{\infty}[V(\tilde{f}'(x)) + \{Bias(\tilde{f}'(x))\}^2]dx$$

$$= \int_{-\infty}^{\infty}\left[\frac{1}{nh^{2d+2}}V\{K'(\frac{x-X_1}{h})\} + [\frac{1}{h^{d+1}}E\{K'(\frac{x-X_1}{h})\} - f'(x)]^2\right] f(x)dx$$

$$= O(n^{-1}h^{-d-2} + h^{2L})$$

under some conditions on f, its derivatives and $K(.)$ including (1.36). The last equality uses

$$E\{K'(\tfrac{x-X_1}{h})\} = \int_{-\infty}^{\infty} K'(\tfrac{x-v}{h}) f(v)dv = h^d \int_{-\infty}^{\infty} K'(u) f(x-hu)du$$

$$= h^{d+1} \int_{-\infty}^{\infty} K(u) f'(x-hu)du = h^{d+1}\left\{ \int_{-\infty}^{\infty} K(u) f'(x)du + O(h^L)\right\}$$

$$= h^{d+1} f'(x) + O(h^{L+d+1}) \ ,$$

$$E\{K'(\tfrac{x-X_1}{h})\}^2 = \int_{-\infty}^{\infty} \{K'(\tfrac{x-v}{h})\}^2 f(v)dv = h^d \int_{-\infty}^{\infty} \{K'(u)\}^2 f(x-hu)du = O(h^d) \ .$$

This implies that the minimum MISE bandwidth for (1.46) is

$$h^* = C n^{-\frac{1}{2L+d+2}} \ , \tag{1.48}$$

where C is a positive constant. This is different from the optimal bandwidth for density estimation. We see in the following subsection the optimal choice of h for the AD estimation is different from these choices.

The kernel function is also user-determined. Additional to the condition that $K(.)$ integrates to unity, we may like to impose $K(u) \geq 0$ which guarantees the resulting density estimate is nonnegative at all points. Then, $K(.)$ is a density. Various alternatives have been suggested in the class of density, among which Epanechnikov kernel has an optimal property that it minimizes the MISE. Higher order kernel function (1.36) does not satisfy $K(u) \geq 0$ so that it can result negative estimate of density at some points, which may be inconvenient. But (1.36) can reduce the estimation bias so that there is a trade-off between $\tilde{f}(x) \geq 0$ and smaller bias of $\tilde{f}(x)$. In the current AD estimation, however, negative density estimate is not so problematic as when density itself is of interest. It rather seems to work quite well in view of the Monte Carlo study in Chapter 5.

There are other principles of nonparametric density estimation such as variable kernel estimation, k-nearest neighbour (k-NN) estimation, orthogonal series estimation and penalized maximum likelihood estimation. We refer to some monographs on density estimation, Tapia and Thompson (1978), Prakasa Rao (1983), Devroye and Györfi (1985), Silverman (1986), Devroye (1987), Scott (1992), and Wand and Jones (1995) as well as review papers by Izenman (1991) and Andrews (1995).

1.1.5 Bandwidth selection for AD estimation

As in the density and its derivatives estimation, h in (1.35) is user-determined in AD estimation. Powell, Stock and Stoker (1989) show the rate of decay required for $\sqrt{n}$ - consistency and asymptotic normality is $n^{-1}h^{-d-1} + n^{1/2}h^{L} \to 0$ as $n \to 0$ under iid environment, while Robinson (1989), under weakly dependent environment, shows various rates could apply depending on the moment condition, mixing condition and the rate of smoothing parameter for nonparametric estimation of power spectrum involved. The above theory only supplies the rate of decay, but we need to use a specific value of h in empirical applications. A principle to select a desirable h is to take it such that the mean squared error (MSE) $E(U - \bar{\mu})^2$ is minimized. Härdle and Tsybakov (1993) derive the following bandwidth choice which minimizes leading terms of the MSE for iid case,

$$h^* = h_0 n^{-\frac{2}{2L+d+2}}, \quad h_0 = \left\{ \frac{Q_2(d+2)}{2LQ_1} \right\}^{\frac{1}{2L+d+2}} \tag{1.49}$$

where Q_1 and Q_2 are constants depending on unknown functionals such as density of X and conditional variances. It is interesting that optimal bandwidth selection for AD estimation (1.49) is different from that for density derivative estimation (1.48). More specifically, we require less smoothing than derivative estimation. (1.49) is optimal but infeasible because of unknown Q_1 and Q_2 . Powell and Stoker (1996) propose plug-in bandwidth choices for more general statistics with the form of density-weighted averages including (1.35).

1.1.6 Estimation of the unknown function $G(.)$ of the single index model

We have shown how we can estimate β up to scale in (1.28) by the density-weighted averaged derivatives. The above asymptotic theory on U may be sufficient when our main interest is, for example, to test linear hypothesis such as $\beta_i=0$ or $\beta_i=\beta_j$ where β_i is the ith element of β. However, if we also would like to know $E(Y|X=x)=G(\beta^T x)$, say, for prediction, we need to estimate the unknown function $G(.)$ also. Since $G(.)$ is not specified parametrically, we adopt a nonparametric regression method. U estimates $\bar{\mu}=c_2\beta$ for some unknown constant c_2 so that we rewrite $G(.)$ correspondingly as $G(\beta^T x)=G(c_2^{-1}\bar{\mu}^T x)=H(\bar{\mu}^T x)$ and consider estimation of $H(.)$.

Nonparametric regression methods have been developed to consistently estimate regression functions when we have no information on the functional form, namely we would like to estimate $E(Y|X=x)=g(x)$ without assuming any parametric form. Let $f(x,y)$ be the joint density of X and Y, and $f(x)=\int_{-\infty}^{\infty}f(x,y)dy$. Since

$$f(x)\,g(x)=\int_{-\infty}^{\infty}y\,f(x,y)dy\ , \tag{1.50}$$

we could consider the following estimator of the regression function $g(x)$.

$$\tilde{g}(x)=\{\tilde{f}(x)\}^{-1}\int_{-\infty}^{\infty}y\tilde{f}(x,y)dy \tag{1.51}$$

where $\tilde{f}(x,y)$ and $\tilde{f}(x)$ are suitable estimates of $f(x,y)$ and $f(x)$ respectively. However this estimator is inconvenient since integration in (1.51) may not be able to work out or even worse it might not exist depending on the estimate of the joint density. Nonparametric kernel regression estimator originally introduced by Nadaraya

(1964) and Watson (1964) independently is

$$\tilde{g}(x) = \left\{ \frac{1}{nh} \sum_{i=1}^{n} K(\tfrac{x-X_i}{h}) \right\}^{-1} \frac{1}{nh} \sum_{i=1}^{n} Y_i K(\tfrac{x-X_i}{h}) \ , \tag{1.52}$$

where $h + n^{-1}h^{-1} \to 0$ as $n \to \infty$ and $K(.)$ integrates to unity. This is called Nadaraya-Watson (NW) estimator. Heuristically we can view this estimator as follows. The inverse factor on the right of (1.52) is the nonparametric kernel estimate for $f(x)$. The expectation of the second factor is, under iid environment and certain smoothness conditions on $g(.)$, $f(.)$ and $K(.)$,

$$h^{-1}E[YK(\tfrac{x-X}{h})] = h^{-1}E[g(X)K(\tfrac{x-X}{h})] = h^{-1}\int_{-\infty}^{\infty} g(v)K(\tfrac{x-v}{h})f(v)dv$$

$$= \int_{-\infty}^{\infty} g(x-hu)K(u)f(x-hu)du \to g(x)f(x) \int_{-\infty}^{\infty} K(u)du = g(x)f(x) \tag{1.53}$$

as $h \to 0$, the third equality using variable transformation. Thus the right hand side of (1.52) will converge to $g(x)$. When X is a scalar random variable, Nadaraya (1970) proves its uniform strong consistency, while Schuster (1972) shows the pointwise asymptotic normality when the estimator is normalized by $\sqrt{nh}$. Since h is chosen such that $h + n^{-1}h^{-1} \to 0$ as $n \to \infty$, convergence rate of the regression estimator is slower than $n^{-1/2}$ of typical parametric estimators. Greblicki (1974) generalizes (1.52) to multidimensional regressors and shows its strong consistency, while Mack and Silverman (1982) prove its weak and strong uniform consistency. Devroye and Wagner (1980a, b) prove $E(D_r) \to 0$ and $D_1 \xrightarrow{p} 0$ where $D_r = \int_{R^d} |\tilde{g}(x) - g(x)|^r dx$. Singh and Tracy (1977) study estimation of higher order conditional moments $E(Y^k|X)$.

In the index model, we can apply this method to estimate $G(.)$ or equivalently $H(.)$. Since U estimate $\bar{\mu}$, we simply regress $U^{\tau}X$ on Y nonparametrically using above method. Namely, we can estimate $H(u)$ by

$$\tilde{H}(u) = \left\{ \frac{1}{na} \sum_{i=1}^{n} K\left(\frac{u - U^{\tau}X_i}{a}\right) \right\}^{-1} \frac{1}{na} \sum_{i=1}^{n} Y_i K\left(\frac{u - U^{\tau}X_i}{a}\right) . \tag{1.54}$$

The bandwidth a used here need not be the same as h used in U. We could consider choosing a and h simultaneously by minimizing MISE of $\tilde{H}(U^{\tau}X)$. This choice of h may be different from the minimum MSE choice (1.49). Which criteria should be used depends on the purpose of the analysis.

Other nonparametric estimators of regression functions such as k-NN, orthogonal series, spline smoothing and local polynomial estimators could be applied to estimation of $H(.)$. We only refer to some good monographs on them, Prakasa Rao (1983), Härdle (1990) and Wand and Jones (1995).

1.1.7 Higher order asymptotic theory of AD estimation

We have seen the first order asymptotic properties of (1.35) is qualitatively identical to the parametric statistics for iid and weakly dependent observations in the sense that they are $\sqrt{n}$ - consistent and asymptotically normally distributed. This is a surprising result in the following sense. Since (1.29) involves a nonparametric kernel estimate of $f'(x)$ and its convergence rate is of order $(nh^{d+1})^{-1/2}$, strictly slower than parametric order of $n^{-1/2}$, it is likely that the estimator (1.35) is affected by the slow convergence property of (1.46). However, as far as the first order properties are concerned, the nonparametric estimate does not affect the properties of the estimator. Then a question arises: are its higher order properties also analogous to parametric statistics? Robinson (1995a) investigates the Berry-Esseen bound of (1.35) for iid observations. The Berry-Esseen bound determines the order of normal approximation error and it is typically of $n^{-1/2}$ for parametric statistics. He establishes

$$\sup_{\lambda:\lambda^T\Sigma\lambda=1} \sup_z |P[n^{1/2}\lambda^T(U-\bar{\mu}) \leq z] - \Phi(z)| = O(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L})$$

where Σ is the asymptotic covariance matrix of U (see (1.39) for the definition of Σ). The bound is not of parametric order in general, however it can attain the bound of $O(n^{-1/2})$ depending on the rate of decay of the bandwidth. He also calculates an optimal choice of bandwidth in terms of minimizing normal approximation error, which is

$$h^* = C n^{-\frac{3}{2(L+d+2)}} \ . \qquad (1.55)$$

This is of smaller order than (1.48) and (1.49). The purpose of this thesis is to extend his study to derive Edgeworth expansions of (1.35). We give a summary on higher order theory of parametric and semiparametric estimators in the last section of this chapter.

1.2 $\sqrt{n}$ - consistent semiparametric estimation

Section 1.1 shows when and how parametric methods can fail and the single index model is a good alternative to some parametric models such as (1.5)-(1.9) and others, then we have introduced the density-weighted AD estimation for it. We first review other estimation methods of single index models in 1.2.1. Subsection 1.2.2 reviews other $\sqrt{n}$ - consistent semiparametric estimators related to (1.5)-(1.8) and other models, specifically,

(1) discrete choice models

(2) censored and truncated regression models

(3) partially linear regression models

(4) linear regression models with heteroscedasticity of unknown form

(5) semiparametric maximum likelihood estimation

(6) other models such as simultaneous equation models.

(1) and (2) can be seen as special cases of index models, and some models seen in (5) and (6) include certain cases of (1) and (2). Also, methods in (3) can be applied to estimate (2). Subsection 1.2.3 generally compares parametric, semiparametric and nonparametric methods in econometrics.

Some review papers on semiparametric econometrics include Robinson (1988a), Newey (1990b), Delgado and Robinson (1992), Powell (1994) and Linton (1995b), while monographs on this topic are Pfanzagl (1990), Bickel *et.al.* (1993), M.-J. Lee(1996) and Horowitz (1998).

1.2.1 Semiparametric estimation of index models

We introduce a more general single index model than (1.28) which allow the index to be nonlinear in X, namely

$$E(Y|X)=g(X)=G(\nu(X;\beta)) \text{ a.s.}, \tag{1.56}$$

where Y is a scalar, X is d dimensional, function $\nu:R^d\to R^s$ $(d>s)$ is known up to unknown parameter β , while function $G(.)$ and distribution of the variables are assumed unknown. We can reduce the dimension of variables from d to s and thus face less curse of dimensionality, which is a strong advantage to nonparametric regression. (1.28) is a special case when $s=1$ and $\nu(X;\beta)=\beta^\tau X$ where β is a $d\times 1$ vector.

We first review AD type estimation when $\nu(X;\beta)=\beta^\tau X$. Stoker's (1986) original work considers averaged derivatives without weight,

$$\mu_1=-E\{\frac{f'(X)}{f(X)}g(X)\}=-E\{l'(X)Y\} \;,\; l(x)=\log f(x) \;. \tag{1.57}$$

Supposing we have iid observations $(X_i^\tau,Y_i), i=1,2,\ldots$, we can estimate it, similarly to the weighted averaged derivatives, by

$$\tilde{\mu}_1 = -\frac{1}{n}\sum_{i=1}^{n} \tilde{l}^{\prime}(X_i)Y_i \ , \ \ \tilde{l}(x) = \log \tilde{f}(x) \tag{1.58}$$

where $\tilde{f}(x)$ is a suitable estimate of $f(x)$. Assuming a parametric family $f(x)=f(x;\theta)$, Stoker (1986) uses $\tilde{f}(x)=f(x;\tilde{\theta})$ where $\tilde{\theta}$ is a consistent estimate of θ. He proves $\sqrt{n}$ - consistency and asymptotic normality of $\tilde{\mu}_1$. This estimation parametrizes the density function, but it does not require any specific form of $g(x)$ so that this is semiparametric. Härdle and Stoker (1989), without parametrizing $f(x)$, plug (1.45) into (1.58) where $K(.)$ is a higher-order kernel function. They show this estimator is $\sqrt{n}$ - consistent and asymptotically normally distributed and derive a consistent estimator for the asymptotic covariance matrix. Stoker (1991) shows this estimator is first order equivalent to other four estimators including the derivative of NW estimator. Chaudhuri, Doksum and Samarov (1991) consider average derivative quantile regression

$$\mu_1(\alpha) = E[\frac{\partial\, \theta_\alpha(X)}{\partial\, x}], \ \theta_\alpha(x) = \inf_y\{y : F_{Y|X}(y|x) \geq \alpha\} \ ,$$

where local polynomial regression estimator is used for θ_α, and prove the estimator is $\sqrt{n}$ - consistent and asymptotically normally distributed under quite weak conditions. Härdle, Hart, Marron and Tsybakov (1992) propose an optimal bandwidth selection which minimizes leading terms of mean squared error (MSE) of Härdle and Stoker's (1989) estimator. They show that the optimal bandwidth is of order $n^{-2/(2L+d)}$ where L is the kernel order. Härdle, Hildenbrand, and Jerison (1991) apply this method to estimate households' income effect from U.K. family expenditure data.

Andrews (1991) proposes a series estimator for AD and proves its $\sqrt{n}$ - consistency and asymptotic normality under a general setup including some other semiparametric estimations. In addition to deriving the efficiency bound for weighted AD (1.42), Newey and Stoker (1993) propose a method to construct an efficient estimator by combining different weighted AD estimators via minimum chi-square.

Ahn (1997) considers the following index model where a part of the regressors is nonparametrically generated,

$$E(Y|X) = G(X_0^T\beta_0 + \lambda(m(X_1))^T\beta_1),\ m(X_1) = E(X_2|X_1)\ . \tag{1.59}$$

Here Y, X_0, X_1, X_2 are observables, parameters β_0, β_1, $G(.)$ and $m(.)$ are unknown, but $\lambda(.)$ is known. He proposes the following two step estimation. We perform nonparametric regression in the first step to estimate $m(.)$, then replacing the unknown $m(.)$ by the first step estimate, we implement AD estimation for β_0, β_1 up to scale in the second step. He proves its $\sqrt{n}$ - consistency and asymptotic normality.

Ichimura and Lee (1991) propose to extend the least squares principle to the estimation of (1.56), namely,

$$\min_{\beta}\sum_{i=1}^{n}\{Y_i - G(\nu(X_i;\beta))\}^2\ . \tag{1.60}$$

They partly specify the function $G(.)$ as multiple index form,

$$G(\nu(X;\beta)) = X_0^T\beta_0(\theta) + \psi(X_1^T\beta_1(\theta),\ldots X_M^T\beta_M(\theta)) \tag{1.61}$$

where $X^T = (X_0^T,\ldots,X_M^T)$, $\beta_i(.)$, $i=0,\ldots,M$ are known functions of basic parameters θ , and $\psi(.)$ is an unknown function. (1.60) is infeasible since we do not know the true functional form of $G(.)$ due to the unknown $\psi(.)$, so that they replace $\psi(.)$ by its nonparametric (kernel) estimate and construct

$$E_n(Y_i|X_i,\theta) = \frac{\sum_{j\neq i}^{n}[Y_j - X_{0j}^T\beta_0(\theta)]K(\frac{[X_{1i}^T - X_{1j}^T]\beta_1(\theta)}{h},\ldots,\frac{[X_{Mi}^T - X_{Mj}^T]\beta_M(\theta)}{h})}{\sum_{k\neq i}^{n}K(\frac{[X_{1i}^T - X_{1k}^T]\beta_1(\theta)}{h},\ldots,\frac{[X_{Mi}^T - X_{Mk}^T]\beta_M(\theta)}{h})}\ ,$$

thus the feasible minimization problem reduces to

$$\min_{\theta}\sum_{i=1}^{n}\{Y_i-X_{0i}^{\tau}\beta_0(\theta)-E_n(Y_i|X_i,\theta)\}^2 \ . \tag{1.62}$$

They prove this estimator is $\sqrt{n}$ - consistent and asymptotically normally distributed, and provide a consistent estimator for the asymptotic covariance matrix. Stern (1996) applies this method to estimate the supply and demand effects of disability on labour force participation. Ichimura (1993) applies the same principle to single index model (1.56) when $s=1$. He allows some weight and proposes an estimator

$$\min_{\theta}\sum_{i=1}^{n}W(X_i)\{Y_i-\hat{E}(Y_i|X_i;\theta)\}^2 \ . \tag{1.63}$$

where

$$\hat{E}(Y_i|X_i,\theta)=\frac{\sum_{j\neq i}^{n}W(X_i)Y_jK(\frac{\nu(X_i;\theta)-\nu(X_j;\theta)}{h})}{\sum_{k\neq i}^{n}W(X_k)K(\frac{\nu(X_i;\theta)-\nu(X_k;\theta)}{h})} \ .$$

He proves the estimator is also $\sqrt{n}$ - consistent and asymptotically normally distributed, and provides a consistent estimator for the asymptotic covariance matrix. He further shows that it attains the semiparametric efficiency bound obtained by Newey (1990b) under the optimal choice of weight function $W(x)=\{V(Y|X=x)\}^{-1/2}$. Though this weight is infeasible, the same asymptotic properties may still hold when we replace it by its suitable consistent estimate. Härdle, Hall and Ichimura (1993) derive an optimal choice of bandwidth for Ichimura's (1993) estimator under linear-in-variable index assumption, i.e. $\nu(X;\beta)=X^{\tau}\beta$. They show the asymptotically minimum MSE choice of bandwidth is of order $n^{-1/5}$. This order equals to that of optimal bandwidth choice for nonparametric regression with one regressor and so differs from the choice for AD estimation.

Samarov (1993) shows that $\sqrt{n}$ - consistent estimates of certain AD functionals such as $E[w(X)\frac{\partial g(X)}{\partial x}\{\frac{\partial g(X)}{\partial x}\}^{\tau}]$ and $E[w(X)\frac{\partial^2 g(X)}{\partial x \partial x^{\tau}}]$ are useful to determine a model out of some alternative non/semiparametric models.

1.2.2 Other $\sqrt{n}$ - consistent semiparametric estimation

We have seen single index model (1.28) is a good alternative in terms of robustness to (1.5)-(1.9) and some other parametric models. There could be, however, a drawback of AD estimator caused by its too much generality. Suppose we know $Y_i = I(\beta^T X_i + u_i \geq 0)$ is the correct specification but we do not know the distribution of u. Obviously AD estimation can be applied to this model. However (1.35) estimates β only up to an unknown constant so that it may be inconvenient when we want to know the level of β itself. Also if we would like predictions of Y, those based on the AD estimation and (1.54) may not be good due to the slow convergence of $\tilde{H}(u)$. Thus we may not want to estimate the model based on (1.28), but based on $Y_i = I(\beta^T X_i + u_i \geq 0)$ just without a specific distribution of u. There have been proposed various ways to estimate the parameters of (1.5)-(1.8) $\sqrt{n}$ - consistently without specifying distribution of disturbances parametrically.

Semiparametric estimation of discrete choice models

For discrete choice models including (1.7) and (1.8), the pioneering work by Manski (1975) considers the following polychotomous choice setup. Supposing the utility function of an individual i under a choice j is $U_{ij} = \beta^T X_{ij}$, i chooses an option j when $\beta^T X_{ij} \geq \beta^T X_{ik}$ for all $k \neq j$. He proposes to estimate β by

$$\max_{\beta} \frac{1}{n} \sum_{i=1}^{n} W(R(j|i,\beta))$$

where $W(.)$ is any monotone increasing function, and $R(j|i,\beta)$ is the rank function presenting the rank of choice j for individual i given a parameter β . This estimate the parameter vector such that it maximizes the sum of the rank of the choice among alternatives each individual takes. He proves the estimator is consistent, while Cavanagh (1987) and Kim and Pollard (1990) study the convergence rate and asymptotic distribution and show this estimator is of slower convergence rate than $\sqrt{n}$. Manski (1987) applies this principle to binary panel data. The objective function is discontinuous with respect to the unknown parameters so that Horowitz (1992) modifies the objective function to be continuous using kernel method and Horowitz (1993) applies this method to the choice between automobile and transit for work trip. Ruud (1983) proposes to apply pseudo maximum likelihood estimation to the discrete choice models, while Cosslett (1983) proposes a distribution-free MLE for the following binary choice model. Suppose individual i faces a choice between two alternatives "1" and "0". Let U_{ji} be the utility of i from choosing j, $j=1$, 0, and it has a parametric form $U_{ji}=v(X_{ji};\theta)+\epsilon_{ji}, j=1,0$ where X_{ji} is a vector of some variables associated with option j, $v(.;.)$ is known up to a parameter vector θ and ϵ is a disturbance term with unknown distribution. Then his choice is characterized by

$$Y_i = I(U_{1i} \geq U_{0i}) = I(\ v(X_{1i};\theta) - v(X_{0i};\theta) \geq \epsilon_{0i} - \epsilon_{1i}\ ) \ .$$

Putting $V(Z_i;\theta) = v(X_{1i};\theta) - v(X_{0i};\theta)$, $Z_i = (X_{1i}, X_{0i})$ and denoting the distribution function of $\epsilon_{0i}-\epsilon_{1i}$ by F, the log-likelihood function is

$$\sum_{i=1}^{n}\left(Y_i \log F[V(Z_i;\theta)] + (1-Y_i)\log\{1-F[V(Z_i;\theta)]\}\right) \ . \tag{1.64}$$

Applying the algorithm by Ayer, Brunk, Ewing, Reid and Silverman (1955) to obtain an estimate of F corresponding to each θ, we maximize the likelihood with respect to θ. He proves its strong consistency. The log-likelihood function is analogous to (1.15). Cosslett (1987) shows its efficiency. Since (1.64) is discontinuous with respect to θ due to the discontinuity of the estimate of $F(.)$ by Ayer et.al., Klein and Spady (1993) extend this idea to invent a continuous objective function and prove the estimator is $\sqrt{n}$ - consistent and asymptotically normally distributed. Matzkin (1991) proposes a semiparametric MLE of polychotomous choice models with nonparametric utility function and normal disturbances. There are some other related works by Manski (1991), Ahn and Manski (1993) and Lewbel (1997).

Semiparametric estimation of censored regression models

There are some studies for semiparametric version of censored and truncated models (1.5) and (1.6) where the disturbance distribution is not parametrized. It can be estimated by distribution-free ML method as Cosslett (1983), namely we replace $\Phi(.)$ in (1.24) and (1.25) by certain estimate of the disturbance distribution and maximize it over the unknown parameters. For (1.5), Powell (1984) proposes least absolute deviation (LAD) estimation by

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} |Y_i - \max(0, \beta^\tau X_i)|$$

and proves its $\sqrt{n}$ - consistency and asymptotic normality. Powell (1986b) extends this principle to estimation of censored regression quantiles, while Honorè (1992) applies LAD to fixed effect panel data models. Horowitz (1986) proposes NLS type estimation for (1.5). Since $E(Y|X) = \int_{-\beta^\tau X}^{\infty} \{1 - F(u)\} du$, his estimator is

$$\tilde{\beta} = \operatorname{argmin}_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i - \int_{-\beta^\tau X_i}^{\infty} [1 - F_n(u;\beta)] du \right\}^2$$

wher $\tilde{\beta} = \operatorname{argmin}_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i - \int_{-\beta^\tau X_i}^{\infty} [1 - F_n(u;\beta)] du \right\}^2$ sturbance distribution by Kaplain and Meier (1958). Powell (1986a) takes a different approach and proposes to use only observations satisfying $E(Y - \beta^\tau X|X) = 0$. Assuming the symmetry of disturbance distribution,

$$E[Y^* | X] = \beta^\tau X \quad \text{where} \quad Y^* = Y I(\beta^\tau X + \epsilon \geq 2\beta^\tau X) \ .$$

He trims the observations with $Y_i \geq 2\beta^\tau X_i$ which are the symmetric equivalent of observations with $\beta^\tau X_i \leq 0$. Based on this idea, the symmetrically trimmed least squares estimator is defined as

$$\tilde{\beta} = \underset{\beta}{argmin} \frac{1}{n} \sum_{i=1}^{n} \{ Y_i - \max(Y_i/2, \beta^\tau X_i) \}^2 \ .$$

This estimator, however, obviously loses some efficiency because of the trimmed observations. Extending this idea, Newey (1991) proposes a method of moment type estimator for the Tobit model under conditional symmetry. When the sample is trimmed as in the above, we have $E[m(Y - \beta^\tau X) | X] = 0$ for any odd function $m(.)$ under the symmetry of ϵ. He proposes an estimator including this information and proves it improves the efficiency of Powell's estimator and attains the semiparametric efficiency bound. Honorè and Powell (1994) propose a different way to estimate β. Putting

$$\epsilon_{ij} = \max(Y_i - \beta^\tau X_i, -\beta^\tau X_j) = \max(\epsilon_i, -\beta^\tau X_i, -\beta^\tau X_j) \ ,$$

since conditional expectation of $\epsilon_{ij} - \epsilon_{ji}$ given X_i, X_j is zero, they propose to estimate β by the following minimization of U process objective function.

$$\min_{\beta} \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j \neq i}^{n} [\max(Y_i - \beta^\tau X_i, -\beta^\tau X_j) - \max(Y_j - \beta^\tau X_j, -\beta^\tau X_i)]^2 \ .$$

Horowitz (1988b) proposes M-estimation for (1.5) with an influence function $\psi(y, x, \beta, F)$ satisfying $E[\psi(Y, X, \beta, F)] = 0$. His estimator is a solution to

$$\frac{1}{n} \sum_{i=1}^{n} \psi(Y_i, X_i, \tilde{\beta}, F_n(\cdot, \tilde{\beta})) = 0$$

where $F_n(., \beta)$ is a smoothed version of Kaplan and Meier's product limit estimate of F. Hall and Horowitz (1990) discusses an asymptotically optimal bandwidth selection for this estimation. Horowitz and Neumann (1987) use this estimator for employment duration models.

Ahn and Powell (1993) and Powell (1998) propose estimators for censored sample selection models

$$Y = (\beta^{\tau}X + \epsilon) I(S \geq 0) \quad , \quad S = g(W) + u$$

where Y, X, S and W are observables, function $g(.)$ is unknown in Ahn and Powell (1993), while Powell (1998) assumes a semiparametric $g(W) = l(\delta^{\tau}W)$. The reduced form is

$$Y = \beta^{\tau}X + \lambda(g(W)) + \epsilon, \quad g(W) = E(S|W) \tag{1.65}$$

where $\lambda(.)$ is unknown function depending on the disturbance distribution. $g(.)$ is estimated nonparametrically by NW kernel regression in Ahn and Powell (1993) and semiparametrically in Powell (1998) in the first step. Letting the estimate $\tilde{g}$, the second step takes pairwise differences, collects the observations such that the nuisance function asymptotically disappears. Writing

$$Y_i - Y_j = \beta^{\tau}(X_i - X_j) + \lambda(\tilde{g}_i) - \lambda(\tilde{g}_j) + \epsilon_i - \epsilon_j \quad ,$$

if we can collect the pairs of observations (i, j) with $\lambda(\tilde{g}_i) - \lambda(\tilde{g}_j) = 0$, we can perform standard regression analysis based on

$$Y_i - Y_j = \beta^{\tau}(X_i - X_j) + \epsilon_i - \epsilon_j \quad .$$

When W is discrete, we can possibly find and pile up these pairs. If not, we regress $Y_i - Y_j$ on $X_i - X_j$ putting more weights on the observations with smaller $\lambda(\tilde{g}_i) - \lambda(\tilde{g}_j)$. The weights are produced by the standard kernel method. These two methods can be viewed as special cases of Ahn (1997) seen in the previous subsection (compare (1.59) and (1.65)).

Cosslett (1991) proposes a two step semiparametric estimation for type 2 Tobit model,

$$Z = I(\beta_1^{\tau}X_1 + \epsilon_1 \geq 0) \quad , \quad Y = (\beta_0^{\tau}X_0 + \epsilon_0) I(\beta_1^{\tau}X_1 + \epsilon_1 \geq 0) \quad , \tag{1.66}$$

where Z,Y,X_0,X_1 are observables. In the first step, distribution-free ML estimation by Cosslett (1983) is applied to estimate β_1 and least squares method is applied to derive an estimate of β_0 in the second step. Chen (1997) proposes a two-step semiparametric least squares estimator for the type 3 Tobit model,

$$Z=(\beta_1^T X_1+\epsilon_1)I(\beta_1^T X_1+\epsilon_1\geq 0) \quad , \quad Y=(\beta_0^T X_0+\epsilon_0)I(\beta_1^T X_1+\epsilon_1\geq 0) \quad , \tag{1.67}$$

where Z,Y,X_0,X_1 are observables. Since

$$E(Y|X_0,\epsilon_1\geq 0,\beta_1^T X_1\geq 0)=\beta^T X_0+\alpha \tag{1.68}$$

for a constant α, given consistent first step estimates of ϵ_1 and β_1 , β_0 can be estimated by

$$\min_{\alpha,\beta_0}\frac{1}{n}\sum_{i=1}^{n} I(\tilde{\epsilon}_{1i}\geq 0,\tilde{\beta}_1^T X_{1i}\geq 0)(Y_i-\beta_0^T X_{1i}-\alpha)^2 \quad .$$

The above estimators are all $\sqrt{n}$ - consistent and asymptotically normal.

Semiparametric estimation of truncated regression models

Tsui *et.al.* (1988), using (1.11), propose to estimate β of truncated model (1.6) by

$$\tilde{\beta}=\underset{\beta}{argmin}\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i-\beta^T X_i-\tilde{B}(\beta^T X_i)\right\}^2 \quad ,$$

where

$$\tilde{B}(\beta^T X)=\int_{-\beta^T X}^{\infty} u dF_n(u;\beta)/[1-F_n(-\beta^T X;\beta)]$$

is an estimate of the bias $E(Y-\beta^T X|X)$. This estimator involves Kaplan and Meier's (1958) estimate of disturbance distribution F_n which is a step function, so that the objective function is discontinuous with respect to the parameter and thus the estimator is computationally inconvenient. Lee (1992) uses the same transformation as in Tsui *et.al.* (1988) and plugs in a kernel smoothed estimate of $B(\beta^T X)$ by Breiman *et.al.*(1987). The estimators by Tsui *et.al.* (1988) as well as Lee (1992) evaluate the "bias" $E(Y-\beta^{\tau}X|X)$ nonparametrically and plug its estimate into the objective function.

A more general truncated regression model allows the truncation criteria to include other variables as in Type 2 and 3 Tobit models. Namely,

$$Y_i = \beta^{\tau}X_i + \eta_i, \quad \eta_i \text{ having density } f_{\eta_i}(x) = \begin{cases} \dfrac{f(x)}{1-F(-\gamma^{\tau}W_i)}, & x > -\gamma^{\tau}W_i \\ 0, & x \le -\gamma^{\tau}W_i \end{cases}$$

where samples with $\gamma^{\tau}W_i + \epsilon_i \ge 0$ are not observed. Chen and Lee (1998) propose a two step semiparametric ML estimation for this sample selection model, where unknown density and distribution functions in the likelihood function are replaced by kernel estimates.

Semiparametric estimation of partially linear regression models

Partially linear regression model

$$E(Y|X_0, X_1) = \beta^{\tau}X_0 + k(X_1) \text{ a.s.} \tag{1.69}$$

where we assume $k(.)$ is unknown is another semiparametric regression model intensively studied. This model includes standard linear regression model and sample selection model (1.65) as special cases. There have been developed two approaches to estimation of this model.

The first approach is the partial smoothing spline estimation and its variants proposed by Wahba (1984, 1986), Engle, Granger, Rice and Weiss (1986), Heckman (1986), Rice (1986), Shiau, Wahba and Johnson (1986), and Chen and Shiau (1991). Assuming X_1 is a deterministic design variable on $[0,1]$, they consider the estimator defined by

$$\min_{\beta,k(.)} \left\{ \frac{1}{n}\sum_{i=1}^{n} \{Y_i - \beta^\tau X_{0i} - k(X_{1i})\}^2 + \lambda \int_0^1 \{k^{(m)}(v)\}^2 dv \right\} , \tag{1.70}$$

where $k^{(m)}(.)$ denotes m-th derivative of $k(.)$. This is an extension of nonparametric spline smoothing estimation of regression function to the semiparametric regression model (1.69). Engle *et.al.* (1986) applies this method to investigate the electricity demand. Rice (1986) shows this estimator can achieve the standard parametric rate of convergence by suitably undersmoothing the nonparametric component $k(.)$. Heckman (1986) proves $\sqrt{n}$ - consistency and asymptotic normality of the estimator when X_1 is a nonstochastic scalar variable on the unit interval. Chen and Shiau (1991) proposes a variant of (1.70) called two-stage spline smoothing where rough parametric component is also penalized. λ raises the same practical problem as the bandwidth choice in nonparametric kernel estimation. Chen and Shiau (1994) propose data-driven choices of λ for the estimator of their 1991 paper.

The second approach is the partial regression estimation by Robinson (1988b) and Speckman (1988). They are based on the regression of partial correlations. Putting $\epsilon = Y - E(Y|X_0,X_1)$, we have

$$Y - E(Y|X_1) = \beta_0^\tau \{X_0 - E(X_0|X_1)\} + \epsilon \ . \tag{1.71}$$

This is free from the unknown (nuisance) function $k(.)$ and it gives

$$\beta_0 = \left(E\left[\{X_0 - E(X_0|X_1)\}\{X_0 - E(X_0|X_1)\}^\tau\right]\right)^{-1}$$
$$E\left[\{X_0 - E(X_0|X_1)\}\{Y - E(Y|X_1)\}^\tau\right] \ . \tag{1.72}$$

Assuming absolutely continuous X_1 , Robinson's (1988b) estimator for β_0 replaces the expectations by their sample analogue and the conditional expectations by their NW kernel estimates in (1.72). He uses a higher order kernel to avoid the bias problem for the first time in semiparametric estimation as mentioned in the previous section, which gets worse as dim(X) increases. He proves the estimator is $\sqrt{n}$ -

consistent and asymptotically normally distributed and provides a consistent estimator for the asymptotic covariance matrix. Delgado and Mora (1995) proposes an estimator which allows discrete regressors based on (1.72). Assuming X_1 is a scalar nonstochastic design variable and X_0 is a vector of nonstochastic variables, Speckman (1988) considers the same type of estimator of β_0 as Robinson's and derives asymptotic expressions of bias and variance of the estimator. The results imply asymptotic unbiasedness of the estimator. He also shows the estimator can coincide with partial smoothing spline estimator depending on the smoother and kernel function used in the two estimations. Li and Stengos (1996) apply this idea to panel data, employing the same transformation as (1.71) then multiply the density of X_{1it} to prevent the stochastic denominator in $E(Y|X_1)$. Ai and McFadden (1997) considers more general partly specified regression model

$$g(E[Y_0|X_0,X_1,X_2,D=1]) = \beta_0^T X_0 + k_1(X_1) + \beta_1 k_2(E[Y_1|X_2]) \ ,$$

where Y_0 and Y_1 are dependent variables, X_0, X_1 and X_2 are vectors of regressors, D is the 0-1 dummy variable, $\beta^T = (\beta_0^T, \beta_1)$ is a vector of unknown parameters, $k_1(.)$ is an unknown function, and g(.) and $k_2(.)$ are assumed to be known. Their estimator is based on the same transformation as (1.72) and plugs nonparametric estimates of unknown functions.

Linear regression with heteroscedasticity of unknown form.

We assume the following linear regression model.

$$Y = \beta' X + \epsilon \ , \ E(\epsilon|X) = 0, \ V(\epsilon|X) = \sigma^2(X) \ . \tag{1.73}$$

If $\sigma^2(.)$ is known, the standard GLS estimator

$$\beta_{GLS} = \{\sum_{i=1}^{n} \sigma(X_i)^{-1} X_i X_i^T\}^{-1} \sum_{i=1}^{n} \sigma(X_i)^{-1} X_i Y_i \tag{1.74}$$

provides an efficient estimate for β under certain regularity conditions. Even if we perform OLS neglecting the heteroscedastic structure, we still get a consistent estimate. The problem there is that the estimator is not efficient and it also produces invalid standard errors, and thus invalid test statistics. If we have a consistent estimate of $\sigma^2(.)$, we can perform feasible GLS estimation and it gives an efficient estimate of β with correct standard errors. There have been a lot of work on linear regression with heteroscedastic disturbance where the form of heteroscedasticity is parametrically specified (see e.g. Goldfeld and Quandt (1965), Rutemiller and Bowers (1968), Glesjer (1969), Box and Hill (1974), Harvey (1976) and Carroll and Ruppert (1982) among others). If the specification of the heteroscedasticity is correct, replacement of the true covariance function by its estimate in (1.74) will yield an asymptotically efficient estimate of β, but if the specification is wrong, the efficiency does not hold any more. Without assuming a parametric form of $\sigma^2(.)$, Carroll (1982) proposes an asymptotically efficient estimator of β using a kernel estimate of $\sigma^2(.)$ under independent disturbance assumptions, while Hidalgo (1992) allows weak dependence. Craig (1983) provides an efficient instrumental variable estimator. His approach is not like that of Carroll, but he proposes to choose a good instrument which can reduce the efficiency loss in the coefficient estimation. Though it depends on the choice of instruments, he does not provide an automatic method of instrument selection, so that it may not be practical. Robinson (1987) plugs a nonparametric k-NN estimate of $\sigma^2(.)$ in (1.74), extending Carroll's estimation to the multivariate regression model, and proves its $\sqrt{n}$ - consistency, asymptotic normality and efficiency. Delgado (1992) considers multivariate nonlinear regression models in the presence of heteroscedasticity of unknown form. He also uses k-NN estimates and shows all the desirable asymptotic properties in Robinson (1987) still hold.

Semiparametric maximum likelihood estimation

We have reviewed semiparametric ML estimation for discrete choice models. We can extend it to more general semiparametric models. Gallant and Nychka (1987) propose an estimation procedure for some nonlinear regression models without assuming specific distribution functions of random terms. They call their method "Semi-nonparametric" ML estimation. The main idea comes from Phillips (1983) who shows an ERA (extended rational approximants) $\phi^2(u|\mu,\Sigma)[P^2(u)/Q^2(u)]$ can approximate any density function arbitrarily closely under certain conditions, where $\phi(u|\mu,\Sigma)$ is the density of multivariate normal distribution with mean μ and covariance matrix Σ , and $P(.)$ and $Q(.)$ are polynomials. They use this density approximation to construct the likelihood function. They are concerned with estimation of sample selection models and Stoker's (1986) functional and prove the estimator of the parameters is strongly consistent under certain conditions. This method, modified suitably, can be applied to other semiparametric regression models. Ai (1997) analyzes a general regression model, where the conditional density of Y given X involves two parts of parameters, a finite-dimensional component θ and a infinite-dimensional (nonparametric) component $\lambda(.)$, and there exists a variable transformation $(y,x) \rightarrow (v_1(y,x,\theta), v_2(x,\theta))$ of known form satisfying

$$f_{Y|X}(y|x,\theta,\lambda) = J(y,x,\theta) f_{V_1|V_2}(v_1(y,x,\theta)|v_2(x,\theta),\theta) ,$$

where J is the known Jacobian of transformation $y \rightarrow v_1(y,x,\theta)$. This class of models includes special cases such as limited dependent variable models, partially specified regression model, selection models, and simultaneous equation models. The ML estimator maximizes

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\log J(Y_i,X_i;\theta)+\log\bar{f}_{V_1|V_2}(v_1(Y_i,X_i;\theta)\,|\,v_2(X_i;\theta))\right\}$$

with respect to θ , where $\bar{f}_{V_1|V_2}$ is the kernel estimate of $f_{V_1|V_2}$. He proves the estimator is $\sqrt{n}$ - consistent, asymptotically normally distributed and efficient in the sense that it attains the efficiency bound in Begun *et.al.*(1983) for multivariate nonlinear regression, simultaneous equations, partially linear regression, index regression, censored regression, switching regression, and disequilibrium models where the error density is unknown.

Other semiparametric estimation

There are some other $\sqrt{n}$ - consistent semiparametric estimation methods which are not classified in the above. Newey (1990a) proposes a $\sqrt{n}$ - consistent and asymptotically normal estimator of nonlinear simultaneous equation systems under iid environment. He replaces the infeasible optimal instruments of Amemiya (1974, 1977) by their nonparametric estimates and proves this is an adaptive situation, and thus the feasible instrumental variable estimator is efficient.

Robinson (1991a) considers more general DGP than Newey (1990a) and derives an efficient three stage least squares estimator for nonlinear simultaneous equations model. Unlike iid observations of Newey (1990a), he considers three different settings; (1) independent error terms and strongly exogenous but possibly serially dependent explanatory variables, (2) independent error terms and explanatory variables which include lagged endogenous variables and (3) parametrically autocorrelated error terms and strongly exogenous but possibly serially dependent explanatory variables. Somewhat different type of estimators are proposed for different settings. Given first step $\sqrt{n}$ - consistent estimates of parameters, he propose an estimator by Gauss-Newton iteration replacing the infeasible instrument by its estimate. He proves the estimator is $\sqrt{n}$ - consistent, asymptotically normally distributed and efficient. He further derives a Berry-Esseen bound when the error terms are independent.

Robinson (1991b) proposes to estimate semiparametric time series models based on first stage nonparametric spectrum estimates. Assuming the variables are covariance stationary, spectrum estimator based on kernel weighted periodogram is proved to be weakly uniformly consistent. He further considers spectrum regression in general semiparametric framework including multivariate linear regression and proves asymptotic normality of the estimator. Automatic choice of minimum MISE bandwidth is also provided.

Lee (1998) proposes a minimum distance estimator for semiparametric simultaneous equation microeconometric models with index restriction. The model is

$$E[g(Z;\beta)\,|X] = E[g(Z;\beta)\,|\delta^{\tau}X]$$

for vectors of random variables Z and X, vectors of unknown parameters β and δ , and a vector of known functions $g(.)$. The minimum distance estimator solves

$$\min_{\beta}\sum_{i=1}^{n}\{E_n[g(Z_i;\beta)\,|X_i] - E_n[g(Z_i;\beta)\,|\delta^{\tau}X_i]\}^{\tau}$$
$$W(X_i)\{E_n[g(Z_i;\beta)\,|X_i] - E_n[g(Z_i;\beta)\,|\delta^{\tau}X_i]\}$$

where $E_n(.\,|.)$ is the NW kernel estimates of $E(.\,|.)$ and $W(.)$ is some matrix of weight functions. This class of estimator can be applied to models such as simultaneous equation sample selection models, multi-market disequilibrium models and the simultaneous equations Tobit model. He shows this estimator is $\sqrt{n}$ - consistent and asymptotically normally distributed. If the conditional variance

$V[g(Z;\beta)|X]$ depends on X only through $\delta^{\tau}X$, the optimal weight is $W_{opt}(X) = V^{-1}(g(Z;\beta)|\delta^{\tau}X)$. This weight is infeasible, but we can estimate it by the kernel method with the first step estimates of β,δ . He proves the estimator using the feasible optimal weight achieves the semiparametric efficiency bound by Chamberlain (1992).

Newey (1994) derives a general formula for the asymptotic variance of semiparametric estimators. He considers a general semiparametric estimator depending on a series estimator of unknown nonparametric functions and shows they are $\sqrt{n}$ - consistent and asymptotically normal under certain high-level assumptions. The estimator considered there includes e.g. polynomial estimators of averaged derivatives and semiparametric panel probit models.

1.2.3 Comparison of nonparametric, semiparametric and parametric methods

Section 1.1 and the previous subsections have shed light on some parametric, semiparametric and nonparametric methods in various regression framework. When we are interested in $E(Y|X) = g(X)$, parametric method assumes certain functional form of $g(.)$, for example linear regression $g(x) = \beta^{\tau}x$, while nonparametric method does not parametrize $g(.)$ at all. Semiparametric method is the intermediate between the two and it partly specifies the function, for example as in (1.28) or models in Subsection 1.2.2. Parametric estimation methods collapse in general when the functional specification and/or assumed underlying distribution is incorrect. In this sense nonparametric or semiparametric estimation based on a less specified model is more robust and reliable when we are not sure of the specification of the model. We discuss here which of the three methods suit econometrics most.

Formally, we can say nonparametric methods aims to consistently estimate some functionals of the underlying joint distribution, such as joint density, conditional density and conditional expectations, without their parametric specification. Nonparametric models typically assume only certain smoothness of the functions so that they are very general in the sense that they include many parametric and semiparametric models as special cases. We, however, typically have to pay the cost of slower convergence rate than the parametric rate of $\sqrt{n}$. Semiparametric methods parametrically specify certain aspects of the joint distribution and estimate the parameters often using nonparametric estimates of the unspecified functions like (1.29). Intuitively speaking, the rate of convergence should be somewhere between nonparametric and parametric convergence rates as discussed in 1.1.7 for (1.29). However, many theoretical works of semiparametric estimation have reported that the estimators of the parametric components attain the parametric convergence rate of $\sqrt{n}$ as reviewed in the previous subsections.

We can generally say that semiparametric methods suit best in many econometric applications among the three methods because of the following reasons. As far as the robustness is concerned, nonparametric method must be the best because of its generality, then semiparametric method follows and parametric method is the worst, while efficiency consideration yields the opposite order. Thus there is a trade-off between the efficiency and robustness. When we have the true information on the DGP, parametric method is obviously the best choice because of its efficiency. On the other hand, when we do not know the DGP, nonparametric or semiparametric methods would be better than parametric ones because parametric estimation will yield an inconsistent estimate if the model is incorrectly specified. If we have a sample of extremely large size so that we do not need to care the efficiency of estimation, nonparametric methods would be the best choice, while if the sample size is not so large, semiparametric method may outperform the nonparametric method. In view that many economic data are typically not of very large sample size[6] because economists normally cannot get observations from experiments unlike in natural science and that it is often the case economic theory does not provide the whole information on the functional form of underlying density and/or regression function, but only some aspects of them, semiparametric methods may have the best balance of efficiency and robustness in econometrics.

1.3 Higher order asymptotic theory

We have seen that the density-weighted AD estimator of single index model is $\sqrt{n}$ - consistent and asymptotically normally distributed in 1.1.2. The purpose of this thesis is to study if the nonparametric density derivative estimate can affect the higher order property of the estimator and if so, how it does. We review related topics on higher order asymptotic theory in this section. 1.3.1 and 1.3.2 explain standard Edgeworth expansions of parametric statistics in econometrics and U-statistics respectively. Results reviewed in 1.3.2 are especially closely related to our work since the estimator of our interest (1.35) has a U-statistic form. We employ similar techniques to those used for U-statistics to validate Edgeworth expansions, however we can point out a significant difference between the AD estimator and the

[6] There are some exceptions like datasets from population survey implemented by overnments, when nonparametric methods may be the most suitable.

standard U-statistics. It is that the variance of kernel of standard U-statistics is assumed to be bounded but it is not the case in (1.35). Therefore none of the established results for U-statistics directly apply to the estimator. The last subsection reviews higher order asymptotic theory of semiparametric statistics related to this work.

1.3.1 Edgeworth expansions of parametric statistics

Given an asymptotically normally distributed statistic, we may perform hypothesis testing or confidence interval estimation using the normal approximation. Suppose X_i, i=1,2,... are iid random variables with mean μ_x and variance σ_x^2 , and let $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ and $S_n=\sqrt{n}\sigma_x^{-1}(\bar{X}-\mu_x)$. Since Lindeberg-Levi's central limit theorem gives $S_n \overset{d}{\to} N(0,1)$, we implement confidence interval estimation for μ_x as follows. Because $P(z_{0.05} \le S_n \le z_{0.95}) \approx 0.9$, z_α denoting the α% quantile of the standard normal distribution, we have a 90% confidence interval estimate

$$P(\bar{X}-\tfrac{\sigma_x z_{0.95}}{\sqrt{n}} \le \mu_x \le \bar{X}-\tfrac{\sigma_x z_{0.05}}{\sqrt{n}}) \approx 0.9 \ . \tag{1.75}$$

This estimation is infeasible because of the unknown σ_x . Given $\tilde{\sigma}_x$, a consistent estimate of σ_x , we have $\tilde{S}_n \overset{d}{\to} N(0,1)$ for the studentized statistic $\tilde{S}_n=\frac{1}{n^{1/2}}\sum_{i=1}^{n}(X_i-\mu_x)/\tilde{\sigma}_x$. Then we can construct a feasible 90% confidence interval estimate

$$P(\bar{X}-\tfrac{\tilde{\sigma}_x z_{0.95}}{\sqrt{n}} \le \mu_x \le \bar{X}-\tfrac{\tilde{\sigma}_x z_{0.05}}{\sqrt{n}}) \approx 0.9 \ . \tag{1.76}$$

The precision of the estimate depends on how well the normal distribution can approximate the exact distribution of $\tilde{S}_n$.

Edgeworth expansions aim to asymptotically improve the normal approximation to the exact distribution. Chebyshev (1890), and independently

Edgeworth (1896, 1905), first provide a formal expansion of the distribution function of S_n ,

$$P(S_n \leq x) = \Phi(x) + \phi(x)\{n^{-1/2}p_1(x) + n^{-1}p_2(x) + ... + n^{-j/2}p_j(x) + ...\} \tag{1.77}$$

for a sequence of functions $p_j(x)$. There is, however, no guarantee that this series converges and it motivates us to investigate the behaviour of

$$\sup_x |P(S_n \leq x) - \Phi(x) - \phi(x)\{n^{-1/2}p_1(x) + n^{-1}p_2(x) + ... + n^{-j/2}p_j(x) + ...\}| \ .$$

Berry (1941) and Esseen (1945) analyze a special case when no correction term is involved, namely the bound for simple normal approximation error, and verify the following inequality under the existence of second moment

$$\sup_x |P(S_n \leq x) - \Phi(x)| \leq \frac{33}{4}\frac{E|X_1 - \mu|^3}{\sigma^3 n^{1/2}} \ . \tag{1.78}$$

This kind of bound characterizing the normal approximation error is called the Berry-Esseen bound. The constant 33/4 has been improved, namely reduced, by various authors. Cramèr (1928, 1946) rigorously derives an asymptotic expansion in powers of $n^{-1/2}$ under some strong conditions including $\limsup_{|t|\to\infty} |E(e^{itX})| < 1$, known as the Cramèr condition. Usually we truncate the series (1.77) after including a certain number of terms then we investigate the order of the remainder. We say the expansion is *valid* when

$$\sup_x |P(S_n \leq x) - \Phi(x) - \phi(x)\{n^{-1/2}p_1(x) + ... + n^{-j/2}p_j(x)\}| = o(n^{-j/2}) \tag{1.79}$$

for a fixed j as $n \to \infty$. It is known that the Cramèr condition and $E|X|^{j+2} < \infty$ are sufficient for the validity when $p_i(x)$ are constructed from the Hermite polynomials,

$$p_1(x) = -\frac{\kappa_3}{6}(x^2-1) \ , \ p_2(x) = -\frac{\kappa_4}{24}(x^3-3x) - \frac{\kappa_3^2}{72}(x^5-10x^3+15x) \ , \ ... \tag{1.80}$$

where κ_3 and κ_4 are the third and fourth order cumulants of $(X-\mu_x)/\sigma_x$ respectively (see e.g. Hall (1992), p.42-44 for the construction of $p_i(x)$). R.Rao (1961) first generalizes the univariate results to the multivariate case under the Cramèr condition, which is extended by von Bahr (1967), Bhattacharya (1968, 1971), Chambers (1967) and Bhattacharya and Rao (1976).

It is known S_n and $\tilde{S}_n$ admit valid one term Edgeworth expansions

$$\sup_x |P(S_n \le x) - \{\Phi(x) - \frac{\kappa_3}{6\sqrt{n}}(x^2 - 1)\phi(x)| = o(n^{-1/2}) \tag{1.81}$$

and

$$\sup_x |P(\tilde{S}_n \le x) - \{\Phi(x) + \frac{\kappa_3}{6\sqrt{n}}(2x^2 + 1)\phi(x)| = o(n^{-1/2}) \tag{1.82}$$

respectively (see e.g. Hall (1992), p.70-72). κ_3 in the above is typically unknown so that we think of replacing it by a strongly consistent estimate $\tilde{\kappa}_3$, then we have

$$\sup_x |P(S_n \le x) - \{\Phi(x) - \frac{\tilde{\kappa}_3}{6\sqrt{n}}(x^2 - 1)\phi(x)| = o(n^{-1/2}) \text{ a.s.} \tag{1.83}$$

and

$$\sup_x |P(\tilde{S}_n \le x) - \{\Phi(x) + \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2x^2 + 1)\phi(x)| = o(n^{-1/2}) \text{ a.s.} \tag{1.84}$$

We can modify (1.76), the 90% confidence interval estimation based on the normal approximation, using (1.84) as follows. Letting w_α be the $100\alpha\%$ quantile of the sampling distribution $P(\tilde{S}_n \le x)$, (1.84) yields

$$\alpha = P(\tilde{S}_n \le w_\alpha) = \Phi(w_\alpha) + \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2w_\alpha^2 + 1)\phi(w_\alpha) + o(n^{-1/2}) \text{ a.s.} \tag{1.85}$$

Expanding the right hand side around $w_\alpha = z_\alpha$, we have

$$\alpha \approx \Phi(z_\alpha) + \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2z_\alpha^2 + 1)\phi(z_\alpha) + [1 - \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2z_\alpha^3 - 3z_\alpha)]\phi(z_\alpha)(w_\alpha - z_\alpha) + o(n^{-1/2}) \text{ a.s.}$$

This yields, since $\Phi(z_\alpha) = \alpha$,

$$\begin{aligned} w_\alpha &\approx z_\alpha - [1 - \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2z_\alpha^3 - 3z_\alpha)]^{-1} \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2z_\alpha^2 + 1) + o(n^{-1/2}) \\ &= z_\alpha - \{1 + O(n^{-1/2})\} \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2z_\alpha^2 + 1) + o(n^{-1/2}) \\ &= z_\alpha - \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2z_\alpha^2 + 1) + o(n^{-1/2}) \text{ a.s.} \end{aligned} \tag{1.86}$$

This is called the Cornish-Fisher expansion (see Hall (1992), p.88). Then putting $\tilde{w}_\alpha = z_\alpha - \frac{\tilde{\kappa}_3}{6\sqrt{n}}(2z_\alpha^2 + 1)$, we have a modified interval estimate

$$P(\bar{X} - \frac{\tilde{\sigma}_x}{\sqrt{n}}\tilde{w}_{0.95} \le \mu_x \le \bar{X} - \frac{\tilde{\sigma}_x}{\sqrt{n}}\tilde{w}_{0.05}) \approx 0.9 \tag{1.87}$$

Noting $w_\alpha = z_\alpha + O(n^{-1/2})$ and $w_\alpha = \tilde{w}_\alpha + o(n^{-1/2})$, (1.87) is an asymptotically better estimate than (1.76). Similar consideration is theoretically possible using (1.81)-(1.83), but interval estimation based on them is obviously infeasible due to the unknown σ_x or κ_3 like the infeasibility of (1.75). Similarly we can use (1.84) to determine critical region of a significance test on μ_x . We call an Edgeworth expansion involving unknown quantities depending on the underlying distribution like κ_3 in (1.81) or (1.82) *a theoretical Edgeworth expansion*, while we call the feasible version like (1.83) or (1.84) in which the unknown quantities are replaced by their estimates *an empirical Edgeworth expansion.* Obviously empirical Edgeworth expansions for studentized statistics are for practical use in interval estimation or hypothesis testing. We provide valid theoretical and empirical Edgeworth expansions for unstudentized averaged derivatives in Chapter 2, while we establish them for the studentized statistics in Chapter 3.

As Bhattacharya and Puri (1995) point out, there are roughly two methods of proving the validity of an Edgeworth expansion. One is the direct method, where we expand the characteristic function of the statistic of interest and rearrange it with respect to the sample size, then invert it. The other exploits the asymptotic expansion of the statistic. For example, Bhattacharya and Ghosh (1978) employ the former method to obtain the Edgeworth expansion of $H(S_n)$, based on the Taylor expansion of $H(.)$, where S_n is the sample average of d-dimensional iid random vector and $H(.)$ is a real valued Borel measurable function on R^d . We employ the first approach to validate Edgeworth expansions in the following chapters.

A pioneering work of higher order asymptotic theory in econometrics is Nagar (1959) who derives expansions for the bias and moment matrix of the k-class estimator in powers of $n^{-1/2}$. This sort of expansion of moments is called a Nagar expansion. Based on economic or econometric models, econometricians have developed various kinds of estimators suitable for the models. The estimators are not necessarily functions of a sum of iid random variables, which have been a main interest of statisticians, so that econometricians have developed higher order asymptotic theory for econometric estimators rather independently. Pioneering works on Edgeworth expansions for econometric statistics are Sargan and Mikhail (1971), Sargan (1974, 1975a, 1975b, 1976, 1980), Sargan and Satchell (1986), Phillips (1977, 1978). Sargan and Mikhail (1971) derive a Gram-Charlier type expansion for a single-equation instrumental variable estimator of a simultaneous equation model. Anderson and Sawa (1973) study Edgeworth and Gram-Charlier expansions of k-class estimators, while Mariano (1973a,b) derive Edgeworth expansions for the OLS, 2SLS and Theil's k-class estimators. Sargan (1974) obtains valid Nagar expansions for rational functions of OLS estimators of the reduced form equation coefficients. Anderson (1974) obtains a valid Edgeworth expansion for the LIML estimator. Sargan (1975b) considers Gram-Charlier type expansion of t-ratio for of k-class estimators. Sargan (1976) proves the validity of Edgeworth expansions for 3SLS and FIML estimators, where he exploits the fact that the estimators can written as a function of the first and second sample moments, which is extended by Phillips (1977) to a more general statistic not necessarily written as a function of sample moments. Morimune (1981) obtains valid Edgeworth expansions for an improved (in terms of the first order efficiency) LIML estimator by Morimune (1978) written as a linear combination of the LIML and k-class estimators. Rothenberg (1984a) obtains a valid Edgeworth expansion for GLS estimator. Nonrandomness of exogenous variables and normality of the disturbance terms were necessary for the Edgeworth expansions derived in the above authors except Phillips (1977) and Rothenberg (1984a). Phillips (1980) and Sargan and Satchell (1986) relax the first assumption and derive a valid Edgeworth expansion for a linear dynamic model. Taniguchi (1983, 1991) derive expansions for estimators in Gaussian time series models. Some review papers include Rothenberg (1984b), Magdalinos (1992) and Bhattacharya and Puri (1995). There are a lot of research on higher order asymptotic properties of parametric estimation in econometrics, but only a few have been developed for semiparametric estimators. We review them in 1.3.3.

1.3.2 Asymptotic theory for U-statistics

(1.35), the statistic of interest of this work, has a U-statistic form. We will validate some Edgeworth expansions for this using some technique developed in U-statistic asymptotic theory. We briefly review some asymptotic results of U-statistics.

U-statistics are introduced in a fundamental paper by Hoeffding (1948) as a generalized class of statistics of sample mean. Given a sample $X_1,\cdots,X_n$, U-statistics of order m are defined as

$$U=\binom{n}{m}^{-1}\sum_{1\le i_1<i_2<\cdots<i_m\le n}\sum\cdots\sum h(X_{i_1},X_{i_2},\cdots,X_{i_m}) \tag{1.88}$$

for a known permutation invariant function $h(.)$ called kernel. This class of statistics includes a large number of statistics considered in statistical theory. For example, if $h=h(x_1)$, U is the sample mean, while if $h=I(x_1\le t)$, U is the sample distribution function. If $h=(x_1-x_2)^2/2$, the corresponding U is the sample variance (see Serfling (1980), Chapter 5, for other statistics). Suppose $X_i, i=1,2,\ldots$ are iid observations with distribution function $F(.)$. This statistic can be viewed as an estimator of

$$\theta(F) = E\{h(X_1, X_2, \cdots, X_m)\}$$
$$= \int \cdots \int h(x_1, x_2, \cdots, x_m) dF(x_1) d(x_2) \cdots dF(x_m) \qquad (1.89)$$

supposing it exists because it is straightforward that $E(U) = \theta(F)$. It is also known that U is an efficient estimator of $\theta(F)$ in the sense $Var(U) \leq Var(\tilde{\theta})$ for any $\tilde{\theta}$ satisfying $E(\tilde{\theta}) = \theta(F)$. Suppose $m=2$. Writing

$$U - \theta = \frac{2}{n}\sum_{i=1}^{n} g(X_i) + \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w(X_i, X_j) \qquad (1.90)$$

where

$$\theta = \theta(F) \ , \quad g(X_j) = E\{h(X_i, X_j) - \theta \,|X_j\} \ (i \neq j) \ ,$$
$$w(X_i, X_j) = h(X_i, X_j) - g(X_i) - g(X_j) - \theta \ ,$$

we can prove the strong consistency of U for θ under $E|h(X_i, X_j)| < \infty$ (see e.g. Serfling (1980), p.190). Hoeffding (1948) proves asymptotic normality of $\sqrt{n}(U - \theta)$. Heuristically, the second term on the right of (1.90) times $\sqrt{n}$ converges to zero but the first term times $\sqrt{n}$ converges to a normal variate with mean zero and variance $4Var\{g(X_1)\}$. Hoeffding (1961) and Berk (1966) prove the forward and reverse martingale structure of U respectively expressing it as in (1.90). These results are extended to $m \geq 3$.

Comparing (1.35) and (1.88) with $m=2$, U_{ij} in (1.35) corresponds to $h(X_i, X_j)$ here. However, in view of Lemma 4 of Robinson (1995a) or (A.5) of Appendix A, we have $E|U_{ij}| = O(h^{-1})$ which may diverge as $n \to \infty$. Therefore though (1.35) has a U-statistic form, it even does not satisfy the condition for consistency of U-statistics so that we cannot appeal to the asymptotic results of U-statistics to investigate (1.35).

Higher order asymptotic theory for U-statistics has been developed, especially for $m=2$. Various authors have investigated the Berry-Esseen bound. Let

$$\sigma_g^2 = Var\{g(X_1)\} \ , \ \ \sigma_U^2 = Var(U) \ ,$$

$$\hat{\sigma}_U^2 = \frac{4}{(n-1)(n-2)^2} \sum_{i=1}^{n} [\frac{1}{n-1} \sum_{j \neq i} h(X_i, X_j) - U]^2 \ ,$$

$$D_g = \sup_z |P[\sqrt{n}(U-\theta)/2\sigma_g \le z] - \Phi(z)| \ ,$$

$$D_U = \sup_z |P[\sqrt{n}(U-\theta)/\sigma_U \le z] - \Phi(z)| \ ,$$

$$D_{\hat{U}} = \sup_z |P[\sqrt{n}(U-\theta)/\hat{\sigma}_U \le z] - \Phi(z)| \ .$$

Note that both $2\sigma_g/\sqrt{n}$ and $\sigma_U/\sqrt{n}$ are valid normalizer for U, satisfying $\sigma_U^2 \to 4\sigma_g^2$ and $|\hat{\sigma}_U^2 - \sigma_U^2| + |\hat{\sigma}_U^2 - 4\sigma_g^2| \xrightarrow{p} 0$ (by Callaert and Veraverbeke (1981)). Grams and Serfling (1973) first provides a Berry-Esseen bound,

$$D_g = O(n^{-\frac{r}{2r+1}}) \ \text{under} \ E\{h^{2r}(X_1, X_2)\} < \infty$$

for a positive integer r. Bickel (1974) validates a bound of order $\sqrt{n}$ for the restrictive case when $X_1, \ldots, X_n$ are iid over [0,1] and $|h(x_1, x_2)| < \infty$. Chan and Wierman (1977) show

$$D_U = O(n^{-\frac{1}{2}} \log^{\frac{1}{3}} n) \ \text{under} \ E|h(X_1, X_2)|^3 < \infty$$

and

$$D_U = O(n^{-\frac{1}{2}}) \ \text{under} \ E\{h^4(X_1, X_2)\} < \infty \ .$$

Callaert and Janssen (1978) further weaken the moment condition and prove

$$D_g = O(n^{-\frac{1}{2}}) \ \text{under} \ E|h(X_1, X_2)|^3 < \infty \ .$$

Helmers and van Zwet (1982) slightly relax the moment condition of Callaert and Janssen (1978) to show

$$\sup_z |P[\sqrt{n-1}(U-\theta)/2\sigma_g \le z] - \Phi(z)| = O(n^{-\frac{1}{2}})$$

under $E|g(X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^r < \infty, r > 5/3$. Callaert and Veraverbeke (1981) consider a studentized statistic and show

$$D_{\hat{U}} = O(n^{-1/2}) \ \text{under} \ E\{g^2(X_1)\} < \infty \ \text{and} \ E|h(X_1, X_2)|^{9/2} < \infty \ ,$$

while Helmers (1985) slightly modifies the moment conditions to $E\{g^2(X_1)\} < \infty$ and $E|h(X_1,X_2)|^{4+\delta} < \infty$, $\delta > 0$ for the same bound of $D_{\dot{U}}$.

Callaert, Janssen and Veraverbeke (1980) (CJV hereafter) derive a valid Edgeworth expansion of degree two for U-statistics of the form

$$\sup_z |P[\sqrt{n}\sigma_U^{-1}(U-\theta) \leq z] - E_n(z)| = o(n^{-1})$$

where

$$E_n(z) = \Phi(z) - \phi(z)[\frac{K_3}{6n^{1/2}}(z^2-1) + \frac{K_4}{24n}(z^3-3z) + \frac{K_3^2}{72n}(z^5-10z^3+15z)] , \tag{1.91}$$

$$K_3 = \sigma_g^{-3}[E\{g^3(X_1)\} + 3E\{g(X_1)g(X_2)w(X_1,X_2)\}] , \tag{1.92}$$

$$K_4 = \sigma_g^{-4}\Big[[E\{g^4(X_1)\} - 3\sigma_g^4 + 12E\{g^2(X_1)g(X_2)w(X_1,X_2)\}] + 12E[g(X_2)g(X_3)w(X_1,X_2)w(X_1,X_3)]\Big] ,$$

under certain moment and Cramèr conditions as well as a complicated condition on the characteristic function of $\sum_{j=m+1}^{n} h(X_1,X_j)$ conditional on $(X_{m+1},\ldots,X_n)$, $m = [n^a], a \epsilon (0,1/8)$. They prove the validity of expansion (1.91) by decomposing $\sqrt{n}\sigma_U^{-1}(U-\theta)$ like (1.90), expanding its characteristic function, rearranging it with respect to the sample size, then inverting it. We prove Theorems 1 and 3 in the following chapters in similar manners to these, especially CJV and Callaert and Veraverbeke. We will see in the following chapter an analogous correction term to that in (1.91) appears in the Edgeworth expansion of the AD estimate. Bickel, Götze and van Zwet (1986) prove validity of the same Edgeworth expansion under milder conditions than CJV. They slightly relax the moment conditions of CJV, and replace the complicated condition on the conditional characteristic function by a condition on the eigenvalues of $h(.,.)$. While the above two articles consider scalar U-statistics, Götze (1987) obtains a valid Edgeworth expansion of degree one for multivariate U-statistics. Helmers (1991) investigates into studentized U-statistics and validates the following Edgeworth expansion of degree one.

$$\tilde{E}_n(z) = \Phi(z) + \frac{\phi(z)}{6n^{1/2}\sigma_g^3}[(2z^2+1)E\{g^3(X_1)\} + 3(z^2+1)E\{g(X_1)g(X_2)w(X_1,X_2)\}] \quad . \qquad (1.93)$$

We will see the correction terms here are analogous to those appearing in the Edgeworth expansion of studentized averaged derivatives in Chapter 3. Bentkus, Götze and van Zwet (1997) consider a more general asymptotically normal statistics than U-statistics which are symmetric function of n iid random variables. They validate an Edgeworth expansion of degree one with remainder of order n^{-1} based on an expansion of the statistics in a series of U-statistics of increasing order.

We refer to Serfling (1980, Chapter 5) and Lee (1990) for comprehensive treatments of U-statistics.

1.3.3 Higher order theory of semiparametric statistics

Consistency, asymptotic distribution and efficiency of semiparametric estimators have been intensively studied as seen in the previous section, however there have not been a lot of studies on their higher order asymptotic properties. Linton (1995a, 1996b) develop both Nagar expansions and asymptotic expansions for estimators of a partially linear model and of a linear regression model with disturbance heteroscedasticity of unknown form. Linton (1996b) finds that the leading terms are of order greater than $n^{-\frac{1}{2}}$ and shows how their contribution might be minimized by a plug-in type choice of bandwidth, attaining second order optimality. Robinson (1995a) studies a Berry-Esseen bound for (1.35) as mentioned in 1.1.7

which is extended in this thesis to obtain Edgeworth expansions. Linton (1996a) is also closely related to this work in the sense that he establishes valid Edgeworth expansions for a wide class of semiparametric estimators. He is concerned with estimators obtained by minimizing the objective function

$$\frac{1}{n}\sum_{i=1}^{n} \Psi\,(X_i;\beta,\tilde{G}(Z_i))$$

where $\Psi(.;.,.)$ is a function satisfying certain regularity conditions, (X,Z) are vectors of observables, β is a vector of unknown parameters of interest and $\tilde{G}(.)$ is a nonparametric estimate of unknown (nuisance) function $G(.)$, having a U-statistic form

$$n^{-1/2}\sum_{i=1}^{n} g_1(X_i)+n^{-\zeta}\sum_{j>i}\sum \rho_{ij}g_2(X_i,X_j)+n^{-\xi}\sum_{k>j>i}\sum\sum \pi_{ijk}g_3(X_i,X_j,X_k)$$

where ρ_{ij} and π_{ijk} are deterministic weights, $1/2<\zeta<\xi$, both g_2 and g_3 are permutation invariant satisfying $E[g_2(X_i,X_j)]=E[g_3(X_i,X_j,X_k)]=0$. Making assumptions of a high-level type, including that the nonparametric estimate converges suitably fast, he shows that the estimator, suitably centred and normalized, possesses a valid theoretical Edgeworth expansion

$$\tilde{F}(z) = \Phi(z) - \phi(z)\left[\frac{\delta_3}{6n^{1/2}}(z^2-1)+\frac{\delta_4}{24n}(z^3-3z) + \frac{\delta_3^2}{72n}(z^5-10z^3+15z)\right]$$

where $\Phi(.)$ and $\phi(.)$ are cumulative distribution function and density function of a standard normal variate respectively, δ_i depends on the weights and moments of $g_1(X_i)$, $g_2(X_i,X_j)$ and $g_3(X_i,X_j,X_k)$. He shows that his assumptions can be satisfied by a version of the partially linear model as well as in models where no smoothing is involved. Comparing this expansion with (1.91), we observe they have the same functional form with respect to z and n, with different coefficients. So the nonparametric estimation has no effect on expansions to order n^{-1}. We closely investigate the relations between Linton (1996b) and the results of this work and show how his results cannot be applied to our specified semiparametric averaged derivative estimation in Chapter 3.

Chapter 2

Edgeworth Expansions for Averaged Derivatives - Normalized Case [1]

This chapter validates Edgeworth expansions of (1.35) suitably normalized by its asymptotic covariance. Section 2.1 introduces some notations and assumptions. Section 2.2 and 2.3 establish valid theoretical and empirical Edgeworth expansions respectively. The expansions involve three correction terms. Two of them are related to the nonparametric density estimate and the rest is "parametric" whose analogue appears in the Edgeworth expansion of U-statistics (1.91). The expansion suggests that some correction term(s) can dominate other(s) asymptotically depending on the bandwidth choice, dimension of regressors and kernel order. We discuss it thoroughly in Section 3.3 in terms of studentized statistics.

2.1 Notations and assumptions

We have reviewed U estimates $\bar{\mu} = -E\{g'(X)f(X)\}$ with first order asymptotic properties described in Section 1.1. Additional to the notations in 1.1.2, we introduce some more of them to describe assumptions for the Edgeworth expansions. For a function $k: R^d \to R$, write

$$k = k(X)\ , \quad k' = \partial k / \partial x\ , \quad k'' = \partial^2 k / \partial x \partial x^\tau\ , \quad h''' = \partial vec\ (k'') / \partial x^\tau$$

and define

$$q = E(Y^2|X)\ , \quad r = E(Y^3|X)\ ,$$

$$\mu = \mu(X, Y) = Yf' - e'\ , \quad e = fg\ ,$$

$$a = g'f + \bar{\mu}\ , \quad a' = \partial a / \partial x^\tau\ ,$$

$$\Sigma = 4E(\mu - \bar{\mu})(\mu - \bar{\mu})^\tau\ .$$

---

[1] This chapter has been revised and written up as a joint paper with my supervisor Professor Peter M. Robinson as Nishiyama and Robinson (2000). It is forthcoming in Econometrica.

We introduce the following assumptions to establish valid Edgeworth expansions.

(i) $E|Y|^3 < \infty$.

(i)' $E|Y|^6 < \infty$.

(i)" $E|Y|^{8+\rho} < \infty$ for some $\rho > 0$.

(ii) Σ is finite and positive definite.

(iii) The underlying measure of (X^τ, Y) can be written as $\mu_X \times \mu_Y$, where μ_X and μ_Y are Lebesgue measure on R^d and R respectively. (X_i^T, Y_i) are iid observations on (X^τ, Y) .

(iv) f is (L+1) times differentiable, and f and its first (L+1) derivatives are bounded for 2L>d+2.

(iv)' f is (L+2) times differentiable, and f and its first (L+2) derivatives are bounded, where 2L>d+2.

(v) g is (L+1) times differentiable, and e and its first (L+1) derivatives are bounded.

(v)' g is (L+2) times differentiable, and e and its first (L+2) derivatives are bounded.

(vi) q is twice differentiable and q', q'', g', g'', g''', $E(|Y|^3 \,|X)f$, and qf' are bounded.

(vi)' q is twice differentiable and q', q'', g', g'', g''', $E(Y^4 \,|X)f$, and qf' are bounded.

(vi)" q is differentiable and qf , $q'f$ and qf' are bounded.

(vii) f, gf, $g'f$, and qf vanish on the boundaries of their convex (possibly infinite) supports.

(vii)' f, gf and qf vanish on the boundaries of their convex (possibly infinite) supports.

(viii) $K(u)$ is even, differentiable,

$$\int_{R^d}\{(1+\|u\|^L)\,|K(u)|+\|K'(u)\|\}\,du + \sup_{u\in R^d}\|K'(u)\| < \infty ,$$

and for the same L as in (iv) and (v),

$$\int_{R^d} u_1^{l_1}\cdots u_d^{l_d} K(u)\,du \begin{cases} = 1, & \text{if } \quad l_1+\cdots+l_d=0 \\ = 0, & \text{if } \quad 0<l_1+\cdots+l_d<L \\ \neq 0, & \text{if } \quad l_1+\cdots+l_d=L \ . \end{cases}$$

(ix) $\dfrac{(\log n)^9}{nh^{d+2}} + nh^{2L} \to 0$ as $n\to\infty$.

(ix)' $\dfrac{(\log n)^9}{nh^{d+3}} + nh^{2L} \to 0$ as $n\to\infty$.

(ix)" $\dfrac{(\log n)^6}{nh^{d+2}} = O(1)$, $nh^{2L} \to 0$ as $n\to\infty$.

(x) $\limsup_{|t|\to\infty} |E\exp[\{i\,t\,2\sigma^{-1}\nu^{\tau}(\mu-\bar{\mu})\}]| < 1$ for any vector ν satisfying $\nu^{\tau}\nu=1$.

(xi) $H(u)$ is even in all arguments u_i, $i=1,\cdots,d$ and (L+1) times differentiable,

$$\int_{R^d} H(u)\,du = 1$$

and

$$\int_{R^d}\|\Delta^{(l_1,\cdots,l_d)}H'(u)\|\,du + \sup_{u\in R^d}\|\Delta^{(l_1,\cdots,l_d)}H'(u)\| < \infty$$

for all integers $l_1,\cdots,l_d$ satisfying $0\le l_1+\cdots+l_d\le L$, where

$$\Delta^{(l_1,\cdots,l_d)}h(x) = \frac{\partial^{(l_1+\cdots+l_d)}}{\partial x_1^{l_1}\cdots\partial x_d^{l_d}}h(x) \text{ for a function } h:\ R^d \to R \ .$$

(xii) $b\to 0$ and $\dfrac{(\log n)^2}{nb^{d+2+2L}} = O(1)$ as $n\to\infty$.

Assumptions (i)-(iv), (vi)", (vii)', and (viii) are identical to corresponding ones of Robinson (1995a), which are discussed there. Assumptions (v)-(vii) and (ix) somewhat strengthen corresponding ones of Robinson (1995a), and assumption (x) is a Cramèr condition (see e.g. Bhattacharya and Rao (1976)). Assumptions with primes somewhat strengthen or weaken those without primes used in Theorem 2, 3, 4 and 5. Notice that H needs only be a second-order kernel, whereas K has to be a higher-order one unless $d=1$. It is possible to choose $H(u)=K(u)$ with assumptions (viii) and (xi) simultaneously satisfied. However, in comparing (xii) with (ix) it seems that b should in general be chosen larger than h.

2.2 A theoretical Edgeworth expansion

Define for $d\times 1$ vector ν, $\sigma^2 = \nu^\tau \Sigma \nu$,

$$Z = n^{1/2}\sigma^{-1}\nu^\tau(U-\bar{\mu}) \ , \ F(z) = P(Z\leq z) \ ,$$

$$\tilde{F}(z) = \Phi(z) - \phi(z)\left\{n^{1/2}h^L\kappa_1 + \frac{\kappa_2}{nh^{d+2}}z + \frac{4(\kappa_3+3\kappa_4)}{3n^{1/2}}(z^2-1)\right\}, \quad (2.1)$$

where z is real-valued, Φ and ϕ are respectively the distribution and density function of a standard normal variate, and

$$\kappa_1 = \frac{2(-1)^L\sigma^{-1}}{L!}\sum_{\substack{0\leq l_1,\cdots,l_d\leq L \\ l_1+\cdots+l_d=L}}\cdots\sum\left\{\int\prod_{i=1}^{d}u_i^{l_i}K(u)\,du\right\}E[(\Delta^{(l_1,\cdots,l_d)}\nu^\tau f')g], \quad (2.2)$$

where

$$\kappa_2 = 2\sigma^{-2}\int\{\nu^\tau K'(u)\}^2du\ E\{(q-g^2)f\} \ ,$$

$$\kappa_3 = \sigma^{-3}E\left[\{r-3(q-g^2)g-g^3\}(\nu^\tau f')^3 - 3(q-g^2)(\nu^\tau f')^2(\nu^\tau a) - (\nu^\tau a)^3\right],$$

$$\kappa_4 = -\sigma^{-3}E\Big[f(q-g^2)(\nu^\tau f')(\nu^\tau a'\nu) - f(\nu^\tau f')\{\nu^\tau(q'-2gg')\}(\nu^\tau a)$$

$$- f(q-g^2)(\nu^\tau a)(\nu^\tau f''\nu) + f(\nu^\tau g')(\nu^\tau a)^2\Big],$$

where $f'' = \dfrac{\partial^2 f(X)}{\partial x\,\partial x^\tau}$ and $a' = \dfrac{\partial a(X)}{\partial x^\tau}$.

THEOREM 1 : *Under assumptions* (i)-(x), *as* $n \to \infty$,

$$\sup_{\nu:\, \nu^T \nu = 1} \sup_{z \in R} |F(z) - \tilde{F}(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \quad .$$

Theorem 1 establishes a valid Edgeworth expansion for a single linear combination of the vector averaged derivative statistics U. The development of full multivariate expansions would require further work (we cannot appeal to the Cramèr-Rao device), however our present setup allows us higher-order inference on individual elements of $\bar{\mu}$, which would be of practical importance, as well as on its arbitrary single linear combination. The normalisation $\nu^T \nu = 1$ employed here differs from that used by Robinson (1995a), namely $\nu^T \Sigma \nu = 1$, which is infeasible.

κ_1 and κ_2 involve the kernel function $K(.)$ and the corresponding correction terms depend on the bandwidth h, thus they are related to nonparametric estimate of $f'(x)$. Also κ_1 and κ_2 are respectively limits of $\sigma_\nu^{-1} \nu^T (EU - \bar{\mu}) / h^L$ and $h^{d+2} E(W_{12}^2)$ (see Lemmas 11 and 12) so that κ_1 and κ_2 are related to first and second moments of U. In standard parametric higher-order theory κ_1 and κ_2 do not arise since unbiased statistics with variance $O(n^{-1})$ are typically considered, not, as here, $O(n^{-1}h^{-d-2})$. κ_3 and κ_4 do not involve $K(.)$ and the corresponding correction term is of parametric order and independent of h. We find the similarity of the last term in the wave bracket of (2.1) to the first term in the square bracket of (1.91) in the following sense. Firstly the functional form with respect to z and n is the same, and secondly since we have

$$\kappa_3+3\kappa_4=E(V_1^3)+3E(V_1V_2W_{12})+o(1)$$

due to Lemmas 13 and 14 (see below for the definition of V_i and W_{ij}), we easily see this structure is analogous to (1.92). The difference of constants 1/6 and 4/3 comes from the different choice of the normalizer. Therefore we can regard the terms involving κ_1 and κ_2 as nonparametric, while that involving κ_3 and κ_4 as parametric.

We follow a similar line to CJV (1980) to prove the theorem. We first decompose the statistic Z into (1.37) (or (1.90) for ordinary U-statistics) plus a bias term. We further decompose the first term of (1.37) into a term of exactly $O_p(n^{-1/2})$ and the remainder. Then we expand the characteristic function of U, rearrange the decomposed terms with respect to the orders of $n^{1/2}h^L$, $n^{-1}h^{-d-2}$ and $n^{-1/2}$. Then we invert it using Esseen's smoothing lemma (see e.g. Bhattacharya and Rao, 1976, Chapter 3).

PROOF OF THEOREM 1

Let C denote a generic, finite, positive constant. The qualification "for n sufficiently large" will sometimes be omitted. Writing $E(\cdot\,|i)=E(\cdot\,|X_i,Y_i)$ and

$\mu_i = \mu(X_i, Y_i)$, let

$$\nu_i = \sigma^{-1}\mathrm{v}^\tau(\mu_i - \bar{\mu}) \quad , \ V_i = \sigma^{-1}\mathrm{v}^\tau(U_i - EU) \quad , \ W_{ij} = \sigma^{-1}\mathrm{v}^\tau(U_{ij} - EU) - V_i - V_j$$

where U_i is defined in (1.38). Z can be decomposed as follows.

$$\begin{aligned} Z &= \frac{2}{n^{1/2}}\sum_{i=1}^{n}\nu_i + \frac{2}{n^{1/2}}\sum_{i=1}^{n}(V_i - \nu_i) \\ &\quad + n^{1/2}\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_{ij} + n^{1/2}\sigma^{-1}\mathrm{v}^\tau(EU - \bar{\mu}) \\ &= \bar{\nu} + \bar{\omega} + \bar{W} + \Delta. \end{aligned}$$

Define

$$\chi(t) = E(e^{itZ}) = e^{it\Delta}E[e^{it(\bar{\nu}+\bar{\omega}+\bar{W})}] , \tag{2.3}$$

$$\begin{aligned} \tilde{\chi}(t) &= \int e^{itz}d\tilde{F}(z) \\ &= e^{-\frac{t^2}{2}}\left\{1 + n^{1/2}h^{L}\kappa_1(it) + \frac{\kappa_2}{nh^{d+2}}(it)^2 + \frac{4(\kappa_3 + 3\kappa_4)}{3n^{1/2}}(it)^3\right\} . \end{aligned} \tag{2.4}$$

By Esseen's smoothing Lemma,

$$\begin{aligned} \sup_{z\in R} |F(z) - \tilde{F}(z)| &\le \int_{-n^{1/2}logn}^{n^{1/2}logn} \left|\frac{\chi(t) - \tilde{\chi}(t)}{t}\right| dt + O\left(\frac{1}{n^{1/2}logn}\right) \\ &\le \int_{-p_1}^{p_1}\left|\frac{\chi(t) - \tilde{\chi}(t)}{t}\right| dt + \int_{p_1 \le |t| \le n^{1/2}logn} \left|\frac{\chi(t)}{t}\right| dt \\ &\quad + \int_{|t| \ge p_1}\left|\frac{\tilde{\chi}(t)}{t}\right| dt + o(n^{-1/2}) \\ &= \text{(I-1)} + \text{(II-1)} + \text{(III-1)} + o(n^{-1/2}) , \end{aligned} \tag{2.5}$$

where $p_1 = \min(\epsilon n^{1/2}, \log\ n)$ for $0 < \epsilon = 1/E|2\nu_1|^3$. Before estimating (2.5), we mention an inequality frequently used hereafter:

$$\left|e^{ix} - 1 - ix - \frac{(ix)^2}{2!} - \dots - \frac{(ix)^k}{k!}\right| \le \frac{|x|^{k+1}}{(k+1)!} \tag{2.6}$$

for real x and integer k .

To estimate (I-1), we represent $\chi(t)$ as $\tilde{\chi}(t)$ plus a remainder. Using (2.6),

$$e^{it\Delta} = 1 + it\Delta + O((t\Delta)^2) , \tag{2.7}$$

$$E\{e^{it(\bar{\nu}+\bar{\omega}+\bar{W})}\} = E\{e^{it(\bar{\nu}+\bar{W})}\} + O(E|t\bar{\omega}|)$$

$$= E\left[e^{it\bar{v}}\left\{1+it\bar{W}+\frac{1}{2}(it\bar{W})^2\right\}\right] + O(E|t\bar{W}|^3) + O(E|t\bar{\omega}|) \quad . (2.8)$$

Writing $w_j(t) = e^{2itn^{-1/2}v_j}$ and $\gamma_v(t) = E[w_j(t)]$,

$$E(e^{it\bar{v}}) = \gamma_v(t)^n \qquad (2.9)$$

$$E(e^{it\bar{v}}\bar{W}) = n^{1/2}\gamma_v(t)^{n-2}E[W_{12}w_1(t)w_2(t)]$$

$$= \gamma_v(t)^{n-2}\left\{\frac{4(it)^2}{n^{1/2}}E(W_{12}v_1v_2) + O\left(\left(\frac{|t|^3}{n}+\frac{t^4}{n^{3/2}}\right)h^{-1}\right)\right\} , (2.10)$$

by Lemma 8, and

$$E(e^{it\bar{v}}\bar{W}^2) = n\binom{n}{2}^{-2}\binom{n}{2}\gamma_v(t)^{n-2}E[W_{12}^2w_1(t)w_2(t)]$$

$$+ 6n\binom{n}{2}^{-2}\binom{n}{3}\gamma_v(t)^{n-3}E[W_{12}W_{13}w_1(t)w_2(t)w_3(t)]$$

$$+ 6n\binom{n}{2}^{-2}\binom{n}{4}\gamma_v(t)^{n-4}[E\{W_{12}w_1(t)w_2(t)\}]^2$$

$$= \frac{2}{n-1}\gamma_v(t)^{n-2}\left\{E(W_{12}^2) + O\left(|t|n^{-\frac{1}{2}}h^{-d-2}\right)\right\}$$

$$+ \gamma_v(t)^{n-3}O\left(|t|^3n^{-\frac{3}{2}}h^{-2}\right)$$

$$+ \gamma_v(t)^{n-4}O(t^4n^{-1}+t^6n^{-2}h^{-2}+t^8n^{-3}h^{-2}) , \qquad (2.11)$$

by Lemmas 8-10 and 14-(a). By Feller (1971, p.534) and Lemma 1-(a), for $|t| \le \epsilon n^{1/2}$ and $m=0,1,2,3,4$,

$$\gamma_v(t)^{n-m} = e^{-\frac{t^2}{2}}\left\{1+\frac{E(2v_1)^3}{6n^{1/2}}(it)^3\right\} + o\left(n^{-1/2}(|t|^3+t^6)e^{-\frac{t^2}{4}}\right) \quad . (2.12)$$

By (2.3), (2.7)-(2.12), (A.7) and Lemma 7,

$$\chi(t) = \left[1+it\Delta+O(t^2\Delta^2)\right]\left[e^{-\frac{t^2}{2}}\left\{1+\frac{4E(v_1^3)}{3n^{1/2}}(it)^3\right\}+o\left(n^{-1/2}(|t|^3+t^6)e^{-\frac{t^2}{4}}\right)\right]$$

$$\times\left[1+\frac{4(it)^3}{n^{1/2}}E(W_{12}v_1v_2)+\frac{(it)^2}{n}E(W_{12}^2)+A_n\right] , \qquad (2.13)$$

where

$$A_n = O\left(\frac{t^4}{nh}+\frac{|t|^5}{n^{3/2}h^2}+\frac{t^6}{n}+\frac{t^8}{n^2h^2}+\frac{t^{10}}{n^3h^2}+|t|h^L+\frac{|t|^3}{(nh^{d+2})^{3/2}}\right)$$

$$= o\left(\frac{t^2+t^{10}}{nh^{d+2}}\right) + O(|t|h^L) \qquad (2.14)$$

by (ix). Expanding the right hand side of (2.13) and using $\Delta^2 = n\{v^\tau(EU-\bar{\mu})/\sigma\}^2 = O(nh^{2L})$ due to (A.1),

$$\chi(t) = e^{-\frac{t^2}{2}}\left\{1+it\Delta+\frac{E(W_{12}^2)}{n}(it)^2+\frac{4E(v_1^3)+12E(W_{12}v_1v_2)}{3n^{1/2}}(it)^3\right\}$$
$$+\ o\left(\frac{1}{n^{1/2}}e^{-\frac{t^2}{4}}\left(|t|^3+|t|^{11}\right)\right)+o\left(\frac{1}{nh^{d+2}}e^{-\frac{t^2}{4}}\left(|t|^3+t^4\right)\right)$$
$$+\ O\left(nh^{2L}t^2e^{-\frac{t^2}{2}}\right)+O\left(A_n\left(e^{-\frac{t^2}{4}}+e^{-\frac{t^2}{2}}\right)(1+t^8)\right). \qquad (2.15)$$

Since the first term on the right of (2.15) is

$$\tilde{\chi}(t)+o\left(e^{-\frac{t^2}{2}}\left(|t|n^{1/2}h^L+t^2n^{-1}h^{-d-2}+|t|^3n^{-1/2}\right)\right)$$

by (2.4) and Lemmas 11, 12, 13-(a) and 14-(a), using (2.14),

$$\text{(I-1)} = \int_{-p_1}^{p_1}\left|\frac{\chi(t)-\tilde{\chi}(t)}{t}\right|dt = o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L)$$

because $\int_{-\infty}^{\infty}|t|^r(e^{-t^2/2}+e^{-t^2/4})\,dt < C$ for any positive constant r.

Next, (III-1) is

$$\int_{|t|\ge p_1}\frac{1}{|t|}e^{-\frac{t^2}{2}}\left|1+n^{1/2}h^L\kappa_1 it+\frac{\kappa_2}{nh^{d+2}}(it)^2+\frac{\kappa_3}{3n^{1/2}}(it)^3\right|dt$$
$$\le C\left[\int_{p_1}^{\infty}\frac{1}{t}e^{-\frac{t^2}{2}}dt+n^{1/2}h^L\int_{p_1}^{\infty}e^{-\frac{t^2}{2}}dt\right.$$
$$\left.+\frac{1}{nh^{d+2}}\int_{p_1}^{\infty}te^{-\frac{t^2}{2}}dt+\frac{1}{n^{1/2}}\int_{p_1}^{\infty}t^2e^{-\frac{t^2}{2}}dt\right].$$

The first integral is smaller than $p_1^{-2}\int_{p_1}^{\infty}te^{-\frac{t^2}{2}}dt = p_1^{-2}e^{-\frac{p_1^2}{2}} = o(n^{-1})$ because $p_1 = \min(\log n, \epsilon n^{1/2})$. The other integrals are clearly $o(1)$ as $n\to\infty$. It follows that (III-1) $= o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L)$.

To estimate (II-1), define, for $m=1,2,\cdots,n$,

$$\bar{v}(m) = \frac{2}{n^{1/2}}\sum_{i=1}^{m}v_i\ ,\quad \bar{\omega}(m) = \frac{2}{n^{1/2}}\sum_{i=1}^{m}(V_i-v_i)\ ,$$

and, for $m=1,2,\cdots,n-1$,

$$\bar{W}(m) = n^{1/2}\binom{n}{2}^{-1}\sum_{i=1}^{m}\sum_{j=i+1}^{n}W_{ij}\ .$$

Note that $\bar{v}(n)=\bar{v}$, $\bar{\omega}(n)=\bar{\omega}$, $\bar{W}(n-1)=\bar{W}$. Using (2.6),

$$|\chi(t)| = |e^{it\Delta}Ee^{it(\bar{v}+\bar{\omega}+\bar{W})}| \le |Ee^{it(\bar{v}+\bar{\omega}+\bar{W})}|$$
$$\le\left|Ee^{it\{\bar{v}+(\bar{\omega}-\bar{\omega}(m))+(\bar{W}-\bar{W}(m))\}}\left\{1+it(\bar{\omega}(m)+\bar{W}(m))+\frac{(it)^2}{2}(\bar{\omega}(m)+\bar{W}(m))^2\right\}\right|$$
$$+\ O(|t|^3E(|\bar{\omega}(m)|^3+|\bar{W}(m)|^3))$$

$$\le |Ee^{it\{\bar{v}+(\bar{\omega}-\bar{\omega}(m))+(\bar{W}-\bar{W}(m))\}}|$$

$$+ |t|\,|Ee^{it\{\bar{v}+(\bar{\omega}-\bar{\omega}(m))+(\bar{W}-\bar{W}(m))\}}\{\bar{\omega}(m)+\bar{W}(m)\}|$$

$$+ \frac{t^2}{2}|Ee^{it\{\bar{v}+(\bar{\omega}-\bar{\omega}(m))+(\bar{W}-\bar{W}(m))\}}\{\bar{\omega}(m)+\bar{W}(m)\}^2|$$

$$+ O(|t|^3E\{|\bar{\omega}(m)|^3+|\bar{W}(m)|^3\}) \ . \tag{2.16}$$

By (iii) and Lemma 2, the first term in (2.16) is bounded by

$$|Ee^{it\bar{v}(m)}Ee^{it\{(\bar{v}-\bar{v}(m))+(\bar{\omega}-\bar{\omega}(m))+(\bar{W}-\bar{W}(m))\}}| \le |Ee^{it\bar{v}(m)}| = |\gamma_v(t)|^m \ . \tag{2.17}$$

Using (iii), Lemma 1-(c), Lemma 2, Lemma 3, (A.5), and Hölder's inequality, the second term in (2.16) is bounded by $|t|$ times

$$|E\{e^{it\bar{v}(m)}\bar{\omega}(m)\}| + |E\{e^{it\bar{v}(m)}\bar{W}(m)\}|$$
$$\le |\gamma_v(t)|^{m-1}\frac{2m}{n^{1/2}}E|V_1-v_1| + |\gamma_v(t)|^{m-2}\frac{m(m-1)}{2}n^{1/2}\binom{n}{2}^{-1}E|W_{12}|$$
$$\le C|\gamma_v(t)|^{m-1}\frac{m}{n^{1/2}}h^L + C|\gamma_v(t)|^{m-2}\frac{m^2}{n^{3/2}h} \ . \tag{2.18}$$

Similarly to the derivation of (2.18), using (iii), Lemmas 1-5, (A.5) and Hölder's inequality, the third term in (2.16) is bounded by $t^2/2$ times

$$|Ee^{it\bar{v}(m)}\bar{\omega}(m)^2|+2|Ee^{it\bar{v}(m)}\bar{\omega}(m)\bar{W}(m)|+|Ee^{it\bar{v}(m)}\bar{W}(m)^2|$$

$$\le |\gamma_v(t)|^{m-1}\frac{4m}{n}E(V_1-v_1)^2 + |\gamma_v(t)|^{m-2}\frac{4m(m-1)}{n}E|V_1-v_1|E|V_2-v_2|$$

$$+2|\gamma_v(t)|^{m-2}\frac{m(m-1)}{2}\binom{n}{2}^{-1}2E|(V_1-v_1)W_{12}|$$

$$+2|\gamma_v(t)|^{m-3}\frac{m(m-1)(m-2)}{2}\binom{n}{2}^{-1}2E|V_1-v_1|E|W_{23}|$$

$$+|\gamma_v(t)|^{m-2}(mn)n\binom{n}{2}^{-2}E(W_{12}^2)$$

$$+|\gamma_v(t)|^{m-3}\{2m(n-1)n+4(m-1)mn\}n\binom{n}{2}^{-2}E|W_{12}W_{13}|$$

$$+6|\gamma_v(t)|^{m-4}\frac{m(m-1)(m-2)(m-3)}{24}n\binom{n}{2}^{-2}E|W_{12}|E|W_{34}|$$

$$\le C\left\{|\gamma_v(t)|^{m-1}m\frac{h^{2L}}{n} + |\gamma_v(t)|^{m-2}m^2\frac{h^{2L}}{n} + |\gamma_v(t)|^{m-2}m^2\frac{h^{L-1}}{n^2}\right.$$

$$+|\gamma_v(t)|^{m-3}m^3\frac{h^{L-1}}{n^2} + |\gamma_v(t)|^{m-2}\frac{mn}{n^3h^{d+2}} + |\gamma_v(t)|^{m-3}\frac{mn^2+m^2n}{n^3h^2}$$

$$\left.+|\gamma_v(t)|^{m-4}\frac{m^4}{n^3h^2}\right\}$$

$$\leq C|\gamma_v(t)|^{m-4}\left\{\left(\frac{h^{2L}}{n}+\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)m+\left(\frac{h^{2L}}{n}+\frac{h^{L-1}}{n^2}+\frac{1}{n^2h^2}\right)m^2\right.$$

$$\left.+\frac{h^{L-1}}{n^2}m^3+\frac{1}{n^3h^2}m^4\right\}, \tag{2.19}$$

because $|\gamma_v(t)|\leq 1$. Substituting (2.17)-(2.19) into (2.16), with $|\gamma_v(t)|\leq 1$ yields

$$|\chi(t)| \leq |\gamma_v(t)|^m+C\left\{|\gamma_v(t)|^{m-1}m\frac{h^L}{n^{1/2}}+|\gamma_v(t)|^{m-2}\frac{m^2}{n^{3/2}h}\right\}|t|$$

$$+C|\gamma_v(t)|^{m-4}\left\{\left(\frac{h^{2L}}{n}+\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)m+\left(\frac{h^{2L}}{n}+\frac{h^{L-1}}{n^2}+\frac{1}{n^2h^2}\right)m^2\right.$$

$$\left.+\frac{h^{L-1}}{n^2}m^3+\frac{1}{n^3h^2}m^4\right\}t^2$$

$$+O(|t|^3E\{|\bar{\omega}(m)|^3+|\bar{W}(m)|^3\})$$

$$\leq C|\gamma_v(t)|^{m-4}\left[1+\left\{\frac{h^L}{n^{1/2}}|t|+\left(\frac{h^{2L}}{n}+\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2\right\}m\right.$$

$$+\left\{\frac{1}{n^{3/2}h}|t|+\left(\frac{h^{2L}}{n}+\frac{h^{L-1}}{n^2}+\frac{1}{n^2h^2}\right)t^2\right\}m^2$$

$$\left.+\frac{h^{L-1}}{n^2}t^2m^3+\frac{1}{n^3h^2}t^2m^4\right]$$

$$+O(|t|^3E\{|\bar{\omega}(m)|^3+|\bar{W}(m)|^3\}). \tag{2.20}$$

Now we evaluate (II-1), partitioning its range of integration into two parts, namely $p_1\leq|t|\leq\epsilon n^{1/2}$ and $\epsilon n^{1/2}\leq|t|\leq n^{1/2}\log n$.

(i) For $p_1\leq|t|\leq\epsilon n^{1/2}$, since $\epsilon=1/E|2v_1|^3\leq\{1/E(2v_1)^2\}^{3/2}=1$ due to Jensen's inequality and so $|t|\leq n^{1/2}$, using (2.6) and Lemma 1-(a),

$$|\gamma_v(t)| = |Ee^{it\frac{2}{\sqrt{n}}v_1}|$$

$$\leq|E\{1+it\frac{2v_1}{n^{1/2}}+\frac{(it)^2}{2!}(\frac{2v_1}{n^{1/2}})^2\}|+\frac{|t|^3}{3!}E|\frac{2v_1}{n^{1/2}}|^3$$

$$\leq|1-\frac{t^2}{2n}|+\frac{t^2}{6n}=1-\frac{t^2}{3n}\leq\exp\left(-\frac{t^2}{3n}\right), \tag{2.21}$$

so using (2.21), Lemma 7, and (A.7) in (2.20),

$$|\chi(t)| \leq C\exp\left\{-\frac{(m-4)t^2}{3n}\right\}\left[1+\left\{\frac{h^L}{n^{1/2}}|t|+\left(\frac{h^{2L}}{n}+\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2\right\}m\right.$$

$$+\left\{\frac{1}{n^{3/2}h}|t|+\left(\frac{h^{2L}}{n}+\frac{h^{L-1}}{n^2}+\frac{1}{n^2h^2}\right)t^2\right\}m^2$$

$$+\frac{h^{L-1}}{n^2}t^2m^3+\frac{1}{n^3h^2}t^2m^4\Big]$$

$$+O\left(|t|^3\left\{\left(\frac{m}{n}\right)^{3/2}h^{3L}+\left(\frac{m}{n^2h^{d+2}}\right)^{3/2}\right\}\right). \tag{2.22}$$

We may take $m=[9n\log n/t^2]$ since $1\le m\le n-1$ holds for $p_1\le|t|\le\epsilon n^{1/2}$ and sufficiently large n. Because $m\ge(9n\log n)/t^2-1$,

$$\exp\left\{-\frac{(m-4)t^2}{3n}\right\}=\exp\left\{-\frac{(m+1)t^2}{3n}\right\}\exp\left(\frac{5t^2}{3n}\right)$$
$$\le C\exp(-3logn)\le\frac{C}{n^3} \tag{2.23}$$

for $|t|\le\epsilon n^{1/2}$. Substituting (2.23) into (2.22) using $m\le(9n\log n)/t^2$, we derive

$$|\chi(t)|\le\frac{C}{n^3}\Big[1+\left\{\frac{h^L}{n^{1/2}}|t|+\left(\frac{h^{2L}}{n}+\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2\right\}\frac{nlogn}{t^2}$$
$$+\left\{\frac{1}{n^{3/2}h}|t|+\left(\frac{h^{2L}}{n}+\frac{h^{L-1}}{n^2}+\frac{1}{n^2h^2}\right)t^2\right\}\frac{n^2(logn)^2}{t^4}$$
$$+\frac{h^{L-1}}{n^2}t^2\frac{n^3(logn)^3}{t^6}+\frac{1}{n^3h^2}t^2\frac{n^4(logn)^4}{t^8}\Big]$$
$$+O\left((logn)^{3/2}h^{3L}+\left(\frac{logn}{nh^{d+2}}\right)^{3/2}\right)$$
$$\le C\Big[\left\{\frac{1+h^{2L}logn}{n^3}+\frac{logn}{n^4h^{d+2}}+\frac{logn}{n^3h^2}+(logn)^{3/2}h^{3L}+\left(\frac{logn}{nh^{d+2}}\right)^{3/2}\right\}$$
$$+\frac{h^Llogn}{n^{5/2}}\frac{1}{|t|}+\left\{\frac{h^{2L}(logn)^2}{n^2}+\frac{h^{L-1}(logn)^2}{n^3}+\frac{(logn)^2}{n^3h^2}\right\}\frac{1}{t^2}$$
$$+\frac{(logn)^2}{n^{5/2}h}\frac{1}{|t|^3}+\frac{h^{L-1}(logn)^3}{n^2}\frac{1}{t^4}+\frac{(logn)^4}{n^2h^2}\frac{1}{t^6}\Big].$$

Therefore, dropping the range of integration $p_1\le|t|\le\epsilon n^{1/2}$ on the right hand side,

$$\int_{p_1\le|t|\le\epsilon n^{1/2}}\left|\frac{\chi(t)}{t}\right|dt$$
$$\le C\Big[\left\{\frac{1+h^{2L}logn}{n^3}+\frac{logn}{n^4h^{d+2}}+\frac{logn}{n^3h}+(logn)^{3/2}h^{3L}+\left(\frac{logn}{nh^{d+2}}\right)^{3/2}\right\}\int\frac{dt}{|t|}$$
$$+\frac{h^Llogn}{n^{5/2}}\int\frac{dt}{t^2}+\left\{\frac{h^{2L}(logn)^2}{n^2}+\frac{h^{L-1}(logn)^2}{n^3}+\frac{(logn)^2}{n^3h^2}\right\}\int\frac{dt}{|t|^3}$$
$$+\frac{(logn)^2}{n^{5/2}h}\int\frac{dt}{t^4}+\frac{h^{L-1}(logn)^3}{n^2}\int\frac{dt}{|t|^5}+\frac{(logn)^4}{n^2h^2}\int\frac{dt}{|t|^7}\Big]$$

$$= o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^{L}) \tag{2.24}$$

by assumption (ix).

(ii) For $\epsilon n^{1/2}\le|t|\le n^{1/2}\log n$, there exists a constant $\eta_1\epsilon(0,1)$ such that $|\gamma_\nu(t)|\le 1-\eta_1$ by (x). Choose m such that $m=[-3\log n/\log(1-\eta_1)]$ since $1\le m\le n-1$ for sufficiently large n. Substituting in (2.20) and applying (A.7) and Lemma 7 bounds $|\chi(t)|$ by

$$C(1-\eta_1)^{-\frac{3\log n}{\log(1-\eta_1)}}\left[1+\left\{\frac{h^L}{n^{1/2}}|t|+\left(\frac{h^{2L}}{n}+\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2\right\}\left\{-\frac{3\log n}{\log(1-\eta_1)}\right\}\right.$$
$$+\left\{\frac{1}{n^{3/2}h}|t|+\left(\frac{h^{2L}}{n}+\frac{h^{L-1}}{n^2}+\frac{1}{n^2h^2}\right)t^2\right\}\left\{-\frac{3\log n}{\log(1-\eta_1)}\right\}^2$$
$$\left.+\frac{h^{L-1}}{n^2}t^2\left\{-\frac{3\log n}{\log(1-\eta_1)}\right\}^3+\frac{1}{n^3h^2}t^2\left\{-\frac{3\log n}{\log(1-\eta_1)}\right\}^4\right]$$
$$+O\left(|t|^3\left\{\left(-\frac{\log n}{n\log(1-\eta_1)}\right)^{3/2}h^{3L}+\left(-\frac{\log n}{n^2h^{d+2}\log(1-\eta_1)}\right)^{3/2}\right\}\right),$$

so that

$$\int_{\epsilon n^{1/2}\le|t|\le n^{1/2}\log n}\left|\frac{\chi(t)}{t}\right|dt$$
$$=O\left(\frac{\log(n^{1/2}\log n)}{n^3}\right.$$
$$+\frac{\log n}{n^3}\left\{\frac{h^L}{n^{1/2}}(n^{1/2}\log n)+\left(\frac{h^{2L}}{n}+\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)n(\log n)^2\right\}$$
$$+\frac{(\log n)^2}{n^3}\left\{\frac{1}{n^{3/2}h}(n^{1/2}\log n)+\left(\frac{h^{2L}}{n}+\frac{h^{L-1}}{n^2}+\frac{1}{n^2h^2}\right)n(\log n)^2\right\}$$
$$+\frac{(\log n)^3}{n^3}\frac{h^{L-1}}{n^2}n(\log n)^2+\frac{(\log n)^4}{n^3}\frac{1}{n^3h^2}n(\log n)^2$$
$$\left.+n^{3/2}(\log n)^3\left\{\left(\frac{\log n}{n}\right)^{3/2}h^{3L}+\left(\frac{\log n}{n^2h^{d+2}}\right)^{3/2}\right\}\right)$$
$$=o(n^{-1/2}+n^{1/2}h^{L}+n^{-1}h^{-d-2}) \tag{2.25}$$

by assumption (ix). Thus by (2.24) and (2.25),

$$\text{(II-1)} = o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^{L}),$$

which completes the proof. □

2.3 An empirical Edgeworth expansion

We derive an empirical Edgeworth expansion by replacing population κ_i in (2.1) by strongly consistent estimates

$$\tilde{\kappa}_1 = \frac{2(-1)^L \hat{\sigma}^{-1}}{L!} \sum_{\substack{0 \le l_1, \cdots, l_d \le L \\ l_1 + \cdots + l_d = L}} \cdots \sum \left\{ \int \prod_{i=1}^{d} u_i^{l_i} K(u)\, du \right\} \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta^{(l_1, \cdots, l_d)} v^{\tau} \tilde{f}'(X_i) \right\} Y_i \ ,$$

$$\tilde{\kappa}_2 = \hat{\sigma}^{-2} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} h^{d+2} \bar{W}_{ij}^2 \ , \quad \tilde{\kappa}_3 = \frac{\hat{\sigma}^{-3}}{n} \sum_{i=1}^{n} \bar{V}_i^3 \ ,$$

$$\tilde{\kappa}_4 = \frac{\hat{\sigma}^{-3}}{n(n-1)} \sum_{i=1}^{n} \sum_{j \ne i}^{n} v^{\tau} U_{ij} \bar{V}_i \bar{V}_j \ ,$$

$$\hat{\sigma}^2 = v^{\tau} \hat{\Sigma} v \ , \quad \hat{\Sigma} = \frac{4}{(n-1)(n-2)^2} \sum_{i=1}^{n} \left\{ \sum_{j \ne i}^{n} (U_{ij} - U) \right\} \left\{ \sum_{k \ne i}^{n} (U_{ik} - U)^{\tau} \right\}$$

where for positive b and a function $H: R^d \to R$

$$\tilde{f}(X_i) = \frac{1}{(n-1) b^d} \sum_{j \ne i}^{n} H\left(\frac{X_i - X_j}{b}\right) \ ,$$

$$\bar{U}_i = \frac{1}{n-1} \sum_{j \ne i}^{n} U_{ij} \ , \tag{2.26}$$

$$\bar{V}_i = v^{\tau}(\bar{U}_i - U) \ , \quad \bar{W}_{ij} = v^{\tau}(U_{ij} - \bar{U}_i - \bar{U}_j + U) \ .$$

$\hat{\Sigma}$ is a jackknife estimate of Σ .

Define

$$\tilde{F}(z) = \Phi(z) - \phi(z) \left\{ n^{1/2} h^L \tilde{\kappa}_1 + \frac{\tilde{\kappa}_2}{nh^{d+2}} z + \frac{4(\tilde{\kappa}_3 + 3\tilde{\kappa}_4)}{3n^{1/2}} (z^2 - 1) \right\} \ . \tag{2.27}$$

THEOREM 2 : *Under* (i)', (ii), (iii), (iv)', (v)', (vi)-(viii), (ix)' *and* (x)-(xii),

$$\sup_{v:\, v^{\tau} v = 1} \sup_{z \in R} |F(z) - \tilde{F}(z)| = o(n^{-1/2} + n^{-1} h^{-d-2} + n^{1/2} h^L) \ \textit{completely.}$$

By the statement $X_n = a + o(g_n)$ completely, for a constant a and a positive decreasing sequence g_n , we mean that for all $\epsilon > 0$, $\sum_{n=1}^{\infty} P(|X_n - a| > \epsilon g_n) < \infty$, and if this holds for $g_n \equiv 1$ we say that $X_n \to a$ completely (see e.g. Serfling, 1980, p.10, Stout, 1974, p.221). We shall frequently use the fact that, by Markov's inequality, $X_n \to a$ completely if $E|X_n - a|^r = O(n^{-1}(\log n)^{-1-\zeta})$ for some $r > 0$, $\zeta > 0$. By the Borel-Cantelli

Lemma, complete convergence implies almost sure convergence. We shall omit the qualification "completely" when referring to this convergence.

Comparing assumptions above with those in Theorem 1, we strengthen (i)', (iv)', (v)' and (ix)' here. We will prove this theorem by showing $\tilde{\kappa}_i \to \kappa_i$, $i = 1, 2, 3, 4$. (i)' is necessary for all four of them since $\hat{\sigma}^2 \to \sigma^2$ entails it. (iv)' and (v)' are used in the proof of $\tilde{\kappa}_1 \to \kappa_1$, while $\tilde{\kappa}_3 \to \kappa_3$ requires (ix)'. Before proving Theorem 2 it is useful to establish the following

PROPOSITION 1: *Under* (i), (iv), (v), (vi), (vii), (viii), *and* (xi), $U \to \bar{\mu}$.

PROOF: Because $EU \to \bar{\mu}$ by (A.1) and ν is arbitrary, it suffices to consider $\sigma^{-1}\nu^{\tau}(U - EU) = \frac{2}{n}\sum_{i=1}^{n} V_i + \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_{ij} = a_1 + a_2$. Since the V_i are independent with zero mean,

$$E|a_1|^3 \leq Cn^{-5/2}\sum_{i=1}^{n} E|V_i|^3 \leq Cn^{-3/2}(E|Y|^3 + 1) = O(n^{-3/2}) ,$$

so $a_1 \to 0$. From (A.6),

$$E(n^2 a_2)^2 \leq C\sum_{i=1}^{n-1} E(\sum_{j=i+1}^{n} W_{ij})^2 \leq Cn^2 h^{-d-2} = O(n^3(\log n)^{-2})$$

from (ix), so $a_2 \to 0$. □

PROOF OF THEOREM 2:

In view of Theorem 1, it suffices to show $\tilde{\kappa}_i \to \kappa_i$, $i = 1, 2, 3, 4$. We first prove $\tilde{\kappa}_3 \to \kappa_3$. In view of Lemma 13 this is implied if

$$\sigma^3 \hat{\sigma}^{-3} \to 1 \tag{2.28}$$

and

$$\frac{1}{n\sigma^3}\sum_{i=1}^{n} \bar{V}_i^3 \to E(v_1^3) . \tag{2.29}$$

By Slutsky's lemma, (2.28) is true if

$$\hat{\sigma}^2 \to \sigma^2 . \tag{2.30}$$

To prove (2.29), put $\lambda=\sigma^{-1}\nu$ and write

$$\frac{1}{n\sigma^3}\sum_{i=1}^{n}\bar{V}_i^3 = \frac{1}{n}\sum_{i=1}^{n}\{\lambda^T(\bar{U}_i-U_i)\}^3 + \frac{1}{n}\sum_{i=1}^{n}V_i^3 - \{\lambda^T(U-EU)\}^3$$
$$+\frac{3}{n}\sum_{i=1}^{n}\{\lambda^T(\bar{U}_i-U_i)\}^2V_i - \frac{3\lambda^T(U-EU)}{n}\sum_{i=1}^{n}\{\lambda^T(\bar{U}_i-U_i)\}^2$$
$$+\frac{3}{n}\sum_{i=1}^{n}\lambda^T(\bar{U}-U_i)V_i^2 + \frac{3\{\lambda^T(U-EU)\}^2}{n}\sum_{i=1}^{n}\lambda^T(\bar{U}_i-U_i)$$
$$-\frac{6\lambda^T(U-EU)}{n}\sum_{i=1}^{n}\lambda^T(\bar{U}_i-U_i)V_i$$
$$-\frac{3\lambda^T(U-EU)}{n}\sum_{i=1}^{n}V_i^2 + \frac{3\{\lambda^T(U-EU)\}^2}{n}\sum_{i=1}^{n}V_i \ . \tag{2.31}$$

We start with the second term on the right of (2.31), writing

$$\frac{1}{n}\sum_{i=1}^{n}V_i^3 = \frac{1}{n}\sum_{i=1}^{n}\nu_i^3 + \frac{1}{n}\sum_{i=1}^{n}\{V_i^3-\nu_i^3-E(V_1^3)+E(\nu_1^3)\} + E(V_1^3-\nu_1^3) \ . \tag{2.32}$$

By (iii) and $E(\nu_1^6)<\infty$ due to (i)' and Lemma 1-(d), $\frac{1}{n}\sum_{i=1}^{n}\nu_i^3 \to E(\nu_1^3)$ from Theorem 4.13 of Stout (1974). The third term on the right of (2.32) is $O(h^L)$ since its modulus is bounded by

$$CE(|V_1-\nu_1||V_1^2+\nu_1^2|) \le Ch^LE[(|Y_1|+1)(|Y_1|+1)^2] \ ,$$

and this is $O(h^L)$ due to (i)', Lemma 1-(d), and (B.1). The second term on the right of (2.32) converges to zero because it has mean zero and variance bounded by

$$\frac{C}{n^2}\sum_{i=1}^{n}E\{V_i^3-\nu_i^3-E(V_1^3)+E(\nu_1^3)\}^2$$
$$\le \frac{C}{n}\{E|V_1^3-\nu_1^3|^2+|E(V_1^3)-E(\nu_1^3)|^2\} \ , \tag{2.33}$$

and because

$$E|V_1^3-\nu_1^3|^2 \le CE|V_1-\nu_1|^2|V_1^4+\nu_1^4| \le Ch^{2L}E(|Y_1|^2+1)(|Y_1|^4+1) \le Ch^{2L} \ ,$$

by (i)', Lemma 1-(d), and (B.1), (2.33) is $O(n^{-1}h^{2L})=O(n^{-2})$ by (ix)'. Thus $n^{-1}\sum_{i=1}^{n}V_i^3 \to E(\nu_1^3)$. Substituting (2.26) into (2.31) and noting $\lambda^T(U_{ij}-U_i)=W_{ij}+V_j$, we have the following typical quantities for the first term on the right of (2.31) :

$$\frac{1}{n(n-1)^3}\sum\sum_{i<j<k<l}\sum\sum(W_{ij}+V_j)(W_{ik}+V_k)(W_{il}+V_l) \ , \tag{2.34}$$

$$\frac{1}{n(n-1)^3}\sum\sum_{i<j<k}\sum(W_{ij}+V_j)^2(W_{jk}+V_k) \ , \tag{2.35}$$

$$\frac{1}{n(n-1)^3}\sum_{i<j}\sum(W_{ij}+V_j)^3 \ . \tag{2.36}$$

(2.34) has mean zero and variance which, by Lemma 1-(b),(c), is bounded by

$$Cn^{-5}\left[nE\{(W_{12}+V_2)^2(W_{13}+V_3)^2(W_{14}+V_4)^2\}\right.$$
$$\left.+n^2E\{(W_{13}+V_3)(W_{14}+V_4)(W_{15}+V_5)(W_{23}+V_3)(W_{24}+V_4)(W_{25}+V_5)\}\right].$$

The first expectation is bounded by

$$CE\left[E(W_{12}^2|1)E(W_{13}^2|1)E(W_{14}^2|1)\right]=O(h^{-3d-6})=O(n^3/(\log n)^3) \quad (2.37)$$

by Lemma 4, (i)', and (ix)' because the terms involving V_i are of smaller orders, e.g.

$$E(W_{12}^2W_{13}^2W_{14}V_4)=E\{W_{12}^2W_{13}^2E(W_{14}V_4|1)\}$$
$$\le CE\{(|Y_1|+1)E(W_{12}^2|1)E(W_{13}^2|1)\}\le Ch^{-2d-4}E(|Y_1|^5+1)=O(h^{-2d-4})$$

where we use

$$|E(W_{14}V_4|1)|\le|E(\lambda^\tau U_{14}V_4|1)|+|E(V_4^2)|\le C(|Y_1|+1), \quad (2.38)$$

due to $|E(\lambda^\tau U_{14}V_4|1)|\le C(|Y_1|+1)$ (see the proof of (A.4)) and $E(V_4^2)\le C$ by (i)' and Lemma 1-(d). Applying Lemma 6 repeatedly and (ix)', the second expectation is

$$E\left[\left\{E(W_{13}W_{23}|1,2)\right\}^3\right]=O(h^{-2d-6})=O((n/\log n)^2),$$

since terms involving V_i are of smaller orders again as in the above. Thus, (2.34) $\to 0$. (2.35) has mean zero and variance

$$\frac{1}{n^2(n-1)^6}\sum_{k=3}^{n}E\left[\sum_{j=2}^{k-1}\sum_{i=1}^{j-1}(W_{ij}+V_j)^2(W_{ik}+V_k)\right]^2$$
$$=\frac{1}{n^2(n-1)^6}\sum_{k=3}^{n}E(\sum_{i=1}^{k-2}P_{ik})^2, \quad (2.39)$$

writing $P_{ik}=(W_{ik}+V_k)\sum_{j=i+1}^{k-1}(W_{ij}+V_j)^2$. Now since terms involving V_i are of smaller orders,

$$E(P_{ik}^2)\le C\left[kE(W_{12}^2W_{13}^4)+k^2E(W_{12}^2W_{13}^2W_{14}^2)\right]\le C(kh^{-4d-6}+k^2h^{-3d-6})$$

because of (2.37) and, by Lemma 4 and (i)',

$$E(W_{12}^2W_{13}^4)\le E\left[W_{13}^4E(W_{12}^2|1)\right]\le Ch^{-d-2}E[W_{13}^4(|Y_1|^2+1)]$$
$$\le Ch^{-d-2}E\left[(|Y_1|^2+1)E(W_{13}^4|1)\right]\le Ch^{-d-2}h^{-3d-4}E(|Y_1|^2+1)(|Y_1|^4+1)$$
$$\le Ch^{-4d-6},$$

while for $l \neq m$,

$$|E(P_{mk}P_{lk})| \leq C\,|\,kE(W_{14}W_{24}W_{13}W_{23}) + k^2E(W_{15}W_{25}W_{13}W_{24})\,| \leq Ck\,h^{-d-4} ,$$

since the first expectation is $O(h^{-d-4})$ by Lemma 6 and the second is zero by Lemma 1-(c). Thus (2.39) is

$$O(\frac{1}{n^8}\sum_{k=3}^{n}(k^2h^{-4d-6}+k^3h^{-3d-6})) = O(1/n(logn)^2)$$

using (ix)', so that (2.35) $\to 0$. Next (2.36) is

$$\frac{1}{n(n-1)^3}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}[(W_{ij}+V_j)^3 - E\{(W_{ij}+V_j)^3|j\}]$$
$$+\frac{1}{n(n-1)^3}\sum_{j=1}^{n}(j-1)\left[E\{(W_{1j}+V_j)^3|j\} - E(W_{12}+V_1)^3\right]$$
$$+\frac{1}{2(n-1)^2}E(W_{12}+V_1)^3 . \tag{2.40}$$

The last expression is $O(n^{-2}h^{-2d-3})$ by (A.5) and Lemma 1-(d) with (i)' . The second term of (2.40) has mean zero and variance

$$\frac{1}{n^2(n-1)^6}\sum_{i=1}^{n-1}E\Big|\sum_{j=i+1}^{n}[(W_{ij}+V_j)^3 - E\{(W_{ij}+V_j)^3|j\}]\Big|^2$$
$$\leq \frac{C}{n^8}\sum_{i=1}^{n-1}\{nE(W_{12}^6) + n^2E(W_{12}^3W_{13}^3)\} \leq C(n^{-6}h^{-5d-6}+n^{-5}h^{-4d-6}) ,$$

by Lemma 4 and (A.5). The first term has mean zero and variance

$$\frac{1}{n^2(n-1)^6}\sum_{j=2}^{n}(j-1)^2E\left[E\{(W_{1j}+V_j)^3|j\} - E(W_{12}+V_1)^3\right]^2$$
$$\leq \frac{C}{n^5}E\{E(|W_{12}|^3|2)\}^2 \leq Cn^{-5}h^{-4d-6} ,$$

by Lemma 4 and (i)'. Thus by (ix)', (2.36) converges to zero. The other terms in (2.31) can be handled like (2.34)-(2.36) and are shown to converge to zero using the Proposition, so that (2.29) is established. To prove (2.30), write $\hat{\sigma}^2 = \sigma^2\frac{4(n-1)}{(n-2)^2\sigma^2}\sum_{i=1}^{n}\bar{V}_i^2$. Similarly to (2.29), we have

$$(n\sigma^2)^{-1}\sum_{i=1}^{n}\bar{V}_i^2 \to E(\nu_1^2) . \tag{2.41}$$

Next we prove $\bar{\kappa}_4 \to \kappa_4$. By (2.28) and Lemma 14, it suffices to show

$$\frac{\sigma^{-3}}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\nu^{\tau}U_{ij}\,\bar{V}_i\bar{V}_j - E(W_{12}\nu_1\nu_2) \to 0 .$$

Since $\sigma^{-1}\bar{V}_i = V_i + \lambda^{\tau}(\bar{U}_i - U_i) - \lambda^{\tau}(U-EU)$,

$$\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sigma^{-3}\nu^{\tau}U_{ij}\,\bar{V}_i\bar{V}_j$$

$$= \binom{n}{2}^{-1} \Big\{ \sum_{i<j}\sum \lambda^T U_{ij} V_i V_j + \sum_{i<j}\sum \lambda^T U_{ij} V_i \lambda^T(\bar{U}_j - U_j)$$

$$- \lambda^T(U-EU) \sum_{i<j}\sum \lambda^T U_{ij} V_i + \sum_{i<j}\sum \lambda^T U_{ij} \lambda^T(\bar{U}_i - U_i) V_j$$

$$+ \sum_{i<j}\sum \lambda^T U_{ij} \lambda^T(\bar{U}_i - U_i) \lambda^T(\bar{U}_j - U_j)$$

$$- \lambda^T(U-EU) \sum_{i<j}\sum \lambda^T U_{ij} \lambda^T(\bar{U}_i - U_i)$$

$$- \lambda^T(U-EU) \sum_{i<j}\sum \lambda^T U_{ij} V_j - \lambda^T(U-EU) \sum_{i<j}\sum \lambda^T U_{ij} \lambda^T(\bar{U}_j - U_j)$$

$$+ \{\lambda^T(U-EU)\}^2 \sum_{i<j}\sum \lambda^T U_{ij} \Big\} . \tag{2.42}$$

Writing

$$Q_i = E(\lambda^T U_{ij} V_i V_j | i) - E(\lambda^T U_{12} V_1 V_2) ,$$

$$Q_{ij} = \lambda^T U_{ij} V_j V_j - Q_i - Q_j + E(\lambda^T U_{12} V_1 V_2) ,$$

the first term on the right of (2.42) is

$$E(\lambda^T U_{12} V_1 V_2) + \frac{2}{n}\sum_{i=1}^{n} Q_i + \binom{n}{2}^{-1} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Q_{ij} . \tag{2.43}$$

Since $\{Q_i\}$ is a stationary martingale difference sequence,

$$E|\frac{1}{n}\sum_{i=1}^{n} Q_i|^3 \le \frac{C}{n^{3/2}} E|Q_1|^3 \le \frac{C}{n^{3/2}} E|E(\lambda^T U_{ij} V_i V_j | i)|^3 \le \frac{C}{n^{3/2}} ,$$

the last inequality using (2.38), Lemma 1-(d) and (i)'. Since $EQ_{ij} = 0$ and $EQ_{ij} Q_{kl} = 0$ unless $i = k$ and $j = l$, the last term in (2.43) has mean zero and variance

$$\frac{4}{n^2(n-1)^2} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E(Q_{ij}^2) \le \frac{C}{n^2} E(Q_{12}^2)$$
$$\le \frac{C}{n^2} E|\lambda^T U_{12} V_1 V_2|^2 \le C n^{-2} h^{-d-2} ,$$

where the last inequality is due, in view of the proof of Lemma 5 of Robinson (1995a), to

$$E(\lambda^T U_{12} V_1 V_2)^2 = E\left[V_1^2 E\{(\lambda^T U_{12})^2 V_2^2 | 1\}\right] \le E\left[V_1^2 C(|Y_1|^2 + 1) h^{-d-2}\right]$$
$$\le C h^{-d-2} E[(|Y_1|^2 + 1)^2] ,$$

by Lemma 1-(d). Since $E(\lambda^T U_{12} V_1 V_2) = E(W_{12} V_1 V_2) = E(W_{12} v_1 v_2) + o(1)$ by (iii), Lemmas 1-(b) and 14, it follows that the first term of (2.42) is $E(W_{12} v_1 v_2) + o(1)$. The proofs that the other terms in (2.42) converge to zero are omitted because of their similarity to the proofs for (2.34)-(2.36) and because of their straightforward use of the Proposition. Thus $\tilde{\kappa}_3 \to \kappa_3$.

Since $\sigma^{-1}\tilde{W}_{ij} = W_{ij} - \lambda^\tau(\bar{U}_i - U_i) - \lambda^\tau(\bar{U}_j - U_j) - \lambda^\tau(U - EU)$,

$$
\begin{aligned}
\tilde{\kappa}_2 = {} & \frac{\sigma^2}{\hat{\sigma}^2} h^{d+2}\{\lambda^\tau(U-EU)\}^2 + \frac{\sigma^2}{\hat{\sigma}^2}\binom{n}{2}^{-1}\Big\{\sum\sum_{i<j} h^{d+2} W_{ij}^2 \\
& + \sum\sum_{i<j} h^{d+2}\{\lambda^\tau(\bar{U}_i - U_i)\}^2 + \sum\sum_{i<j} h^{d+2}\{\lambda^\tau(\bar{U}_j - U_j)\}^2 \\
& - 2\sum\sum_{i<j} h^{d+2} W_{ij}\, \lambda^\tau(\bar{U}_i - U_i) - 2\sum\sum_{i<j} h^{d+2} W_{ij}\, \lambda^\tau(\bar{U}_j - U_j) \\
& - 2\lambda^\tau(U-EU)\sum\sum_{i<j} h^{d+2} W_{ij} \\
& + 2\sum\sum_{i<j} h^{d+2}\lambda^\tau(\bar{U}_i - U_i)\,\lambda^\tau(\bar{U}_j - U_j) \\
& + 2\lambda^\tau(U-EU)\sum\sum_{i<j} h^{d+2}\lambda^\tau(\bar{U}_i - U_i) \\
& + 2\lambda^\tau(U-EU)\sum\sum_{i<j} h^{d+2}\lambda^\tau(\bar{U}_j - U_j)\Big\}. \qquad (2.44)
\end{aligned}
$$

Because of (2.30), Lemma 12 and because $E(W_{12}^4) = O(h^{-3d-4})$ due to (A.5) under (i)', it follows much as in the proof for (2.42) that the first term in the wave brackets on the right of (2.44) converges to κ_2 . We omit the proof that the remaining terms converge to zero because it is similar to the proofs for (2.34)-(2.36) and straightforwardly uses the Proposition.

We finally show $\tilde{\kappa}_1 \to \kappa_1$. In view of (2.2) and (2.30) it suffices to show that

$\frac{1}{n}\sum_{i=1}^{n}\{\Delta^{(l_1,\cdots,l_d)}\nu^\tau \tilde{f}'(X_i)\}Y_i$ converges to $E[(\Delta^{(l_1,\cdots,l_d)}\nu^\tau f')\,g]$. We have

$$
\frac{1}{n}\sum_{i=1}^{n}\Delta^{(l_1,\cdots,l_d)}\nu^\tau \tilde{f}'(X_i)Y_i = \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} H_{ij} \qquad (2.45)
$$

where

$$
H_{ij} = b^{-d-1-L}\Delta^{(l_1,\cdots,l_d)}\nu^\tau H'\Big(\frac{X_i - X_j}{b}\Big)(Y_i - Y_j) \ .
$$

Since we may choose an even L, the kernel order, without loss of generality, $\Delta^{(l_1,\cdots,l_d)}\nu^\tau H'(u)$ is an odd function by (xi) and thus (2.45) has a U-statistic form. As is standard in U-statistic theory, define $H_i = E(H_{ij}\,|i)$, then (iv)', (vii), (xi), and integration by parts give

$$
\begin{aligned}
H_i &= E[E(H_{ij} \mid X_i, Y_i, X_j) \mid X_i, Y_i] \\
&= b^{-d-1-L} E[\Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(\frac{X_i - X_j}{b}) \{Y_i - g(X_j)\} \mid X_i, Y_i] \\
&= b^{-1-L} \int \{\Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(u)\} \{Y_i - g(X_i - bu)\} f(X_i - bu)\, du \\
&= Y_i \int H(u) \{\Delta^{(l_1, \cdots, l_d)} \nu^\tau f'(X_i - bu)\} du \\
&\qquad - \int H(u) \{\Delta^{(l_1, \cdots, l_d)} \nu^\tau e'(X_i - bu)\} du
\end{aligned}
$$

and

$$
\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} H_{ij} - EH_{12} = a_3 + a_4
$$

where

$$
a_3 = \frac{2}{n} \sum_{i=1}^{n} (H_i - EH_{12}) \quad , \quad a_4 = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (H_{ij} - H_i - H_j + EH_{12}) \ .
$$

Noting that the $H_i - EH_{12}$ are iid with zero mean and bounded variance due to (i)', (iv)', (v)' and (xi), and

$$
\begin{aligned}
E(H_{12}^2) &\le b^{-2d-2-2L} E\left(E[\{\Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(\frac{X_1 - X_2}{b})\}^2 (Y_1^2 + Y_2^2) \mid X_1, Y_1] \right) \\
&= b^{-2d-2-2L} E \int \{\Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(\frac{X_1 - x}{b})\}^2 \{Y_1^2 + q(x)\} f(x)\, dx \\
&= b^{-d-2-2L} E \int \{\Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(u)\}^2 \{Y_1^2 + q(X_1 - bu)\} f(X_1 - bu)\, du \\
&\le C b^{-d-2-2L} E(Y_1^2 + 1) = O(b^{-d-2-2L}) \ ,
\end{aligned}
$$

due to (i)', (iv)', (vii) and (xi), $a_3 \to 0$ and $a_4 \to 0$ under (i)', (iii), (iv)', (v)', (xi), and (xii) similarly to the proof of the proposition. By (iv)', (vii), (xi), and integration by parts

$$
\begin{aligned}
E(H_{12}) &= E\{b^{-d-1-L} \Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(\frac{X_1 - X_2}{b}) Y_1\} \\
&= b^{-d-1-L} E \int \Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(\frac{X_1 - x}{b}) Y_1 f(x)\, dx \\
&= b^{-1-L} E\left[Y \int \{\Delta^{(l_1, \cdots, l_d)} \nu^\tau H'(u)\} f(X - bu)\, du \right] \\
&= E\left[Y \int H(u) \{\Delta^{(l_1, \cdots, l_d)} \nu^\tau f'(X - bu)\} du \right] \\
&\to E[Y \Delta^{(l_1, \cdots, l_d)} \nu^\tau f'] = E[(\Delta^{(l_1, \cdots, l_d)} \nu^\tau f') g]
\end{aligned}
$$

so that $\tilde{\kappa}_1 \to \kappa_1$. □

Chapter 3

Edgeworth Expansions for Averaged Derivatives And Bandwidth Selections - Studentized Case [1]

Results in the previous chapter are theoretically interesting, however they do not suit for practical use because the statistics there are not studentized (see 1.3.1). We provide valid theoretical and empirical Edgeworth expansions for studentized U in Section 3.1 and 3.2 respectively. The former expansion is still infeasible due to the unknown quantities it involves, but the latter is feasible and so for practical use. It can be used for confidence interval estimation or hypthesis tests as described in (1.87) for the sum of i.i.d. random variables. Section 3.3 discusses the following points. Firstly, we show that a number of situations results depending on the dimension of the explanatory variables, the kernel order and the rate of decay of the bandwidth, and especially that an Edgeworth expansion of parametric order is possible. Secondly, we propose bandwidth selections which minimize the normal approximation error. Thirdly, we discuss differences between our results and those of Linton (1996a).

3.1 A theoretical Edgeworth expansion

Our studentized statistic is $\hat{Z} = n^{1/2}\hat{\sigma}^{-1}\nu^{\tau}(U-\bar{\mu})$ and we are concerned with approximating

$$\hat{F}(z) = P(\hat{Z} \leq z)$$

by

[1] This Chapter has been revised and written up as a joint paper with my supervisor Professor Peter M. Robinson as Nishiyama and Robinson (1998). A part of Section 3.3 is included in Nishiyama and Robinson (2000).

$$F^{+}(z) = \Phi(z) - \phi(z)\left[n^{1/2}h^{L}\kappa_1 - \frac{\kappa_2}{nh^{d+2}}z - \frac{4}{3n^{1/2}}\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\}\right] . \tag{3.1}$$

THEOREM 3 : Under (i)', (ii)-(x), as $n \to \infty$

$$\sup_{v:\, v^{\tau}v=1} \sup_{z \in R} |\hat{F}(z) - F^{+}(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) .$$

The assumptions are the same as those of Theorem 1 except (i)' strengthens the third moment assumption. The correction term in $F^{+}(z)$ contains terms of the same orders as those in Theorem 1. Moreover, the "bias" component, of order $n^{1/2}h^{L}$ is identical to that of Theorem 1, while the "variance" component, of order $n^{-1}h^{-d-2}$ is the negative of that in Theorem 1, though the remaining component, of order $n^{-1/2}$, differs from that of Theorem 1. The first two correction terms are related to nonparametric density estimate as the corresponding ones in (2.1) while the last term is also "parametric" as the last term of (2.1). We find a similarity between the last term of (3.1) and the correction term of studentized ordinary U-statistics (1.93) like the similarity of the last correction term in (2.1) and the first correction term of (1.91). We discuss this "parametric" case in Section 3.3.

We use the same bandwidth h to estimate σ^2 here, but it is not necessary. If we use a different bandwidth choice, we will have the same correction terms as in Theorem 1 and other correction terms related to the variance estimation.

We handle the studentization like Callaert and Veraverbeke (1981), namely we expand $\hat{\sigma}^{-1}$ around $\hat{\sigma}^2 = \sigma^2$. Then, we have cross terms of the expansion and the decomposition of U. Based on this, we prove this theorem in a similar manner to Theorem 1 by investigating the characteristic function.

PROOF OF THEOREM 3

As is standard in U-statistic theory, we write

$$n^{1/2}\sigma^{-1}v^{\tau}(U - \bar{\mu}) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n} V_i + n^{1/2}\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} W_{ij} + n^{1/2}\sigma^{-1}v^{\tau}(EU - \bar{\mu})$$

$$= \bar{V} \quad + \quad \bar{W} \quad + \quad \Delta . \tag{3.2}$$

Writing $S=4Var(U_i)$, $s^2=\sigma^{-2}\nu^\tau S\nu$, Taylor's theorem gives

$$\sigma\hat{\sigma}^{-1} = s^{-1} - \frac{s^{-3}}{2}(\sigma^{-2}\hat{\sigma}^2 - s^2) + \frac{3}{8}\{s^2+\theta(\sigma^{-2}\hat{\sigma}^2-s^2)\}^{-5/2}(\sigma^{-2}\hat{\sigma}^2-s^2)^2$$
$$= s^{-1} \quad + \quad \tilde{R} \quad + \quad \tilde{\tilde{R}} \tag{3.3}$$

for some $\theta \in [0, 1]$. Similarly to Callaert and Veraverbeke (1981), we expand $\tilde{R}$ as follows. With $\tilde{V}_i = E(V_j W_{ij} | i)$, $\tilde{W}_{jk} = E(W_{ij} W_{ik} | j, k)$, we have

$$\tilde{R} = T+Q+R \ , \quad T=T_1+T_2+T_3 \ , \quad Q=Q_1+Q_2 \ , \quad R=R_1+R_2+R_3+R_4+R_5$$

where

$$T_1 = \frac{4\delta n}{(n-2)^2}E(W_{12}^2)\ , \ T_2 = \frac{\delta}{n}\sum_{i=1}^{n}\left\{(4V_i^2-s^2)+8\tilde{V}_i\right\},\ T_3 = 4\delta\binom{n-1}{2}^{-1}\sum_{i<j}^{n}\tilde{W}_{ij}\ ,$$

$$Q_1 = 4\delta\binom{n}{2}^{-1}\sum_{i<j}\left\{(V_i+V_j)W_{ij} - \tilde{V}_i - \tilde{V}_j\right\},$$

$$Q_2 = -\frac{8\delta}{n}\binom{n-1}{2}^{-1}\sum_{i=1}^{n}\sum_{k<m}^{n\ (i)} V_i W_{km}\ ,$$

$$R_1 = -4\delta\binom{n}{2}^{-1}\sum_{i<j}V_iV_j\ , \ R_2 = \frac{4\delta}{n-2}\binom{n-1}{2}^{-1}\sum_{i=1}^{n}\sum_{k<m}^{n\ (i)}(W_{ik}W_{im} - \tilde{W}_{km})\ ,$$

$$R_3 = \frac{4\delta n}{(n-2)^2}\binom{n}{2}^{-1}\sum_{i<j}\{W_{ij}^2 - \tilde{W}_{ii} - \tilde{W}_{jj} + E(W_{12}^2)\}\ ,$$

$$R_4 = \frac{8\delta}{(n-2)^2}\sum_{i=1}^{n}\{\tilde{W}_{ii} - E(W_{12}^2)\}\ , \ R_5 = -\frac{4\delta n(n-1)}{(n-2)^2}\left\{\binom{n-1}{2}^{-1}\sum_{i<j}W_{ij}\right\}^2\ ,$$

where $\delta = -s^{-3}/2$ and $\sum_{k<m}^{n\ (i)}$ denotes summation with respect to k and m for $1 \le k < m \le n$ excluding $k=i$ and $m=i$. Lemmas 24-33 show moment bounds of T, Q and R, and $E|T|^r < E|Q|^r < E|R|^r$ asymptotically for such r that the expectations exist. Because

$$\hat{Z} = (s^{-1}+\tilde{R}+\tilde{\tilde{R}})(\bar{V}+\bar{W}+\Delta)\ ,$$

by a standard inequality

$$\sup_z |\hat{F}(z) - F^+(z)| \le \sup_z |P((s^{-1}+T+Q)(\bar{V}+\bar{W}) + s^{-1}\Delta \le z) - F^+(z)|$$
$$+P\left(|(R+\tilde{\tilde{R}})(\bar{V}+\bar{W}+\Delta) + (T+Q)\Delta| \ge a_n\right) + O(a_n) \tag{3.4}$$

for $a_n>0$, where here and subsequently we drop reference to $\sup_{\nu:\ \nu^\tau\nu=1}$. Taking $a_n = \frac{1}{logn}\max(n^{-1/2}, n^{-1}h^{-d-2}, n^{1/2}h^{L})$, we bound the second term on the right of (3.4) by

$$P\left(|(R+\tilde{\tilde{R}})(\bar{V}+\bar{W}+\Delta)| \geq \frac{a_n}{2}\right) + P\left(|(T+Q)\Delta| \geq \frac{n^{1/2}h^{L}}{2logn}\right)$$

$$\leq P\left(|R+\tilde{\tilde{R}}| \geq \frac{a_n}{2\log n}\right) + P(|\bar{V}+\bar{W}+\Delta| \geq \log n) + P\left(|(T+Q)\Delta| \geq \frac{n^{1/2}h^{L}}{2\log n}\right) .$$

(3.5)

The first term in (3.5) is, by elementary inequalities, bounded by

$$P(|R| \geq \frac{a_n}{4\log n}) + P\left(\frac{|\tilde{\tilde{R}}|}{\tilde{R}^2} \geq C_0\right) + P\left(\tilde{R}^2 \geq \frac{a_n}{4C_0\log n}\right) \tag{3.6}$$

for a constant C_0 determined later. The third term of (3.6) is bounded by

$$P\left(T_2^2 \geq \frac{a_n}{12C_0\log n}\right) + P\left(|T_1+T_3|^2 \geq \frac{a_n}{12C_0\log n}\right) + P\left(|Q+R|^2 \geq \frac{a_n}{12C_0\log n}\right)$$

$$= \quad (a) \quad + \quad (b) \quad + \quad (c) .$$

Lemmas 24-33 and Markov's inequality give, for $\zeta > 0$,

$$(a) \leq \frac{E|T_2|^{2(1+\zeta)}}{\left(\frac{a_n}{12C_0\log n}\right)^{1+\zeta}} \leq \frac{Cn^{-(1+\zeta)}(\log n)^{2(1+\zeta)}}{n^{-\frac{1}{2}(1+\zeta)}} = o(n^{-\frac{1}{2}}) ,$$

$$(b) \leq \frac{E|T_1+T_3|^{2(1+\zeta)}}{\left(\frac{a_n}{12C_0\log n}\right)^{1+\zeta}} \leq \frac{C(n^{-1}h^{-d-2})^{2(1+\zeta)}(\log n)^{2(1+\zeta)}}{(n^{-1}h^{-d-2})^{1+\zeta}} = o(n^{-1}h^{-d-2}) ,$$

$$(c) \leq \frac{E|R+Q|^2}{\frac{a_n}{12C_0\log n}} \leq \frac{Cn^{-2}h^{-d-2}(\log n)^2}{n^{-1/2}} = o(n^{-1}h^{-d-2}) ,$$

where $\zeta \geq \frac{2}{7}$ suffices in (b) under (ix), and ζ arbitrarily small suffices in (a).

The first term of (3.6) is, using Markov's inequality, (ix) and Lemmas 29-33, bounded by

$$\frac{16\,E(R^2)(\log n)^2}{a_n^2} \leq C(n^{-1}+n^{-2}h^{-2d-4})(\log n)^4 = o(n^{-1/2}+n^{-1}h^{-d-2}) .$$

Now, in view of (3.3), $\tilde{\tilde{R}} = \frac{3}{2}s(1-2\theta s\tilde{R})^{-5/2}\tilde{R}^2$ so that because $\tilde{\tilde{R}} \geq 0$ and

$0 \leq \theta \leq 1$,

$$P\left(\frac{\tilde{\tilde{R}}}{\tilde{R}^2} \geq C_0\right) = P\left(\frac{3}{2}s(1-2\theta s\tilde{R})^{-5/2} \geq C_0\right)$$

$$\leq P\left(|\tilde{R}| \geq \frac{1}{2s}\{1 - (\frac{3s}{2C_0})^{2/5}\}\right) . \tag{3.7}$$

Taylor's expansion of s^r around $s^2 = 1$, (ii) and (A.2) give for integer r,

$$s^r = 1 + O(\sigma^{-2}\nu^T(S-\Sigma)\nu) = 1 + O(h^L) \tag{3.8}$$

so that we can choose C_0 such that $C_0 > \frac{3}{2}s$ for sufficiently large n. Then by (3.8) and Markov's inequality, (3.7) is bounded by a constant times $E|T+Q+R|^3 = O(n^{-3/2} + n^{-3}h^{-3d-6})$ from Lemmas 24-33, so that the second term of (3.6) is $O(n^{-3/2} + n^{-3}h^{-3d-6})$. Therefore,

$$P\left(|R+\tilde{\tilde{R}}| \geq \frac{a_n}{\log n}\right) = o(n^{-\frac{1}{2}} + n^{-1}h^{-d-2}) . \tag{3.9}$$

Writing $F(z) = P[n^{1/2}\sigma^{-1}\nu^T(U-\bar{\mu}) \leq z]$ as in Theorem 1, and noting (3.2), we have

$$P(|\bar{V}+\bar{W}+\Delta| \geq \log n) = 1 - F(\log n) + F(-\log n) . \tag{3.10}$$

Theorem 1 implies for any z

$$1 - F(z) + F(-z) = 1 - \tilde{F}(z) + \tilde{F}(-z) + o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) . \tag{3.11}$$

Now by (2.1),

$$\begin{aligned} 1 - \tilde{F}(z) + \tilde{F}(-z) &= 1 - \Phi(z) + \Phi(-z) + \phi(z)\frac{2\kappa_2}{nh^{d+2}}z \\ &= 2 - 2\Phi(z) + \phi(z)\frac{2\kappa_2}{nh^{d+2}}z . \end{aligned} \tag{3.12}$$

Substituting (3.12) into (3.11) and putting $z = \log n$, because $1 - \Phi(\log n) = o(n^{-1/2})$ and $\phi(\log n)\log n = o(n^{-1/2})$, we have

$$1 - F(\log n) + F(-\log n) = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) . \tag{3.13}$$

By (3.10) and (3.13),

$$P(|\bar{V}+\bar{W}+\Delta| \geq \log n) = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) . \tag{3.14}$$

Finally, Markov's inequality, (ix), (A.1), and Lemmas 24-28 bound the last term of (3.5) by

$$\frac{\Delta^2 E|T+Q|^2(2\log n)^2}{nh^{2L}} \leq C(n^{-1} + n^{-2}h^{-2d-4})(\log n)^2 = o(n^{-1/2} + n^{-1}h^{-d-2}) . \tag{3.15}$$

Substituting (3.9), (3.14) and (3.15) into (3.5),

$$P\left(|(R+\tilde{\tilde{R}})(\bar{V}+\bar{W}+\Delta) + (T+Q)\Delta| \geq a_n\right) = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L) . \tag{3.16}$$

To deal with the first term on the right of (3.4), write $b_2 = s^{-1}\bar{V}$, $b_3 = s^{-1}\bar{W}$, $\tilde{b}_2 = (T+Q)\,\bar{V}$, $\tilde{b}_3 = (T+Q)\,\bar{W}$, $b_1 = b_2 + b_3$, $\tilde{b}_1 = \tilde{b}_2 + \tilde{b}_3$, $B = b_1 + \tilde{b}_1$, and define

$$\chi^+(t) = \int e^{itz} dF^+(z)$$

$$= e^{-\frac{t^2}{2}}\left[1 + \left\{n^{1/2}h^L\kappa_1 - \frac{4(\kappa_3+2\kappa_4)}{n^{1/2}}\right\}(it) - \frac{\kappa_2}{nh^{d+2}}(it)^2 - \frac{4(2\kappa_3+3\kappa_4)}{3n^{1/2}}(it)^3\right].$$

Esseen's smoothing lemma gives for $N_0 = \log n \min(\tilde{\varepsilon} n^{1/2}, nh^{d+2})$ with

$\tilde{\varepsilon} = (E|2s^{-1}V_1|^3)^{-1}$,

$$\sup_z |P((s^{-1}+T+Q)(\bar{V}+\bar{W}) + s^{-1}\Delta \le z) - F^+(z)|$$

$$\le \int_{-N_0}^{N_0} \left| \frac{Ee^{it(B+s^{-1}\Delta)} - \chi^+(t)}{t} \right| dt + O(N_0^{-1}),$$

which, for $p_2 = \min(\log n, \tilde{\varepsilon} n^{1/2}, nh^{d+2}) \le N_0$, is bounded by

$$\int_{-p_2}^{p_2} \left| \frac{Ee^{it(B+s^{-1}\Delta)} - \chi^+(t)}{t} \right| dt + \int_{p_2 \le |t| \le N_0} \left| \frac{Ee^{it(B+s^{-1}\Delta)}}{t} \right| dt$$

$$+ \int_{|t| \ge p_2} \left| \frac{\chi^+(t)}{t} \right| dt + o(n^{-1/2} + n^{-1}h^{-d-2})$$

$$= \text{(I-2)} + \text{(II-2)} + \text{(III-2)} + o(n^{-1/2} + n^{-1}h^{-d-2}).$$

Here $\tilde{\varepsilon}$ is bounded away from zero due to (3.8) and Lemma 1-(d) for sufficiently large n. (III-2) is $o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^L)$ as (III-1) in Theorem 1.

To estimate (I-2), we proceed by writing $\chi(t)$ as $\chi^+(t)$ plus a remainder. Since $s^{-1}\Delta$ is nonstochastic,

$$E\{e^{it(B+s^{-1}\Delta)}\} = e^{its^{-1}\Delta} E(e^{itB}), \tag{3.17}$$

where (2.6) and (3.8) yield

$$e^{its^{-1}\Delta} = 1 + its^{-1}\Delta + (e^{its^{-1}\Delta} - 1 - its^{-1}\Delta)$$

$$= 1 + its^{-1}\Delta + O(t^2 s^{-2}\Delta^2)$$

$$= 1 + it\Delta + O(t^2\Delta^2 + |t|h^L\Delta) \quad . \tag{3.18}$$

Writing $\bar{b}_2 = \bar{b}_2' + \bar{b}_2''$ where $\bar{b}_2' = T\bar{V}$ and $\bar{b}_2'' = Q\bar{V}$, and applying (2.6) repeatedly, we have

$$\begin{aligned}
E(e^{itB}) &= E(e^{itb_1}) + E(e^{itB} - e^{itb_1}) \\
&= E(e^{itb_1}) + \left\{Ee^{it(b_1+\bar{b}_2+\bar{b}_3)} - Ee^{it(b_1+\bar{b}_2')}\right\} + \left\{Ee^{it(b_1+\bar{b}_2')} - E(e^{itb_1})\right\} \\
&= E(e^{itb_1}) + O(|t|E|\bar{b}_2''+\bar{b}_3|) + \left\{Ee^{it(b_1+\bar{b}_2')} - E(e^{itb_1}) - itE(\bar{b}_2'e^{itb_1})\right\} \\
&\quad + itE(\bar{b}_2'e^{itb_1} - \bar{b}_2'e^{itb_2}) + itE(\bar{b}_2'e^{itb_2}) \\
&= E(e^{itb_1}) + itE(\bar{b}_2'e^{itb_2}) + O(|t|E|\bar{b}_2''+\bar{b}_3| + t^2(E|\bar{b}_2'|^2 + E|\bar{b}_2'b_3|)) \quad .
\end{aligned} \tag{3.19}$$

Using (2.6), write

$$\begin{aligned}
E(e^{itb_1}) &= E\left[e^{itb_2}\left\{1 + itb_3 + \frac{(it)^2}{2}b_3^2\right\}\right] \\
&\quad + E\left[e^{itb_2}\left\{e^{itb_3} - 1 - itb_3 - \frac{(it)^2}{2}b_3^2\right\}\right] \\
&= E\left[e^{itb_2}\left\{1 + itb_3 + \frac{(it)^2}{2}b_3^2\right\}\right] + O(|t|^3E|b_3|^3) \quad ,
\end{aligned} \tag{3.20}$$

and put $\gamma(t) = E(e^{it\frac{2}{\sqrt{n}s}V_1})$. As in (2.9)-(2.11),

$$E(e^{itb_2}) = \gamma(t)^n \quad , \tag{3.21}$$

$$\begin{aligned}
E(b_3e^{itb_2}) &= \gamma(t)^{n-2}\left[\frac{4(it)^2}{n^{1/2}}E(W_{12}V_1V_2)\right. \\
&\quad \left. + O\left(\frac{t^2h^L}{n^{1/2}} + \left(\frac{t^4}{n^{3/2}} + \frac{|t|^3}{n}\right)h^{-\frac{2}{3}d-1}\right)\right] \quad ,
\end{aligned} \tag{3.22}$$

$$\begin{aligned}
E(b_3^2e^{itb_2}) &= \frac{2}{n-1}\gamma(t)^{n-2}\left[E(W_{12}^2) + O\left(\frac{h^L}{nh^{d+2}} + |t|n^{-1/2}h^{-\frac{4}{3}d-2}\right)\right] \\
&\quad + \gamma(t)^{n-3}O(|t|n^{-3/2}h^{-\frac{4}{3}d-2}) \\
&\quad + \gamma(t)^{n-4}O(t^4n^{-1} + t^8n^{-3}h^{-\frac{4}{3}d-2} + t^6n^{-2}h^{-\frac{4}{3}d-2}) \quad .
\end{aligned} \tag{3.23}$$

Since for m=0,1,2,3,

$$\gamma(t)^{n-m} = e^{-\frac{t^2}{2}}\left\{1 + \frac{E(2V_1)^3}{6n^{1/2}s^3}(it)^3\right\} + o\left(n^{-1/2}(|t|^3 + t^6)e^{-\frac{t^2}{4}}\right) \quad , \tag{3.24}$$

by (A.1), Appendix C-(a) and (3.17)-(3.24),

$$E\{e^{it(B+s^{-1}\Delta)}\} = \{1+it\Delta+O(t^2nh^{2L}+|t|n^{1/2}h^{2L})\}$$
$$\times\left\{\left[e^{-\frac{t^2}{2}}\left\{1+\frac{4E(V_1^3)}{3n^{1/2}s^3}(it)^3\right\}+o\left(n^{-1/2}(|t|^3+t^6)e^{-\frac{t^2}{4}}\right)\right]\right.$$
$$\times\left[1+\frac{4(it)^3}{n^{1/2}}E(W_{12}V_1V_2)+\frac{(it)^2}{n}E(W_{12}^2)-\frac{2(it)^2}{n}E(W_{12}^2)\right.$$
$$\left.-\frac{(it)^3+(it)}{n^{1/2}}\{4E(V_1^3)+8E(W_{12}V_1V_2)\}+O(A_n')\right]$$
$$\left.+\left(|t|E|\breve{b}_2''+\breve{b}_3|+t^2(E|\breve{b}_2'|^2+E|\breve{b}_2'b_3|)\right)\right\} \qquad (3.25)$$

where

$$A_n' \le C\left\{\frac{|t|^3h^L}{n^{1/2}}+\left(\frac{|t|^5}{n^{3/2}}+\frac{t^4}{n}\right)h^{-\frac{2}{3}d-1}+\frac{t^2}{n^2h^{d+2}}+\frac{t^2h^L}{n^2h^{d+2}}+\frac{|t|^3}{n^{3/2}h^{\frac{4}{3}d+2}}\right.$$
$$+\frac{t^6}{n}+\frac{t^{10}}{n^3h^{\frac{4}{3}d+2}}+\frac{t^8}{n^2h^{\frac{4}{3}d+2}}+\frac{|t|^3}{(nh^{d+2})^{3/2}}$$
$$+\frac{t^2}{n^2h^{d+2}}+\frac{|t|^3}{n^{3/2}h^{d+2}}+\frac{t^2h^L}{nh^{d+2}}+\frac{t^2}{n}$$
$$\left.+\frac{|t|^3+t^4}{n}+\frac{|t|^3}{n^{3/2}}+\frac{|t|^7}{n^3h^{d+2}}+\frac{t^6}{n^{5/2}h^{d+2}}+\frac{|t|^5+t^4}{n^2h^{d+2}}\right\}$$
$$=o\left(\frac{t^2+t^{10}}{nh^{d+2}}+\frac{t^2+t^6}{n^{1/2}}\right).$$

Expanding (3.25), we have

$$E\{e^{it(B+s^{-1}\Delta)}\}=e^{-\frac{t^2}{2}}\left[1+\left\{\Delta-\frac{4E(V_1^3)+8E(W_{12}V_1V_2)}{n^{1/2}}\right\}(it)-\frac{E(W_{12}^2)}{n}(it)^2\right.$$
$$\left.-\frac{4\{2E(V_1^3)+3E(W_{12}V_1V_2)\}}{3n^{1/2}}(it)^3\right]+D_n(t)$$
$$=\chi^+(t)+e^{-\frac{t^2}{2}}\left[\frac{4\{\kappa_3-E(V_1^3)\}+8\{\kappa_4-E(W_{12}V_1V_2)\}}{n^{1/2}}(it)\right.$$
$$-(n^{1/2}h^L\kappa_1-\Delta)(it)+\frac{\kappa_2-h^{d+2}E(W_{12}^2)}{nh^{d+2}}(it)^2$$
$$\left.+\frac{4[\{\kappa_3-2E(V_1^3)\}+3\{\kappa_4-E(W_{12}V_1V_2)\}]}{3n^{1/2}}(it)^3\right]+D_n(t)$$
$$=\chi^+(t)+C_n(t)+D_n(t), \qquad (3.26)$$

using Lemmas 11, 12, 13-(b), 14-(b), where

$$C_n(t) = o\left(e^{-\frac{t^2}{2}}\left(\frac{|t|+|t|^3}{n^{1/2}} + |t|n^{1/2}h^L + \frac{t^2}{nh^{d+2}}\right)\right) ,$$

$$D_n(t) = O\left(\left\{e^{-\frac{t^2}{2}}\frac{|t|^3}{n^{1/2}} + o\left(n^{-1/2}(t^6+|t|^3)e^{-\frac{t^2}{4}}\right)\right\}\left\{\frac{t^2}{n^2h^{d+2}} + \frac{|t|^3+|t|}{n^{1/2}} + A_n'\right\}\right.$$

$$+ e^{-\frac{t^2}{2}}\left(|t|n^{1/2}h^L + t^2h^{2L} + |t|n^{1/2}h^{2L}\right)\left(\frac{t^2}{n^2h^{d+2}} + \frac{|t|^3+|t|}{n^{1/2}} + A_n'\right)$$

$$+ \left(|t|n^{1/2}h^L + t^2nh^{2L}\right)\left\{e^{-\frac{t^2}{2}}\frac{|t|^3}{n^{1/2}} + o\left(n^{-1/2}(t^6+|t|^3)e^{-\frac{t^2}{4}}\right)\right\}$$

$$+ \left(|t|n^{1/2}h^L + t^2nh^{2L}\right)\left\{e^{-\frac{t^2}{2}}\frac{|t|^3}{n^{1/2}} + o\left(n^{-1/2}(t^6+|t|^3)e^{-\frac{t^2}{4}}\right)\right\}$$

$$\times\left(\frac{t^2}{n^2h^{d+2}} + \frac{|t|^3+|t|}{n^{1/2}} + A_n'\right)$$

$$+ \left(|t| + t^2n^{1/2}h^L + |t|^3nh^{2L}\right)E|\tilde{b}_2'' + \tilde{b}_3|$$

$$\left.+ \left(t^2 + |t|^3n^{1/2}h^L + t^4nh^{2L}\right)\left(E|\tilde{b}_2'|^2 + E|\tilde{b}_2'b_3|\right)\right) . \tag{3.27}$$

By Hölder's inequality, (A.7) and Lemmas 23-28,

$$E|\tilde{b}_2''| = E|Q\bar{V}| \leq (E|Q|^2E|\bar{V}|^2)^{1/2} = O(n^{-1}h^{-\frac{d+2}{2}}) \tag{3.28}$$

$$E|\tilde{b}_3| = E|(T+Q)\bar{W}| \leq (E|T+Q|^2E|\bar{W}|^2)^{1/2}$$

$$= O((n^{-1/2}+n^{-1}h^{-d-2})(n^{-1}h^{-d-2})^{1/2}) . \tag{3.29}$$

Writing $E|\tilde{b}_2'|^2 \leq C(|T_1|^2E|\bar{V}|^2 + E|T_2\bar{V}|^2 + E|T_3\bar{V}|^2)$, Lemmas 23, 24, 26 and Hölder's inequality give

$$|T_1|^2E|\bar{V}|^2 + E|T_3\bar{V}|^2 \leq |T_1|^2E|\bar{V}|^2 + (E|T_3|^4E|\bar{V}|^4)^{1/2} = O(n^{-2}h^{-2d-4}) ,$$

and (3.8), (i)', (iii), Lemma 1-(d) and (B.5) give

$$E|T_2\bar{V}|^2 \leq \frac{C}{n^3}E\left|\sum_{i=1}^{n}(4V_i^2 - s^2 + 8\tilde{V}_i)\sum_{j=1}^{n}V_j\right|^2$$

$$= \frac{C}{n^3}\left\{nE|(4V_1^2 - s^2 + 8\tilde{V}_1)V_1|^2 + n(n-1)E|(4V_1^2 - s^2 + 8\tilde{V}_1)V_2|^2\right\}$$

$$= O(n^{-1}) .$$

Thus

$$E|\tilde{b}_2'|^2 = E|T\bar{V}|^2 = O(n^{-1} + n^{-2}h^{-2d-4}) . \tag{3.30}$$

Hölder's inequality, (3.30) and (A.7) yield

$$E|\tilde{b}_2'b_3| = (E|\tilde{b}_2'|^2E|b_3|^2)^{1/2} = O((n^{-1/2}+n^{-1}h^{-d-2})(n^{-1}h^{-d-2})^{1/2}) . \tag{3.31}$$

Using (3.26)-(3.31),

$$\text{(I-2)} \le \int_{-\log n}^{\log n} \left| \frac{E e^{it(B+s^{-1}\Delta)} - \chi^{+}(t)}{t} \right| dt$$
$$\le \int_{-\log n}^{\log n} \left| \frac{C_n(t) + D_n(t)}{t} \right| dt = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) .$$

To estimate (II-2), put $b_3' = T\bar{W}$, $b_3'' = Q\bar{W}$, then noting that $b_3 = b_3' + b_3''$ and $B = b_1 + b_2 + (b_3' + b_3'')$, we have, using (2.6),

$$|Ee^{itB}| \le |Ee^{itB} - Ee^{it(b_1+b_2+b_3')} - it\, Eb_3'' e^{it(b_1+b_2+b_3)}| + |Ee^{it(b_1+b_2+b_3')}|$$
$$+ |t|\, |Eb_3'' e^{it(b_1+b_2+b_3')}|$$
$$\le |t|^2 E|b_3''|^2 + |Ee^{it(b_1+b_2+b_3')}| + |t|\, |Eb_3'' e^{it(b_1+b_2+b_3')}| . \qquad (3.32)$$

Writing $E|b_3''|^2 \le C(E|Q_1\bar{W}|^2 + E|Q_2\bar{W}|^2)$, Hölder's inequality, (A.7) and Lemma 28 give

$$E|Q_2\bar{W}|^2 \le (E|Q_2|^6)^{1/3}(E|\bar{W}|^3)^{2/3} = O((n^{-6}h^{-5d-6})^{1/3}n^{-1}h^{-d-2}) \quad (3.33)$$

and

$$E|Q_1\bar{W}|^2 \le \frac{C}{n^7} E\Big|\sum\sum_{i<j} \{ (V_i+V_j) W_{ij} - \tilde{V}_i - \tilde{V}_j \} \sum\sum_{k<l} W_{kl} \Big|^2$$
$$\le \frac{C}{n^7} E\Big|\sum\sum_{i<j<k<l}\sum\sum \{ (V_i+V_j) W_{ij} - \tilde{V}_i - \tilde{V}_j \} W_{kl} \Big|^2$$
$$+ \frac{C}{n^7} E\Big|\sum\sum_{i<j<l}\sum \{ (V_i+V_j) W_{ij} - \tilde{V}_i - \tilde{V}_j \} W_{il} \Big|^2$$
$$+ \frac{C}{n^7} E\Big|\sum\sum_{i<j} \{ (V_i+V_j) W_{ij} - \tilde{V}_i - \tilde{V}_j \} W_{ij} \Big|^2 ,$$
$$\le \frac{C}{n^7} \sum\sum_{i<j<k<l}\sum\sum E|\{ (V_i+V_j) W_{ij} - \tilde{V}_i - \tilde{V}_j \} W_{kl} |^2$$
$$+ \frac{C}{n^7} \sum\sum_{j>l\ge 2} n^2 E|\{ (V_1+V_j) W_{1j} - \tilde{V}_1 - \tilde{V}_j \} W_{1l} |^2$$
$$+ \frac{C}{n^7} n^4 E|\{ (V_1+V_2) W_{12} - \tilde{V}_1 - \tilde{V}_2 \} W_{12} |^2$$
$$= O(n^{-3}h^{-3d-4}) , \qquad (3.34)$$

where the third inequality uses the Theorem of DFJ, and the equality uses nested conditional expectation, Lemmas 1-(d), 4, 16, and (A.5). Therefore by (3.33) and (3.34),

$$E|b_3''|^2 = E|Q\bar{W}|^2 = O(n^{-3}h^{-3d-4}) . \qquad (3.35)$$

To investigate the second term of (3.32), let

$$d_i = (4V_i^2 - s^2) + 8\tilde{V}_i \quad , \; e_{ij} = 4\left\{(V_i + V_j)W_{ij} - \tilde{V}_i - \tilde{V}_j + \frac{n}{(n-2)}\tilde{W}_{ij}\right\} \; , \tag{3.36}$$

then

$$\bar{b}_2 = -\frac{s^{-3}}{2}\left\{\frac{4n}{(n-2)^2}E(W_{12}^2) + \frac{1}{n}\sum_{i=1}^{n} d_i + \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e_{ij} - \frac{4}{n}\binom{n-1}{2}^{-1}\sum_{i=1}^{n}\sum_{k=1}^{n-1\,(i)}\sum_{l=k+1}^{n\,(i)} V_i W_{kl}\right\}\bar{V} \; ,$$

$$\bar{b}_3' = -\frac{s^{-3}}{2}\left\{\frac{4n}{(n-2)^2}E(W_{12}^2) + \frac{1}{n}\sum_{i=1}^{n} d_i + \binom{n}{2}^{-1}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} 4\tilde{W}_{jk}\right\}\bar{W} \; .$$

Define

$$b_{3m} = s^{-1} n^{\frac{1}{2}}\binom{n}{2}^{-1}\sum_{i=1}^{m}\sum_{j=i+1}^{n} W_{ij} \quad ,$$

$$\bar{b}_{2m} = -\frac{s^{-3}}{2}\Bigg[\frac{8n^{1/2}}{(n-2)^2}E(W_{12}^2)\sum_{i=1}^{m} V_i + \frac{2}{n^{3/2}}\left(\sum_{i=1}^{n}\sum_{s=1}^{m} d_i V_s + \sum_{i=1}^{m}\sum_{s=m+1}^{n} d_i V_s\right)$$

$$+ \frac{2}{n^{1/2}}\binom{n-1}{2}^{-1}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{s=1}^{m} e_{ij} V_s + \sum_{i=1}^{m}\sum_{j=i+1}^{n}\sum_{s=m+1}^{n} e_{ij} V_s\right)$$

$$- \frac{8}{n^{3/2}}\binom{n-1}{2}^{-1}\Bigg(\sum_{i=1}^{n}\sum_{k<l}^{(i)}\sum_{s=1}^{m} V_i W_{kl} V_s + \sum_{i=1}^{n}\sum_{k=1}^{m\,(i)}\sum_{l=k+1}^{n\,(i)}\sum_{s=m+1}^{n} V_i W_{kl} V_s$$

$$+ \sum_{i=1}^{m}\sum_{k=m+1}^{n-1\,(i)}\sum_{l=k+1}^{n\,(i)}\sum_{s=m+1}^{n} V_i W_{kl} V_s\Bigg)\Bigg] \; ,$$

$$\bar{b}_{3m}' = -\frac{s^{-3}}{2}\Bigg[\frac{4n^{3/2}}{(n-2)^2}E(W_{12}^2)\binom{n}{2}^{-1}\sum_{l=1}^{m}\sum_{s=l+1}^{n} W_{ls}$$

$$+ \frac{1}{\sqrt{n}}\binom{n}{2}^{-1}\left(\sum_{i=1}^{m}\sum_{l=1}^{n-1}\sum_{s=l+1}^{n} d_i W_{ls} + \sum_{i=m+1}^{n}\sum_{l=1}^{m}\sum_{s=l+1}^{n} d_i W_{ls}\right)$$

$$+ n^{1/2}\binom{n}{2}^{-2}\left(\sum_{j=1}^{m}\sum_{k=j+1}^{n}\sum_{l=1}^{n-1}\sum_{s=l+1}^{n} 4\tilde{W}_{jk} W_{ls} + \sum_{j=m+1}^{n-1}\sum_{k=j+1}^{n}\sum_{l=1}^{m}\sum_{s=l+1}^{n} 4\tilde{W}_{jk} W_{ls}\right)\Bigg]$$

for $m=1,\ldots,n$-1. Note that $b_1 - b_{3m}$, $\bar{b}_2 - \bar{b}_{2m}$ and $\bar{b}_3' - \bar{b}_{3m}'$ are independent of $(X_1^T, Y_1), \cdots, (X_m^T, Y_m)$. Putting $\bar{B}_m = (b_1 - b_{3m}) + (\bar{b}_2 - \bar{b}_{2m}) + (\bar{b}_3' - \bar{b}_{3m}')$, and using (2.6) repeatedly, we have

$$|Ee^{it(b_1 + \bar{b}_2 + \bar{b}_3')}| \le \frac{t^2}{2}E|\bar{b}_{2m} + \bar{b}_{3m}'|^2 + |Ee^{it(\bar{B}_m + b_{3m})}|$$

$$+ |t|\,|Ee^{it(\bar{B}_m + b_{3m})}(\tilde{b}_{2m} + \tilde{b}'_{3m})|$$

$$\leq \frac{t^2}{2}E|\tilde{b}_{2m} + \tilde{b}'_{3m}|^2 + \left[\frac{|t|^3}{6}E|b_{3m}|^3 + |Ee^{it\bar{B}_m}\{1 + it\,b_{3m} + \frac{(it)^2}{2}b_{3m}^2\}|\right]$$

$$+ \left[t^2E|b_{3m}|\,|\tilde{b}_{2m} + \tilde{b}'_{3m}| + |t|\,|Ee^{it\bar{B}_m}(\tilde{b}_{2m} + \tilde{b}'_{3m})|\right]$$

$$= \left[\frac{t^2}{2}E|\tilde{b}_{2m} + \tilde{b}'_{3m}|^2 + t^2E|b_{3m}|\,|\tilde{b}_{2m} + \tilde{b}'_{3m}|\right]$$

$$+ \left[\frac{|t|^3}{6}E|b_{3m}|^3 + |Ee^{it\bar{B}_m}\{1 + it\,b_{3m} + \frac{(it)^2}{2}b_{3m}^2\}|\right]$$

$$+ \left[|t|\,|Ee^{it\bar{B}_m}(\tilde{b}_{2m} + \tilde{b}'_{3m})|\right] . \tag{3.37}$$

By elementary inequalities, (ix), Appendix C-(d), (e) and (A.7), the first bracketed term is bounded by

$$Ct^2\{E|\tilde{b}_{2m}|^2 + E|\tilde{b}'_{3m}|^2 + (E|b_{3m}|^2)^{1/2}(E|\tilde{b}_{2m}|^2 + E|\tilde{b}'_{3m}|^2)^{1/2}\}$$
$$\leq Cmt^2\left\{\frac{1}{n^3h^{2d+4}} + \frac{1}{n^2} + \frac{1}{n^4h^{3d+6}}\right.$$
$$\left. + \frac{1}{(n^2h^{d+2})^{1/2}}\left(\frac{1}{(n^3h^{2d+4})^{1/2}} + \frac{1}{n} + \frac{1}{(n^4h^{3d+6})^{1/2}}\right)\right\}$$
$$\leq Cmt^2\left(\frac{1}{n^2h^{\frac{d+2}{2}}} + \frac{1}{n^{\frac{5}{2}}h^{\frac{3d+6}{2}}}\right) . \tag{3.38}$$

The second bracketed term on the right of (3.37) is bounded by

$$C\left[|t|^3\left(\frac{m}{n^2h^{d+2}}\right)^{3/2} + \left\{1 + \frac{m|t|}{n^{1/2}h} + \frac{m|t|}{n^2h^{d+2}} + \frac{t^2m^2}{nh^2}\right\}|\gamma(t)|^{m-4}\right] , \tag{3.39}$$

which is verified as in equations (13)-(19) of Robinson (1995a), because s^{-1} is bounded due to (3.8) and $\bar{B}_m$ is the sum of $\frac{2}{\sqrt{ns}}\sum_{i=1}^{n}V_i$ and $(b_3 - b_{3m}) + (\tilde{b}_2 - \tilde{b}_{2m}) + (\tilde{b}'_3 - \tilde{b}'_{3m})$, the latter being independent of (X_1^T, Y_1) , … , (X_m^T, Y_m) . Appendix C-(b), (c) bound the last term in (3.37) by

$$\frac{Cm|t|}{n^{1/2}h^3}|\gamma(t)|^{m-4} . \tag{3.40}$$

Now, we investigate the third term on the right of (3.32). Using elementary inequalities, (2.6), (3.35), (A.7), Appendix C-(d), (e), (f),

$$|E\tilde{b}''_3e^{it(b_1 + \tilde{b}_2 + \tilde{b}'_3)}|$$
$$\leq |E\tilde{b}''_3e^{it(b_1 + \tilde{b}_2 + \tilde{b}'_3)} - E\tilde{b}''_3e^{it(b_1 - b_{3m} + \tilde{b}_2 - \tilde{b}_{2m} + \tilde{b}'_3 - \tilde{b}'_{3m})}| + |E\tilde{b}''_3e^{it\bar{B}_m}|$$

$$\le |t| \, |E| b_3''| \, |b_{3m}+b_{2m}+b'_{3m}| + |E b_3'' e^{it\bar{B}_m}|$$

$$\le C|t| \, |(E|b_3''|^2)^{1/2}[(E|b_{3m}|^2)^{1/2}+(E|b_{2m}|^2)^{1/2}+(E|b'_{3m}|^2)^{1/2}] + |E b_3'' e^{it\bar{B}_m}|$$

$$\le \frac{C|t|h}{(nh^{d+2})^{3/2}}\left[(\frac{m}{n^2h^{d+2}})^{1/2}+(\frac{m}{n^3h^{2d+4}})^{1/2}+(\frac{m}{n^2})^{1/2}+(\frac{m}{n^4h^{3d+6}})^{1/2}\right]$$

$$+ \frac{Cn^{1/2}}{h^2}|\gamma(t)|^{m-5}$$

$$\le \frac{C|t|hm^{1/2}}{n^{1/2}(nh^{d+2})^2} + \frac{Cn^{1/2}}{h^2}|\gamma(t)|^{m-5} . \qquad (3.41)$$

Therefore, by (3.32), (3.35), (3.37)-(3.41),

$$|Ee^{itB}| \le \frac{Ct^2}{n^3h^{3d+4}} + Cmt^2\left(\frac{1}{n^2h^{\frac{d+2}{2}}} + \frac{1}{n^{\frac{5}{2}}h^{\frac{3d+6}{2}}}\right)$$

$$+ C\left[|t|^3(\frac{m}{n^2h^{d+2}})^{3/2} + \left\{1+\frac{m|t|}{n^{1/2}h}+\frac{m|t|}{n^2h^{d+2}}+\frac{t^2m^2}{nh^2}\right\}|\gamma(t)|^{m-4}\right]$$

$$+ \frac{Cm|t|}{n^{1/2}h^3}|\gamma(t)|^{m-4}$$

$$+ \frac{Chm^{1/2}t^2}{n^{1/2}(nh^{d+2})^2} + \frac{Cn^{1/2}|t|}{h^2}|\gamma(t)|^{m-5} . \qquad (3.42)$$

Now, divide (3.42) by $|t|$ and integrate over $p_2 \le |t| \le N_0$, where we partition the range of integration into two parts, $p_2 \le |t| \le N_1$ and $N_1 \le |t| \le N_0$, for $N_1 = \min(\bar{\epsilon} n^{1/2}, nh^{d+2})$.

(i) $p_2 \le |t| \le N_1$

We can choose $m = [9n\log n / t^2]$ to satisfy $1 \le m \le n-1$ for large n. For this m , since $E(2V_1/s)=0$ and $Var(2V_1/s)=1$, as in (2.21),

$$|\gamma(t)|^{m-5} \le \exp(-\frac{m-5}{3n}t^2) \le C\exp(-3\log n) = \frac{C}{n^3} . \qquad (3.43)$$

By (3.42), (3.43), $|\gamma(t)|<1$ and (ix), we obtain

$$\int_{p_2\le|t|\le N_1}\left|\frac{Ee^{itB}}{t}\right|dt$$

$$\le \frac{C}{n^3h^{3d+4}}\int_{p_2\le|t|\le nh^{d+2}}|t|dt$$

$$+ C\left(\frac{n\log n}{n^2h^{\frac{d+2}{2}}}+\frac{n\log n}{n^{\frac{5}{2}}h^{\frac{3d+6}{2}}}\right)\int_{p_2\le|t|\le\bar{\epsilon}n^{1/2}}\frac{dt}{|t|}$$

$$+ C\left(\frac{n\log n}{n^2h^{d+2}}\right)^{3/2}\int_{p_2\le|t|\le\bar{\epsilon}n^{1/2}}\frac{dt}{|t|}$$

$$+\frac{C}{n^3}\int_{p_2\le|t|\le\bar{\epsilon}n^{1/2}}\left\{\frac{1}{|t|}+\frac{n\log n}{n^{1/2}ht^2}+\frac{(n\log n)^2}{nh^2|t|^3}\right\}dt$$

$$+ \frac{Cnlogn}{n^{7/2}h^3}\int_{P_2\le|t|\le\bar{\varepsilon}n^{1/2}}\frac{dt}{t^2}$$

$$+ \frac{Ch(nlogn)^{1/2}}{n^{1/2}(nh^{d+2})^2}\int_{P_2\le|t|\le nh^{d+2}}dt + \frac{C}{n^{5/2}h^2}\int_{P_2\le|t|\le\bar{\varepsilon}n^{1/2}}dt$$

$$= o(n^{-1/2} + n^{-1}h^{-d-2}) \quad . \qquad (3.44)$$

(ii) $N_1 \le |t| \le N_0$

For sufficiently large n, there exists $\eta_2 \in (0, 1)$ such that $|\gamma(t)| < 1-\eta_2$ since $|\gamma_v(t)| \le \eta_1$ by assumption (x) and

$$|\gamma(t) - \gamma_v(t)| = |E[e^{it\frac{2v_1}{\sqrt{n}}}\{e^{it(\frac{2V_1}{\sqrt{ns}}-\frac{2v_1}{\sqrt{n}})} - 1\}]| \le CE|\frac{tV_1}{\sqrt{ns}} - \frac{tv_1}{\sqrt{n}}|$$

$$\le CE|\frac{tV_1}{\sqrt{ns}} - \frac{tV_1}{\sqrt{n}}| + CE|\frac{tV_1}{\sqrt{n}} - \frac{tv_1}{\sqrt{n}}| = O(\frac{|t|h^L}{\sqrt{n}}) = O(h^L logn) = o(1) \, ,$$

using (2.6) in the first inequality, (3.8), Lemma 1-(d) and (B.1) in the second equality and (ix) in the last equality. We may take $m=[-3logn / \log(1-\eta_2)]$ to satisfy $1 \le m \le n-1$ for sufficiently large n. Since $|\gamma(t)|^{m-5} \le Cn^{-3}$,

$$\int_{N_1\le|t|\le N_0}|\frac{Ee^{itB}}{t}|dt$$

$$\le \frac{C}{n^3h^{3d+4}}\int_{N_1\le|t|\le nh^{d+2}logn}|t|dt$$

$$+ C\left(\frac{\log n}{n^2h^{\frac{d+2}{2}}} + \frac{\log n}{n^{\frac{5}{2}}h^{\frac{3d+6}{2}}}\right)\int_{N_1\le|t|\le\bar{\varepsilon}n^{1/2}logn}|t|dt$$

$$+ C(\frac{\log n}{n^2h^{d+2}})^{3/2}\int_{N_1\le|t|\le\bar{\varepsilon}n^{1/2}logn}|t|^2dt$$

$$+ \frac{C}{n^3}\int_{N_1\le|t|\le\bar{\varepsilon}n^{1/2}logn}\{\frac{1}{|t|} + \frac{logn}{n^{1/2}h} + \frac{(logn)^2}{nh^2}|t|\}dt$$

$$+ \frac{C\log n}{n^{7/2}h^3}\int_{N_1\le|t|\le\bar{\varepsilon}n^{1/2}logn}dt$$

$$+ \frac{Ch(\log n)^{1/2}}{n^{1/2}(nh^{d+2})^2}\int_{N_1\le|t|\le nh^{d+2}logn}|t|dt$$

$$+ \frac{C}{n^{5/2}h^2}\int_{N_1\le|t|\le\bar{\varepsilon}n^{1/2}logn}dt$$

$$= o(n^{-1/2} + n^{-1}h^{-d-2}) \qquad (3.45)$$

by (ix). Therefore, by (3.44) and (3.45),

$$\text{(II-2)} = o(n^{-1/2} + n^{-1}h^{-d-2}) \, . \qquad \square$$

3.2 An empirical Edgeworth expansion

The κ_i are unknown, but as in Theorem 2, we can construct a feasible, empirical Edgeworth expansion

$$\hat{F}^{+}(z) = \Phi(z) - \phi(z)\left[n^{1/2}h^{L}\tilde{\kappa}_1 - \frac{\tilde{\kappa}_2}{nh^{d+2}}z - \frac{1}{3n^{1/2}}\{(2z^2+1)\tilde{\kappa}_3 + 3(z^2+1)\tilde{\kappa}_4\}\right],$$

THEOREM 4 : *Under* (i)', (ii), (iii), (iv)', (v)', (vi) - (viii), (ix)' *and* (x)-(xii),

$$\sup_{v:\, v^{\tau}v=1}\ \sup_{z\in R} |\hat{F}(z) - \hat{F}^{+}(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \quad \textit{completely.}$$

PROOF OF THEOREM 4

Straightforward from Theorem 3 and $\tilde{\kappa}_i \to \kappa_i$ completely for $i = 1, 2, 3, 4$ (see the proof of Theorem 2). □

We have seen in Section 2.3 that (i)', (iv)', (v)' and (ix)' are necessary to prove $\tilde{\kappa}_i \to \kappa_i$ completely for $i = 1, 2, 3, 4$. It is obvious that we need these four assumptions here additional to those assumed in Theorem 3. Eventually the same set of assumptions as that in Theorem 2 is sufficient here.

3.3 Discussion

This section discusses the followings regarding the Edgeworth expansions we have verified. 3.3.1 shows which correction term(s) is dominant in the expansion depending on the dimension of explanatory variables, kernel order and bandwidth choice. 3.3.2 proposes optimal bandwidth choices which minimizes the normal approximation error. 3.3.3 compares the results here with those in Linton (1996b).

3.3.1 Comparison of the order of Edgeworth correction terms.

Theorem 3 covers a number of situations, depending on the choice of kernel order L, relative to dimension d, and on the rate of decay of the bandwidth h. We classify these according to L and then h. A similar classification was used by Robinson (1995b) in expansions for Nadaraya-Watson regression estimates allowing higher order kernels under heteroscedastic but conditionally normal errors, with similar interpretations. Let C_i, $i=1,2,3,4$, be finite positive constants.

I. $\frac{d+2}{2} < L < 2(d+2)$.

(a) If $n^3 h^{2(L+d+2)} \to 0$,

$$\hat{F}(z) = \Phi(z) + \frac{\kappa_2 z\,\phi(z)}{nh^{d+2}}\{1 + o(1)\} .$$

(b) If $h \sim C_1 n^{-\frac{3}{2(L+d+2)}}$,

$$\hat{F}(z) = \Phi(z) - (C_1^L \kappa_1 - \frac{\kappa_2 z}{C_1^{d+2}})\,\phi(z)\, n^{\frac{-2L+d+2}{2(L+d+2)}}\{1 + o(1)\} .$$

(c) If $n^3 h^{2(L+d+2)} \to \infty$,

$$\hat{F}(z) = \Phi(z) - \kappa_1 \phi(z)\, n^{1/2} h^L \{1 + o(1)\} .$$

II. $L = 2(d+2)$.

(a) If $n^{1/2} h^{d+2} \to 0$,

$$\hat{F}(z) = \Phi(z) + \frac{\kappa_2 z\,\phi(z)}{nh^{d+2}}\{1 + o(1)\} .$$

(b) If $h \sim C_2 n^{-\frac{1}{2(d+2)}}$,

$$\hat{F}(z) = \Phi(z) - \left[C_2^L \kappa_1 - \frac{\kappa_2 z}{C_2^{d+2}} - \frac{4\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\}}{3} \right] \frac{\phi(z)}{n^{1/2}}\{1 + o(1)\}$$

(c) If $n^{1/2} h^{d+2} \to \infty$,

$$\hat{F}(z) = \Phi(z) - \kappa_1 \phi(z)\, n^{1/2} h^L \{1 + o(1)\} .$$

III. $L > 2(d+2)$.

(a) If $nh^L + \frac{1}{n^{1/2}h^{d+2}} \to 0$,

$$\hat{F}(z) = \Phi(z) + 4\{(2z^2+1)\kappa_3 + 3(z^2+1)\kappa_4\}\frac{\phi(z)}{3n^{1/2}}\{1 + o(1)\} .$$

(b) If $h \sim C_3 n^{-\frac{1}{L}}$,

$$\hat{F}(z) = \Phi(z) - \left[C_3^L\kappa_1 - \frac{4\{(2z^2+1)\kappa_3+3(z^2+1)\kappa_4\}}{3}\right]\frac{\phi(z)}{n^{1/2}}\{1 + o(1)\} .$$

(c) If $h \sim C_4 n^{-\frac{1}{2(d+2)}}$,

$$\hat{F}(z) = \Phi(z) + \left[\frac{\kappa_2 z}{C_4^{d+2}} + \frac{4\{(2z^2+1)\kappa_3+3(z^2+1)\kappa_4\}}{3}\right]\frac{\phi(z)}{n^{1/2}}\{1 + o(1)\} .$$

(d) If $n^{1/2}h^{d+2} \to 0$,

$$\hat{F}(z) = \Phi(z) + \frac{\kappa_2 z\phi(z)}{nh^{d+2}}\{1 + o(1)\} .$$

(e) If $nh^L \to \infty$,

$$\hat{F}(z) = \Phi(z) - \kappa_1\phi(z)n^{1/2}h^L\{1 + o(1)\} .$$

In each of the seven cases I(a)-(c), II(a), II(c), III(d), and III(e), the correction term in the expansion is of larger order than $n^{-1/2}$. In the other four cases it is of exact order $n^{-1/2}$, but of these the cases I(b), II(b), III(b), and III(c), which involve a knife-edge choice of bandwidth, include κ_1 or κ_2 (which depend on the kernel K) or both in the correction term. It is case III(a) which corresponds in detail to the "parametric" situation in the sense of discussion in Section 2.2. We can also derive analogous expressions based on Theorem 1. For example, for (L, d, h) satisfying III(a), we have

$$F(z) = \Phi(z) - \frac{4(\kappa_3+3\kappa_4)}{3n^{1/2}}(z^2-1)\phi(z)\{1 + o(1)\} . \qquad (3.46)$$

3.3.2 Bandwidth selection

For U and related statistics, Härdle, Hart, Marron, and Tsybakov (1992), Härdle and Tsybakov (1993), and Powell and Stoker (1996) derived h that are optimal in the sense of asymptotically minimizing leading terms in the mean squared error (MSE). As seen in (1.49), these optimal h are of form,

$$h^* = C^* n^{-2/(2L+d+2)}, \quad 0 < C^* < \infty, \tag{3.47}$$

where we are in one of the cases I(c), II(c) or III(e), in each of which the leading correction term is $-\kappa_1 \phi(z) n^{1/2} h^L$, so that bias correction has the greatest impact in improving the quality of the normal approximation. However, the conventional approach of relating choice of h to MSE is not directed towards producing a version of the statistic which, in some sense, makes the normal approximation especially good, and in the present context the latter goal is relevant. Under (3.47)

$$F(z) = \Phi(z) - C^{*L}\kappa_1 \phi(z) n^{-\frac{2L-d-2}{2(2L+d+2)}} \{1 + o(1)\} . \tag{3.48}$$

Here, the order of the correction term can be as large as $n^{-1/2(2d+5)}$ when $L=(d+3)/2$ (see Assumption (iv)) and tends to $n^{-1/2}$ only as $L/d \to \infty$, so (3.47) is certainly not optimal in the sense of minimizing the error in the normal approximation. As shown in 1.1.4, the h which minimizes the integrated MSE of nonparametric derivative-of-density estimates is of form $h^+ = C^+ n^{-1/(2L+d+2)}$, for $0 < C^+ < \infty$, but this is even larger than (3.47) and thus provides an even larger correction term than (3.48). Robinson (1995a) calculated the rate of decay of h that minimizes the order of the normal approximation error. This exceeds $n^{-1/2}$ due to choosing $L<2(d+2)$, and the more detailed information provided by our Edgeworth expansion allows us to discuss the choice of h itself. In particular, the optimal rate of h here is that in I(b) as described by Robinson (1995a), but we would like to know how to choose C_1 in

$$h = C_1 n^{-\frac{3}{2(L+d+2)}} . \tag{3.49}$$

One possibility is to minimize the maximal deviation from the normal approximation, by

$$C_1^A = \underset{C}{argmin} \; \underset{z \in R}{max} \left| \left(C^L \kappa_1 - \frac{\kappa_2 z}{C^{d+2}} \right) \phi(z) \right| .$$

Because $\kappa_2 > 0$ this equals

$$\underset{C}{argmin} \; \underset{z \in R}{max} (C^L |\kappa_1| + \frac{\kappa_2 z}{C^{d+2}}) \phi(z)$$
$$= \underset{C}{argmin} \; \{ C^L |\kappa_1| + \frac{\kappa_2 Z^*(C)}{C^{d+2}} \} \phi(Z^*(C)) \; ,$$

where

$$Z^*(C) = C^{d+2} \{ (C^{2L} \kappa_1^2 + 4\kappa_2^2 / C^{2d+4})^{1/2} - C^L |\kappa_1| \} / 2\kappa_2 \; .$$

Using the envelope theorem, the first order condition of minimization with respect to C is

$$\left\{ L C^{L-1} |\kappa_1| - \frac{(d+2)\kappa_2}{C^{d+3}} Z^*(C) \right\} \phi(Z^*(C)) = 0 \; . \tag{3.50}$$

Solving (3.50), we derive

$$C_1^A = \left\{ \frac{(d+2)^2 \kappa_2^2}{L(L+d+2)\kappa_1^2} \right\}^{\frac{1}{2(L+d+2)}} \tag{3.51}$$

The second order condition is easily verified using (3.50) and $Z^{*\prime}(C) < 0$. Though (3.51) is infeasible since it involves unknown κ_1 and κ_2, we can replace κ_1 and κ_2 by their estimates $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ in Section 2.3 to give the feasible version

$$\tilde{C}_1^A = \left\{ \frac{(d+2)^2 \tilde{\kappa}_2^2}{L(L+d+2)\tilde{\kappa}_1^2} \right\}^{\frac{1}{2(L+d+2)}} \tag{3.52}$$

The estimates $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$, introduced to provide empirical Edgeworth expansions (Theorems 2 and 4), are consistent under the conditions stated there, so that $\tilde{C}_1^A$ is consistent for the optimal C_1^A .

One could consider variants of this idea for bandwidth choice, for example maximizing with respect to z over some desired proper subset of R , such as $\{z : |z| > a\}$ for some $a > 0$, perhaps to stress one of the usual critical regions. However, the simple forms (3.51) and (3.52) seem appealing. Hall and Sheather (1988) (see also Hall, 1992, p.321) used an Edgeworth expansion for studentized sample quantiles, especially the median, to determine a choice of the bandwidth employed in the studentization. In their problem, the basic $n^{1/2}$ - consistent statistic of interest, the sample quantile, does not involve a bandwidth. In our case, on the other hand, though we also consider studentization involving a bandwidth, it is the bandwidth in the basic statistic of interest, the averaged derivative, that is to be chosen using the Edgeworth expansion. Moreover, unlike us, Hall and Sheather (1988) did not maximize over the argument z, but simply balanced the mean and variance terms of the expansion for given z, so that their data-dependent bandwidth is z-dependent (and thus a 'local' bandwidth). It might be anticipated that the step of maximizing over z, which is incorporated in our procedure, would lead to a more complicated, perhaps only implicitly-defined, formula for the optimal C, and the emergence nevertheless of the simple closed form (3.51) is of some interest. We believe our 'global' approach could be employed in choosing the bandwidth in other semiparametric and nonparametric problems involving smoothing.

3.3.3 Comparison of the results with Linton (1996b)

We now precisely compare Theorems 1-4 with the results in Linton (1996b). The statistics considered here and in Linton are quite similar in the sense that both estimators are written in the form of standard U statistics, of order two here and of order up to three in Linton, and both established valid Edgeworth expansions. However, the two are different in the regularity conditions as follows. In order for the comparison, we rewrite our problem in Linton's setup. The objective function is

$$\Psi_n(Y_i, X_i^T; \bar{\mu}, \tilde{f}^{\prime}(x)) = \frac{1}{n}\sum_{i=1}^{n}\left\{\bar{\mu} - Y_i\tilde{f}^{\prime}(X_i)\right\}^T\left\{\bar{\mu} - Y_i\tilde{f}^{\prime}(X_i)\right\} .$$

1. Linton did not handle the bias explicitly thinking of mean subtracted U statistics claiming that bias is deterministic and can be handled analytically. However, we show in the above that the bias term affects an optimal bandwidth selection which minimizes normal approximation error. Therefore we should incorporate the bias effect explicitly into the Edgeworth expansions in the current problem.

2. Linton's regularity conditions require a better than $n^{1/4}$ - consistent estimate $\tilde{f}^{\prime}(x)$ of $f^{\prime}(x)$. Since $\tilde{f}^{\prime}(x)$ is $\sqrt{nh^{d+2}}$ -consistent for $f^{\prime}(x)$, $h = O(n^{-\frac{1}{2(d+2)}+\epsilon})$, where $\epsilon > 0$, is necessary, which corresponds to the cases I(c),

II(c) or III(a), (b), (e) depending on the values of ϵ, L and d. The bias related term has the primary effect on the expansion in four cases out of the five so that this condition on the order of derivatives estimation entails explicit treatment of the bias in our problem.

3. Linton worked with a fixed design case unlike us. It could be extended to conditional arguments, but we would need additional conditions for conditional validity of the Edgeworth or asymptotic expansions.

4. Linton's Assumption B4(2) assumes the orthogonality condition

$$E\left\{\frac{\partial^2 \Psi(\bar{\mu}, f')}{\partial \bar{\mu}\, \partial f'^{\tau}}\right\} = 0 \quad \text{where} \quad \Psi(\bar{\mu}, f') = \{\bar{\mu} - Yf'(X)\}^{\tau}\{\bar{\mu} - Yf'(X)\} \quad ,$$

which ensures that there is no "information loss" resulting from estimation of $f'(x)$, but this is not satisfied unless $E(Y) = E\{G(\beta^{\tau}X)\} = 0$ in the current problem.

5. Linton's Theorem 4.1 looks to validate Edgeworth expansion with error $o(n^{-1})$, however as he addresses in the note 4, what he established was the validity of an order $n^{-1/2}$ Edgeworth expansion. In view of the proof of Theorem 4.1 of Linton, we see that Assumption D4 is very crucial and high level. He follows the proof of Callaert, Janssen and Veraverbeke (1981) or Bickel, Götze and van Zwet (1986), where one of the main points is to find low level conditions under which D4 is satisfied. We actually even do not know D4 is consistent with the other assumptions in general. In this sense both Linton and this work are concerned with Edgeworth expansions of degree one.

6. Linton was more concerned with the situation where the U statistic type semiparametric estimators have qualitatively the same Edgeworth expansion as that of standard U statistics obtained in Callaert *et. al.* (1981) and Bickel *et.al.* (1986), namely the correction terms are of exactly order $n^{-1/2}$ and n^{-1} . What we have established in the previous chapters are more concerned with how it can be different from standard U statistics due to the first stage nonparametric density estimates,

namely how Edgeworth correction terms of larger-than- $n^{-1/2}$ can appear.

7. Like Theorems 1 and 2 in the previous chapters, Linton only investigated the statistic normalized by its asymptotic variance in which sense it is infeasible, however we also validate Edgeworth expansions of a studentized statistic in Theorem 3 and 4, amongst the latter is fully feasible so that we can apply the result for higher order inference of empirical data.

Chapter 4

Bootstrap Distribution for Averaged Derivatives

Bootstrapping is one of the resampling methods intensively studied lately. Given asymptotically normally distributed statistics, it has been shown in some cases that the bootstrap distribution can approximate the exact distribution as good as its Edgeworth expansion up to certain order and thus better than the normal distribution. We review established results on bootstrap distribution in Section 4.1. Section 4.2 compares bootstrap distribution of the density-weighted AD with its exact distribution for normalized case, where we prove the former approximate the latter as good as the Edgeworth expansions in Theorems 1 and 2 and so better than the normal distribution. This property is useful to improve the approximation to the exact distribution of the estimator. This is however infeasible due to the unstudentization. We further conjecture the possibility of analogous property for studentized statistics.

4.1 Bootstrap Distribution

Quenouille (1949, 1956) and Tukey (1958) consider nonparametric methods to estimate bias and variance of estimators by resampling called jackknife method. Efron (1979) proposes a different resampling principle called the bootstrap method for the same purpose. Efron (1982), Efron and Tibshirani (1993) and Hall (1992, Chapter 1) gives good summary of the jackknife and bootstrap methods. We explain the principles of bootstrap method following Hall's account. Suppose we have an iid sample X_i, $i = 1, 2, \ldots, n$ of a variate X with distribution function F_0 and would like to estimate a parameter $\theta_0 = \theta(F_0)$. Let F_1 be the empirical distribution function of the sample, then $\theta_1 = \theta(F_1)$ is a sensible estimator of θ_0 . We would like to know the properties of θ_1 , for example, the bias $E_{F_0}(\theta_1 - \theta_0)$ where E_F

indicates expectation with respect to distribution function F. The core idea of bootstrap method is that we may infer some aspects of (F_0, F_1) from (F_1, F_2) where F_2 is the empirical distribution function of a sample drawn form F_1 . For instance, $E_{F_0}[\theta(F_1) - \theta(F_0)]$, the true bias of $\theta(F_1)$, can be estimated by $\tilde{b} = E_{F_1}[\theta(F_2) - \theta(F_1)]$. Then $\theta_1 - \tilde{b}$ is a bias reduced estimator of θ_0 and typically the first order bias of $O(n^{-1})$ is removed. This principle can be used to estimate the confidence interval and other quantities related to the distribution of θ_1 .

Suppose $\theta(F_1)$ is asymptotically normally distributed. We can expect that $H_2(x) \equiv P(\theta(F_2) \leq x | X_1, \ldots, X_n)$, called the bootstrap distribution of $\theta(F_1)$, estimates $H_1(x) \equiv P(\theta(F_1) \leq x)$, the distribution function of $\theta(F_1)$. Its asymptotic properties have been studied and it is shown that $H_2(x)$ not only converges to a normal distribution but also approximates $H_1(x)$ at least as well as Edgeworth expansions of $\theta(F_1)$ in some cases. Bickel and Freedman (1981) consider the sample average. Supposing X_i, $i = 1, 2, \ldots$ is an iid sample of d dimensional random variate X and $\{X_1^*, \ldots, X_m^*\}$ is its bootstrap sample of size m, they prove the conditional distribution function of $m^{1/2}(\frac{1}{m}\sum^{m} X_{j}^* - \frac{1}{n}\sum^{n} X_{i})$

give$m$, they prove the conditional distribution function of $m^{1/2}(\frac{1}{m}\sum_{j=1} X_j^* - \frac{1}{n}\sum_{i=1} X_i)$

zero as n and m tends to infinity. They also extends the result to von Mises functionals. Singh (1981) proves that the bootstrap distribution of the suitably normalized sample average is at least as good as its one term Edgeworth expansion uniformly when $d=1$. Beran (1982) proves both the bootstrap distribution and one term Edgeworth expansion of asymptotically normally distributed real valued statistics constructed from an iid sample with unknown distribution function F are optimal in an asymptotically minimax sense among possible estimates of its sampling distribution when $d=1$. Beran (1984) shows the asymptotic optimality of the bootstrap distribution and the jackknife Edgeworth expansion estimate where the unknown quantities in the Edgeworth expansion are replaced by their jackknife estimates when $d=1$. Babu and

Singh (1984) show the bootstrap distribution and the Edgeworth expansion of functions of sample average of multivariate iid observations are asymptotically equivalent up to the term of order $n^{-s/2+1}$ under the existence of s-th absolute moment. Bhattacharya and Qumsiyeh (1989) prove the bootstrap distribution of studentized statistics constructed from iid random vector can approximate even better than the two term empirical Edgeworth expansion under certain moment conditions. Helmers (1991) proves the asymptotic equivalence of bootstrap distribution and one term Edgeworth expansion of studentized U-statistics.

4.2 Bootstrap distribution of U

Section 4.1 reviews that bootstrap distribution of some asymptotically normally distributed statistics attains the same precision of approximation to their exact distribution as their Edgeworth expansions of certain order. Since U is also asymptotically normally distributed, it naturally raises a question if the distribution of U also has the same property. We show in 4.2.1 that bootstrap distribution of U can uniformly approximate $F(z)$ asymptotically as good as the Edgeworth expansions shown in Theorems 1 and 2 and thus it gives a better approximation than the normal. Theorem 5 below, the main result of this chapter, describes this situation. We also state a possibility of bootstrap distribution for the studentized statistics analogous to Theorem 5 in 4.2.2. We report Mote Carlo results on bootstrap and Edgeworth approximation for both unstudentized and studentized statistics based on a Tobit model in Chapter 5.

4.2.1 Unstudentized statistics

We first give some notations and definitions to describe the theorem. We consider bootstrapped U based on the bootstrap sample (X_i^{*T}, Y_i^*), $i=1,\ldots,n$,

$$U^* = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} U_{ij}^* \quad ,$$

where

$$U_{ij}^* = K_{ij}^{*\prime}(Y_i^* - Y_j^*)\, h^{-d-1} \quad , \quad K_{ij}^{*\prime} = K'\left(\frac{X_i^* - X_j^*}{h}\right) \quad .$$

Define bootstrap distribution of Z, $F^*(z) = P^*(Z^* \le z)$,

where for a $d \times 1$ vector ν

$$Z^* = n^{1/2}\sigma^{*-1}\nu^{\tau}(U^* - \bar{\mu}^*) \quad ,$$

$$\sigma^{*2} = E^*[\,2\nu^{\tau}(U_1^* - E^*U^*)\,]^2 \quad ,$$

$$U_i^* = E^*(U_{ij}^* \mid i^*) = \frac{1}{n}\sum_{j=1}^{n} h^{-d-1}K'\left(\frac{X_i^* - X_j}{h}\right)(Y_i^* - Y_j) \quad .$$

$$\nu^{\tau}\bar{\mu}^* = E^*\nu^{\tau}U^* - h^{L}\sigma^*\tilde{\kappa}_1 = \frac{n-1}{n}\nu^{\tau}U - h^{L}\sigma^*\tilde{\kappa}_1$$

where $P^*(\cdot)$ is the conditional probability measure given $(X_1^T, Y_1), \cdots,$ (X_n^T, Y_n), $E^*(\cdot)$ is the conditional expectation given $(X_1^T, Y_1), \cdots, (X_n^T, Y_n)$ and $E^*(\cdot | i_1^*, \cdots, i_r^*) = E^*\{\cdot | (X_i^{*T}, Y_i^*), i = i_1, \cdots, i_r\}$. Note

$$E^*U^* = \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E^*(U_{ij}^*) = E^*(U_{12}^*) = h^{-d-1}E^*\left\{K_{ij}^*(Y_i^* - Y_j^*)\right\}$$

$$= \frac{1}{n^2h^{d+1}}\sum_{i=1}^{n}\sum_{j=1}^{n}K_{ij}(Y_i - Y_j) = \frac{n-1}{n}U \quad .$$

We will prove the following theorem.

THEOREM 5

Under (i)", (ii), (iii), (iv)', (v)', (vi)', (vii), (viii), (ix)' and (x)-(xii), as $n \to \infty$,

$$\sup_{\nu:\, \nu^{\tau}\nu = 1}\ \sup_{z \in R} |F(z) - F^*(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \quad \text{a.s.}$$

In view of the theorem of Robinson (1995a) and Theorems 2 and 5, we easily see that $F^*(z)$ and $\tilde{\tilde{F}}(z)$ approximate $F(z)$ equally well and better than $\Phi(z)$. Therefore we could modify interval estimates or critical region of a test for the parameters of interest based on normal approximation using $F^*(z)$ or $\tilde{\tilde{F}}(z)$. We have described how to do it using Edgeworth expansions in 1.3.1, where we have also mentioned the problem of studentization. When we would like to implement, say, a modified interval estimation based on the approximation to $F(z)$ by $F^*(z)$ based on Theorem 5, we have two problems. Firstly, we are concerned with the unstudentized statistics Z so that interval estimate based on it involves the unknown variance. Secondly, $F^*(z)$ is unknown so that we need to estimate it. One way of the estimation is that we resample repeatedly from the original sample and calculate U^* from each resample and construct an empirical distribution based on the accumulated U^* (see 5.1.2 for detail). In order for empirical applications, we need to establish a theorem with respect to the studentized statistics $\hat{Z}$. We state a conjecture for bootstrap distribution of $\hat{Z}$ in the next subsection. We are currently working on the bootstrap distribution of $\hat{Z}$ to analytically prove a studentized version of Theorem 5, where we have found Theorem 5 plays an important role as Theorem 1 does in the proof of Theorem 3. Therefore, Theorem 5 is still worth establishing despite its infeasibility, in addition that this itself is of theoretical interest.

We need somehow stronger assumptions than the previous theorems. (i)" is a moment condition which is required for a sufficiently fast convergence rate of σ^{*2} (Lemma 34-(d)). (iv)' and (v)' are necessary for $\bar{\kappa}_1 \to \kappa_1$ which are assumed in Theorem 2 and 4 also. (vi)' and (ix)' are necessary in the proof of Lemma 36 and Lemma 34-(e) respectively.

PROOF OF THEOREM 5

It suffices to show

$$\sup_{v:\, v^T v=1} \sup_{z \in R} |P^*(Z^* \le z) - \tilde{F}(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \quad \text{a.s.}$$

in view of Theorem 1.

The qualification "almost surely" or "almost surely for n sufficiently large" will be omitted. Let

$$V_i^* = \sigma^{*-1}\nu^\tau(U_i^*-E^*U^*) \quad , \quad W_{ij}^* = \sigma^{*-1}\nu^\tau(U_{ij}^*-E^*U^*)-V_i^*-V_j^* \quad ,$$

$$\Delta^* = n^{1/2}h^L\tilde{\kappa}_1 \quad .$$

Z^* can be decomposed as follows.

$$Z^* = \frac{2}{n^{1/2}}\sum_{i=1}^{n}V_i^* + n^{1/2}\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}W_{ij}^* + \Delta^*$$

$$= \quad \bar{V}^* \quad + \quad \bar{W}^* \quad + \quad \Delta^* \ .$$

Define

$$\chi^*(t) = E^*(e^{itZ^*}) = e^{it\Delta^*}E^*\left[e^{it(\bar{V}^*+\bar{W}^*)}\right] \ . \tag{4.1}$$

By Esseen's smoothing lemma,

$$\sup_z |F^*(z)-\tilde{F}(z)| \le \int_{-n^{1/2}logn}^{n^{1/2}logn}\left|\frac{\chi^*(t)-\tilde{\chi}(t)}{t}\right|dt + O\left(\frac{1}{n^{1/2}logn}\right)$$

$$\le \int_{-p_2}^{p_2}\left|\frac{\chi^*(t)-\tilde{\chi}(t)}{t}\right|dt + \int_{p_2\le|t|\le n^{1/2}logn}\left|\frac{\chi^*(t)}{t}\right|dt$$

$$+ \int_{|t|\ge p_2}\left|\frac{\tilde{\chi}(t)}{t}\right|dt + o(n^{-1/2})$$

$$= \text{(I-3)} + \text{(II-3)} + \text{(III-3)} + o(n^{-1/2}) \ , \tag{4.2}$$

where p_2=min $(\bar{\epsilon}n^{1/2}, \log\ n)$. $\bar{\epsilon}$ can be chosen such that

$$0<\bar{\epsilon}\le\epsilon^* \equiv (E^*|2V_1^*|^3)^{-1} \le \{E^*(2V_1^{*2})\}^{-3/2}=1$$

almost surely for sufficiently large n by Jensen's inequality and Lemma 34-(a). It was shown in the proof of Theorem 1 that

$$\text{(III-3)} = o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^L) \ . \tag{4.3}$$

To evaluate (I-3), we proceed by representing $\chi^*(t)$ as $\tilde{\chi}(t)$ plus a remainder. Since $\Delta^* = n^{1/2}h^L\tilde{\kappa}_1 = n^{1/2}h^L\kappa_1 + o(n^{1/2}h^L)$ due to $\tilde{\kappa}_1 \to \kappa_1$ a.s.

(see the proof of Theorem 2), using (2.6),

$$e^{it\Delta^*} = 1 + it\Delta^* + (e^{it\Delta^*} - 1 - it\Delta^*) = 1 + it\Delta^* + O((t\Delta^*)^2)$$

$$= 1 + itn^{1/2}h^L\kappa_1 + o(n^{1/2}h^L(|t| + t^2)), \quad (4.4)$$

and

$$E^*\{e^{it(\bar{V}^* + \bar{W}^*)}\} = E^*\left[e^{it\bar{V}^*}\left\{1 + it\bar{W}^* + \frac{1}{2}(it\bar{W}^*)^2\right\}\right]$$

$$+ E^*\left[e^{it\bar{V}^*}\left\{e^{it\bar{W}^*} - 1 - it\bar{W}^* - \frac{1}{2}(it\bar{W}^*)^2\right\}\right]$$

$$= E^*\left[e^{it\bar{V}^*}\left\{1 + it\bar{W}^* + \frac{1}{2}(it\bar{W}^*)^2\right\}\right] + O(E^*|t\bar{W}^*|^3) \quad . \quad (4.5)$$

Writing $w_j^*(t) = e^{2itn^{-1/2}V_j^*}$ and $\gamma^*(t) = E^*[w_j^*(t)]$,

$$E^*(e^{it\bar{V}^*}) = \gamma^*(t)^n \quad (4.6)$$

$$E^*(e^{it\bar{V}^*}\bar{W}^*) = n^{1/2}\gamma^*(t)^{n-2}E^*[W_{12}^* w_1^*(t) w_2^*(t)]$$

$$= \{\gamma^*(t)\}^{n-2}\left\{\frac{4(it)^2}{n^{1/2}}E(W_{12}V_1V_2) + o\left(\frac{t^2}{n}\right) + O\left(\left(\frac{|t|^3}{n} + \frac{t^4}{n^{3/2}}\right)h^{-\frac{2}{3}d-1}\right)\right\} \quad (4.7)$$

by Lemma 38, and

$$E^*(e^{it\bar{V}^*}\bar{W}^{*2}) = n\binom{n}{2}^{-2}\binom{n}{2}\{\gamma^*(t)\}^{n-2}E^*[W_{12}^{*2}w_1^*(t)w_2^*(t)]$$

$$+ 6n\binom{n}{2}^{-2}\binom{n}{3}\{\gamma^*(t)\}^{n-3}E^*[W_{12}^*W_{13}^*w_1^*(t)w_2^*(t)w_3^*(t)]$$

$$+ 6n\binom{n}{2}^{-2}\binom{n}{4}\{\gamma^*(t)\}^{n-4}E^*[W_{12}^*w_1^*(t)w_2^*(t)]^2$$

$$= \frac{2}{n-1}\{\gamma^*(t)\}^{n-2}\left\{h^{-d-2}\kappa_2 + o(h^{-d-2}) + O\left(|t|n^{-\frac{1}{2}}h^{-\frac{4}{3}d-2}\right)\right\}$$

$$+ \{\gamma^*(t)\}^{n-3}O\left(|t|^3n^{-\frac{3}{2}}h^{-\frac{4}{3}d-2}\right)$$

$$+ \{\gamma^*(t)\}^{n-4}O\left(t^4n^{-1} + t^6n^{-2}h^{-\frac{4}{3}d-2} + t^8n^{-3}h^{-\frac{4}{3}d-2}\right) \quad (4.8)$$

by Lemmas 39-41. Since $\bar{\epsilon}$ can be chosen arbitrarily small, for $|t| \le \bar{\epsilon}n^{1/2}$ and $m=0,1,2,3,4$,

$$\{\gamma^*(t)\}^{n-m} = e^{-\frac{t^2}{2}}\left\{1 + \frac{E^*(2V_1^*)^3}{6n^{1/2}}(it)^3\right\} + o\left(n^{-1/2}(|t|^3 + t^6)e^{-\frac{t^2}{4}}\right)$$

$$= e^{-\frac{t^2}{2}}\left\{1 + \frac{E(2v_1)^3}{6n^{1/2}}(it)^3\right\} + o\left(n^{-1/2}(|t|^3 + t^6)e^{-\frac{t^2}{4}}\right) \quad . \quad (4.9)$$

The first equality is similarly proved as Theorem 1, Section 41 of Gnedenko and Kolmogorov (1954) using Lemma 34-(a), (e), while the second equality is because

$$E^*(2V_1^*)^3 = \frac{8}{n\sigma^{*3}}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} \nu^\tau U_{ij} - \frac{n-1}{n}\nu^\tau U\Big)^3 = \frac{8(n-1)^3}{n^4\sigma^{*3}}\sum_{i=1}^{n}\bar{V}_i^3 \to 8E(\nu_1^3)$$

by Lemma 34-(d) and (2.29). By (4.1), (4.4)-(4.9), and Lemma 42,

$$\begin{aligned}\chi^*(t) &= \Big[1 + itn^{1/2}h^L\kappa_1 + O(t^2nh^{2L})\Big]\\ &\times\Big[e^{-\frac{t^2}{2}}\Big\{1+\frac{4E(\nu_1^3)}{3n^{1/2}}(it)^3\Big\} + o\Big(n^{-1/2}(|t|^3+t^6)e^{-\frac{t^2}{4}}\Big)\Big]\\ &\times\Big[1 + \frac{4(it)^3}{n^{1/2}}E(W_{12}V_1V_2) + \frac{(it)^2}{nh^{d+2}}\kappa_2 + A_n''\Big] \qquad (4.10)\end{aligned}$$

where

$$\begin{aligned}A_n'' &= O\Bigg(\frac{t^4}{nh^{\frac{2}{3}d+1}} + \frac{|t|^3}{n^{\frac{3}{2}}h^{\frac{4}{3}d+2}} + \frac{t^2}{n^2h^{d+2}} + \frac{|t|^5}{n^{\frac{3}{2}}h^{\frac{4}{3}d+2}}\\ &\qquad + \frac{t^6}{n} + \frac{t^{10}}{n^3h^{\frac{4}{3}d+2}} + \frac{t^8}{n^2h^{\frac{4}{3}d+2}} + \frac{|t|^3}{(nh^{d+2})^{\frac{3}{2}}}\Bigg) + o\Big(\frac{t^2}{nh^{d+2}}\Big)\\ &= o\Big(\frac{t^2+t^{10}}{nh^{d+2}}\Big) \text{ by (ix)'.}\end{aligned}$$

Expanding the right hand side of (4.10),

$$\begin{aligned}\chi^*(t) &= e^{-\frac{t^2}{2}}\Big\{1 + itn^{1/2}h^L\kappa_1 + \frac{\kappa_2}{nh^{d+2}}(it)^2\\ &\qquad + \frac{4E(\nu_1^3)+12E(W_{12}V_1V_2)}{3n^{1/2}}(it)^3\Big\}\\ &+ o\Big(\frac{1}{n^{1/2}}e^{-\frac{t^2}{4}}(|t|^3+|t|^{11})\Big) + o\Big(\frac{1}{nh^{d+2}}e^{-\frac{t^2}{4}}(|t|^3+t^4)\Big)\\ &+ O\Big(nh^{2L}|t|^2e^{-\frac{t^2}{2}}\Big) + O\Big(A_n''(e^{-\frac{t^2}{4}}+e^{-\frac{t^2}{2}})(1+t^8)\Big). \qquad (4.11)\end{aligned}$$

Lemmas 13 and 14(b) give $E(\nu_1^3) + 3E(W_{12}V_1V_2) = \kappa_3 + 3\kappa_4 + o(1)$ so that the first term on the right of (4.11) is $\tilde{\chi}(t) + o(e^{-\frac{t^2}{2}}n^{-1/2}|t|^3)$. Thus

$$(\mathrm{I}\text{-}3) = \int_{-p}^{p} \left| \frac{\chi^*(t) - \tilde{\chi}(t)}{t} \right| dt = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \ . \qquad (4.12)$$

To estimate (II-3), define $\bar{V}^*(m) = \frac{2}{n^{1/2}} \sum_{i=1}^{m} V_i^*$ for $m = 1, \dots, n$,

and for $m = 1, \dots, n\text{-}1$

$$\bar{W}^*(m) = n^{1/2} \binom{n}{2}^{-1} \sum_{i=1}^{m} \sum_{j=i+1}^{n} W_{ij}^* \ .$$

Note that $\bar{V}^*(n) = \bar{V}^*$ and $\bar{W}^*(n-1) = \bar{W}^*$. Using (2.6),

$$|\chi^*(t)| = |e^{it\Delta^*} E^* e^{it(\bar{V}^* + \bar{W}^*)}| \le |E^* e^{it(\bar{V}^* + \bar{W}^*)}|$$

$$\le \left| E^* e^{it\{\bar{V}^* + (\bar{W}^* - \bar{W}^*(m))\}} \left\{ 1 + it\bar{W}^*(m) + \frac{(it)^2}{2} \bar{W}^*(m)^2 \right\} \right|$$

$$+ O(|t|^3 E^* |\bar{W}^*(m)|^3)$$

$$\le |E^* e^{it\{\bar{V}^* + (\bar{W}^* - \bar{W}^*(m))\}}| + |t| \, |E^* e^{it\{\bar{V}^* + (\bar{W}^* - \bar{W}^*(m))\}} \bar{W}^*(m)|$$

$$+ \frac{t^2}{2} |E^* e^{it\{\bar{V}^* + (\bar{W}^* - \bar{W}^*(m))\}} \bar{W}^*(m)^2| + O(|t|^3 E^* |\bar{W}^*(m)|^3) \ . \qquad (4.13)$$

By (iii) and Lemma 35, the first term on the right of (4.13) can be bounded by

$$|E^* e^{it\bar{V}^*(m)} E^* e^{it\{(\bar{V}^* - \bar{V}^*(m)) + (\bar{W}^* - \bar{W}^*(m))\}}|$$

$$\le |E^* e^{it\bar{V}^*(m)}| = |\gamma^*(t)|^m \ . \qquad (4.14)$$

Using (iii), Lemmas 35, 37-(a), and (A.5), the second term on the right of (4.13) can be bounded by $|t|$ times

$$|E^* \{ e^{it\bar{V}^*(m)} \bar{W}^*(m) \}| \le |\gamma^*(t)|^{m-2} \frac{m(m-1)}{2} n^{1/2} \binom{n}{2}^{-1} E^* |W_{12}^*|$$

$$\le C |\gamma^*(t)|^{m-2} \frac{m^2}{n^{3/2} h} \ . \qquad (4.15)$$

Similarly to the derivation of (4.15), the third term on the right of (4.13) other than $t^2/2$ can be bounded as follows using (iii), Lemmas 35, 37 and (A.5).

$$|E^* e^{it\{\bar{V}^* + (\bar{W}^* - \bar{W}^*(m))\}} \bar{W}^*(m)^2|$$

$$\le |E^* e^{it\bar{V}^*(m)} \bar{W}^*(m)^2|$$

$$\le |\gamma^*(t)|^{m-2} (mn) n \binom{n}{2}^{-2} E^* (W_{12}^{*2})$$

$$+ |\gamma^*(t)|^{m-3} \{ 2m(n-1)n + 4(m-1)mn \} n \binom{n}{2}^{-2} E^* |W_{12}^* W_{13}^*|$$

$$+6|\gamma^*(t)|^{m-4}\frac{m(m-1)(m-2)(m-3)}{24}n\binom{n}{2}^{-2}E^*|W_{12}^*|E^*|W_{34}^*|$$

$$\le C\left\{|\gamma^*(t)|^{m-2}\frac{mn}{n^3h^{d+2}}+|\gamma^*(t)|^{m-3}\frac{mn^2+m^2n}{n^3h^2}+|\gamma^*(t)|^{m-4}\frac{m^4}{n^3h^2}\right\}$$

$$\le C|\gamma^*(t)|^{m-4}\left\{\left(\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)m+\frac{m^2}{n^2h^2}+\frac{m^4}{n^3h^2}\right\} \tag{4.16}$$

because $|\gamma^*(t)|\le 1$. Substituting (4.14)-(4.16) into (4.13), with $|\gamma^*(t)|\le 1$ yields

$$|\chi^*(t)| \le |\gamma^*(t)|^m+C\left\{|\gamma^*(t)|^{m-2}\frac{m^2}{n^{3/2}h}\right\}|t|$$

$$+C|\gamma^*(t)|^{m-4}\left\{\left(\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)m+\frac{m^2}{n^2h^2}+\frac{m^4}{n^3h^2}\right\}t^2$$

$$+O(|t|^3E^*|\bar{W}^*(m)|^3)$$

$$\le C|\gamma^*(t)|^{m-4}\left[1+\left(\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2m+\left(\frac{|t|}{n^{3/2}h}+\frac{t^2}{n^2h^2}\right)m^2+\frac{t^2m^4}{n^3h^2}\right]$$

$$+O(|t|^3E^*|\bar{W}^*(m)|^3) . \tag{4.17}$$

Now we evaluate (II-3), partitioning its range of integration into two parts, namely $p_2\le|t|\le\epsilon^*n^{1/2}$ and $\epsilon^*n^{1/2}\le|t|\le n^{1/2}\log n$.

(i) For $p_2\le|t|\le\epsilon^*n^{1/2}$,

$$|\gamma^*(t)| = |E^*e^{it\frac{2}{\sqrt{n}}V_1^*}|$$

$$\le |E^*\{e^{it\frac{2}{\sqrt{n}}V_1^*}-1-it\frac{2}{\sqrt{n}}V_1^*-\frac{1}{2!}(it\frac{2}{\sqrt{n}}V_1^*)^2\}|$$

$$+|E^*\{1+it\frac{2}{\sqrt{n}}V_1^*+\frac{1}{2!}(it\frac{2}{\sqrt{n}}V_1^*)^2\}|$$

$$\le\left|1+\frac{(it)^2}{2n}\right|+\frac{|t|^3}{6n^{3/2}}E^*|2V_1^*|^3$$

$$\le 1-\frac{t^2}{3n}\ \text{, since } |t|\le\epsilon^*n^{1/2}=\frac{n^{1/2}}{E^*|2V_1^*|^3}$$

$$\le \exp\left(-\frac{t^2}{3n}\right) , \tag{4.18}$$

where the second inequality uses (2.6) and Lemma 34-(a). Substituting (4.18) in (4.17) and applying Lemma 42,

$$|\chi^*(t)| \le \exp\left\{-\frac{(m-4)t^2}{3n}\right\}$$
$$\times\left[1+\left\{\left(\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2\right\}m+\left\{\frac{|t|}{n^{3/2}h}+\frac{t^2}{n^2h^2}\right\}m^2+\frac{t^2m^4}{n^3h^2}\right]$$
$$+O\left(|t|^3\left(\frac{m}{n^2h^{d+2}}\right)^{3/2}\right). \tag{4.19}$$

We may take $m=[9nlogn/t^2]$ since $1\le m\le n-1$ holds for $p_2\le|t|\le\epsilon^* n^{1/2}$ and sufficiently large n. Because $m\ge(9nlogn/t^2)-1$,

$$\exp\left[-\frac{(m-4)t^2}{3n}\right]=\exp\left(-\frac{(m+1)t^2}{3n}\right)\exp\left(\frac{5t^2}{3n}\right)$$
$$\le C\exp(-3logn)\le\frac{C}{n^3} \tag{4.20}$$

for $|t|\le\epsilon^* n^{1/2}$. Substituting (4.20) into (4.19) using $m\le 9nlogn/t^2$, we derive

$$|\chi^*(t)|\le\frac{C}{n^3}\left[1+\left(\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2\frac{nlogn}{t^2}+\left\{\frac{|t|}{n^{3/2}h}+\frac{t^2}{n^2h^2}\right\}\frac{n^2(logn)^2}{t^4}\right.$$
$$\left.+\frac{t^2}{n^3h^2}\frac{n^4(logn)^4}{t^8}\right]+O\left(\left(\frac{logn}{nh^{d+2}}\right)^{3/2}\right)$$
$$\le C\left[\left\{\frac{1}{n^3}+\frac{logn}{n^4h^{d+2}}+\frac{logn}{n^3h^2}+\left(\frac{logn}{nh^{d+2}}\right)^{3/2}\right\}+\frac{(logn)^2}{n^3h^2}\frac{1}{t^2}\right.$$
$$\left.+\frac{(logn)^2}{n^{5/2}h}\frac{1}{|t|^3}+\frac{(logn)^4}{n^2h^2}\frac{1}{t^6}\right]$$
$$\le C\left\{\left(\frac{logn}{nh^{d+2}}\right)^{3/2}+\frac{(logn)^2}{n^3h^2}\frac{1}{t^2}+\frac{(logn)^2}{n^{5/2}h}\frac{1}{|t|^3}+\frac{(logn)^4}{n^2h^2}\frac{1}{t^6}\right\}.$$

Therefore, dropping the range of integration $p_2\le|t|\le Cn^{1/2}$ on the right of the second inequality,

$$\int_{p_2\le|t|\le\epsilon^* n^{1/2}}\left|\frac{\chi(t)}{t}\right|dt\le\int_{p_2\le|t|\le Cn^{1/2}}\left|\frac{\chi(t)}{t}\right|dt$$
$$\le C\left\{\left(\frac{\log n}{nh^{d+2}}\right)^{3/2}\int\frac{dt}{|t|}+\frac{(\log n)^2}{n^3h^2}\int\frac{dt}{|t|^3}\right.$$
$$\left.+\frac{(\log n)^2}{n^{5/2}h}\int\frac{dt}{t^4}+\frac{(\log n)^4}{n^2h^2}\int\frac{dt}{|t|^7}\right\}$$

$$\leq C\left\{\frac{(\log n)^{5/2}}{(nh^{d+2})^{3/2}}+\frac{1}{n^3h^2}+\frac{1}{n^{5/2}h\log n}+\frac{1}{n^2h^2(\log n)^2}\right\}$$

$$=o(n^{-1/2}+n^{-1}h^{-d-2}) \tag{4.21}$$

by assumption (ix)'.

(ii) For $\epsilon^* n^{1/2}\leq|t|\leq n^{1/2}\log n$, noting $\sigma^{-1}v^\tau(U_{ij}-EU)=W_{ij}+V_i+V_j$,

$$\gamma^*(t)=E^*(e^{\frac{2it}{\sqrt{n}}V_1^*})=\frac{1}{n}\sum_{k=1}^n e^{\frac{2it}{\sqrt{n}\sigma^*}v^\tau\left\{\frac{1}{nh^{d+2}}\sum_{j=1}^n K'(\frac{X_k-X_j}{h})(Y_k-Y_j)-E^*U^*\right\}}$$

$$=\frac{1}{n}\sum_{k=1}^n e^{\frac{2it}{\sqrt{n}\sigma^*}\left\{\frac{\sigma}{n}\sum_{j=1}^n(W_{kj}+V_k+V_j)+v^\tau(EU-E^*U^*)\right\}}$$

$$=\frac{1}{n}\sum_{k=1}^n e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}+\frac{1}{n}\sum_{k=1}^n e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}\frac{2it\sigma}{\sqrt{n}\sigma^*}\left\{\frac{1}{n}\sum_{j=1}^n(W_{kj}+V_j)+\sigma^{-1}v^\tau(EU-E^*U^*)\right\}$$

$$+\frac{1}{n}\sum_{k=1}^n e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}\left[e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}\{\frac{1}{n}\sum_{j=1}^n(W_{kj}+V_j)+\sigma^{-1}v^\tau(EU-E^*U^*)\}}-1\right.$$

$$\left.-\frac{2it\sigma}{\sqrt{n}\sigma^*}\{\frac{1}{n}\sum_{j=1}^n(W_{kj}+V_j)+\sigma^{-1}v^\tau(EU-E^*U^*)\}\right].$$

Writing $\gamma_v(t)=E(e^{\frac{2it}{\sqrt{n}}v_1})$ as in the proof of Theorem 1, using (2.6) and $|t|\leq n^{1/2}\log n$, we have

$$|\gamma^*(t)-\gamma_v(t)|\leq\left|\frac{1}{n}\sum_{k=1}^n e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}-E(e^{\frac{2it}{\sqrt{n}}v_1})\right|$$

$$+\left|\frac{2\sigma\log n}{n^2\sigma^*}\sum_{k=1}^n\sum_{j=1}^n e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}(W_{kj}+V_j)\right|$$

$$+\left|\frac{2\log n}{\sigma^*}v^\tau(EU-E^*U^*)\right|\times\left|\frac{1}{n}\sum_{k=1}^n e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}\right|$$

$$+\frac{C}{n}\sum_{k=1}^n\left\{\frac{2\sigma\log n}{n\sigma^*}\sum_{j=1}^n(W_{kj}+V_j)\right\}^2+C\left\{\frac{\log n}{\sigma^*}v^\tau(EU-E^*U^*)\right\}^2. \tag{4.22}$$

The first term on the right of (4.22) is bounded by

$$\left|\frac{1}{n}\sum_{k=1}^n(e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}-e^{\frac{2it}{\sqrt{n}}V_k})\right|+\left|\frac{1}{n}\sum_{k=1}^n(e^{\frac{2it}{\sqrt{n}}V_k}-e^{\frac{2it}{\sqrt{n}}v_k})\right|$$

$$+\left|\frac{1}{n}\sum_{k=1}^n\{e^{\frac{2it}{\sqrt{n}}v_k}-E(e^{\frac{2it}{\sqrt{n}}v_1})\}\right|. \tag{4.23}$$

The first term of (4.23) is smaller than

$$\frac{C|t|}{\sqrt{n}}\left|\frac{\sigma}{\sigma^*}-1\right|\left(\frac{1}{n}\sum_{k=1}^n|Y_k|+1\right)=o\left(\frac{|t|}{\sqrt{n}(\log n)^3}\right)=o((\log n)^{-2})$$

by (2.6), SLLN with (i)", and Lemmas 1-(d), 34-(d) for $|t|\leq n^{1/2}\log n$. The second term of (4.23) is smaller than

$$\frac{1}{n}\sum_{k=1}^{n}\frac{|t|}{\sqrt{n}}|V_k-v_k|\le\frac{h^{L}\log n}{n}\sum_{k=1}^{n}(|Y_k|+1)=o(n^{1/2}h^{L})$$

by (2.6), (B.1), SLLN with (i)", and Lemma 1-(d) for $|t|\le n^{1/2}\log n$. Theorem 2.6 of Feuerverger and Mureika (1977) applies to show the last term of (4.23) converges to zero for $|t|\le n^{1/2}\log n$, thus (4.23) converges to zero. The second term of (4.22) converges to zero because

$$\begin{aligned}
&E\left|\frac{\log n}{n^2}\sum_{k=1}^{n}\sum_{j=1}^{n}e^{\frac{2it}{\sqrt{n}}V_k}(W_{kj}+V_j)\right|^2\\
&=\frac{(\log n)^2}{n^4}\sum\sum_{k<j}E\{(W_{kj}+V_j)^2+(W_{jk}+V_k)^2\\
&\quad+e^{2it(V_k-V_j)}(W_{kj}+V_j)(W_{jk}+V_k)+e^{2it(V_j-V_k)}(W_{kj}+V_j)(W_{jk}+V_k)\}\\
&\quad+\text{ smaller order terms with } k=j\\
&\le\frac{C(\log n)^2}{n^2}E|W_{12}+V_2|^2\le\frac{C(\log n)^2}{n^2h^{d+2}}=O(n^{-1}(\log n)^{-4})\ ,
\end{aligned}$$

where the first equality uses Lemma 1-(c), the second inequality uses (A.5) and Lemma 1-(d), while the last equality uses (ix)'. The third and last terms of (4.22) converge to zero because of Lemma 34-(c), (d) and

$$\left|\frac{1}{n}\sum_{k=1}^{n}e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}\right|\le\frac{1}{n}\sum_{k=1}^{n}\left|e^{\frac{2it\sigma}{\sqrt{n}\sigma^*}V_k}\right|=1\ .$$

The fourth term of (4.22) converges to zero because of Lemma 34-(d) and

$$\frac{(\log n)^3}{n}\sum_{k=1}^{n}\left\{\frac{1}{n}\sum_{j=1}^{n}(W_{kj}+V_j)\right\}^2\to 0 \text{ (see the proof of Lemma 34-(d))}.$$

Therefore

$$|\gamma^*(t)-\gamma_v(t)|\to 0 \text{ a.s.}$$

so that there exists a constant $\eta_3\epsilon(0,1)$ such that

$$|\gamma^*(t)|\le 1-\eta_3\ , \tag{4.24}$$

since $|\gamma_v(t)|<1-\eta_1$ for some $\eta_1\epsilon(0,1)$ (see the proof of Theorem 1). We can choose m such that

$$m=\left[-\frac{3\log n}{\log(1-\eta_3)}\right] \tag{4.25}$$

since $1\le m\le n-1$ for sufficiently large n. Substituting (4.24) and (4.25) in (4.17) and applying Lemma 42 bounds $|\chi^*(t)|$ by

$$C(1-\eta_3)^{-\frac{3\log n}{\log(1-\eta_3)}}\left[1+\left\{\left(\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)t^2\right\}\left\{-\frac{3\log n}{\log(1-\eta_3)}\right\}\right.$$
$$\left.+\left\{\frac{|t|}{n^{3/2}h}+\frac{t^2}{n^2h^2}\right\}\left\{-\frac{3\log n}{\log(1-\eta_3)}\right\}^2+\frac{t^2}{n^3h^2}\left\{-\frac{3\log n}{\log(1-\eta_3)}\right\}^4\right]$$
$$+O\left(|t|^3\left\{-\frac{\log n}{n^2h^{d+2}\log(1-\eta_3)}\right\}^{3/2}\right). \tag{4.26}$$

Because $(1-\eta_3)^{-\frac{3\log n}{\log(1-\eta_3)}}=n^{-3}$, dividing (4.26) by $|t|$ and integrating over the range $\epsilon^*n^{1/2}\le|t|\le n^{1/2}\log n$, we obtain

$$\int_{\epsilon^*n^{1/2}\le|t|\le n^{1/2}\log n}\left|\frac{\chi(t)}{t}\right|dt$$
$$=O\left(\frac{\log(n^{1/2}\log n)}{n^3}+\frac{\log n}{n^3}\left\{\left(\frac{1}{n^2h^{d+2}}+\frac{1}{nh^2}\right)n(\log n)^2\right\}\right.$$
$$+\frac{(\log n)^2}{n^3}\left\{\frac{1}{n^{3/2}h}(n^{1/2}\log n)+\frac{n(\log n)^2}{n^2h^2}\right\}$$
$$\left.+\frac{(\log n)^4}{n^3}\frac{1}{n^3h^2}n(\log n)^2+n^{3/2}(\log n)^3\left(\frac{\log n}{n^2h^{d+2}}\right)^{3/2}\right)$$
$$=o(n^{-1/2}+n^{-1}h^{-d-2}) \tag{4.27}$$

by assumption (ix)'. Thus by (4.21) and (4.27),

$$\text{(II-3)}=o(n^{-1/2}+n^{-1}h^{-d-2}+n^{1/2}h^{L}). \tag{4.28}$$

Substituting (4.3), (4.12) and (4.28) into (4.2) gives the required result. □

4.2.2. Studentized statistics

Theorem 5 proves that the bootstrap distribution $F^*(z)$ can approximate $F(z)$ as good as the Edgeworth expansions in Theorems 1 and 2. It is theoretically an interesting result, however it does not suit for empirical use. We would like to obtain an equivalent result for studentized statistics $\hat{Z}$ analyzed in Chapter 3 for the sake of practical application. In view of the above Theorems, we can naturally conjecture the following.

Let $\hat{F}^*(z) = P^*(\hat{Z}^* \le z)$, (4.29)

where for a $d \times 1$ vector ν

$$\hat{Z}^* = n^{1/2}\hat{\sigma}^{*-1}\nu^{\tau}(U^* - \bar{\mu}^*) \quad , \quad \hat{\sigma}^{*2} = \nu^{\tau}\hat{\Sigma}^*\nu \ ,$$

$$\hat{\Sigma}^* = \frac{4}{(n-1)(n-2)^2}\sum_{i=1}^{n}\left\{\sum_{j \ne i}^{n}(U_{ij}^* - U^*)\right\}\left\{\sum_{k \ne i}^{n}(U_{ik}^* - U^*)^{\tau}\right\}$$

CONJECTURE

Under certain conditions including assumptions in Theorem 5, we conjecture

$$\sup_{\nu:\ \nu^{\tau}\nu=1}\ \sup_{z \in R} |\hat{F}(z) - \hat{F}^*(z)| = o(n^{-1/2} + n^{-1}h^{-d-2} + n^{1/2}h^{L}) \quad \text{a.s.} \quad (4.30)$$

The proof for this is currently under way and it seems to entail existence of higher order moments than in Theorem 5. Theorem 5 seems to play an important role to establish (4.30) as Theorem 1 does in the proof of Theorem 3. We give Monte Carlo results based on (4.30) as well as Theorems 1-5 in the next chapter, which give an encouraging support to this conjecture. It is interesting that we found the bootstrap distribution seems to be even better than the Edgeworth expansions in approximating the exact distribution in the simulation study.

Chapter 5

A Monte Carlo Study [1]

This chapter presents results from a Monte Carlo study for a Tobit model based on the Theorems shown in the previous chapters. The model and kernel choice of the estimator are described in Section 5.1. We compare empirical distributions of the density-weighted AD for various choices of sample size, bandwidth and kernel order with the normal distribution, the empirical and theoretical Edgeworth expansions in Section 5.2 as well as the bootstrap distributions in Section 5.3. We compare them for both unstudentized and studentized statistics based on Theorems 1-5 and the Conjecture in Chapter 4. Roughly speaking, we find that the empirical Edgeworth expansion and bootstrap distribution approximate the empirical distribution quite well, and better than the normal in many cases. We also implement confidence interval estimation using Cornish-Fisher expansion explained in 1.3.1 (see (1.86)), where we find naturally that the interval estimate is good when the corresponding Edgeworth expansion performs well.

5.1 Model and estimator

We report results from a Monte Carlo study for the Tobit model $Y_i = (\beta^\tau X_i + \epsilon_i)\, I(\beta^\tau X_i + \epsilon_i \geq 0)$ where $X_i = (X_{1i}, X_{2i})^\tau$ is bivariate. We took $(X_i^\tau, \epsilon_i) \sim N(0, I_3)$ so that $g(x) = \beta^\tau x\{1 - \Phi(-\beta^\tau x)\} + \phi(-\beta^\tau x)$ and $\bar{\mu} = -\beta/(8\pi)$. We took $\beta = (1, 1)'$. There is no closed form formula for $\Sigma, \kappa_1, \kappa_2, \kappa_3, \kappa_4$, the first being needed in the expansions of Theorems 1 and 2, and the last four in the expansions of Theorems 1 and 3, so they were calculated

[1] Some figures in this chapter are included in Nishiyama and Robinson (2000) and discussed there.

by simulation, with 100,000 replications, to be

$$\Sigma = 10^{-5}\begin{pmatrix} 887 & 458 \\ 458 & 887 \end{pmatrix}, \ \kappa_1 = 0.397 \ , \ \kappa_2 = 1.724 \ , \ \kappa_3 = -0.144 \ , \ \kappa_4 = -0.266 \ ,$$

for example $\Sigma = 10^{-5}\sum_{i=1}^{10^5} 4\{\mu(X_i, Y_i) - \bar{\mu}\}\{\mu(X_i, Y_i) - \bar{\mu}\}^{\tau}$ where (X_i^T, Y_i), $i = 1, \ldots, 10^5$ are generated independently and identically following the above Tobit model. We employed three values of L, L=4, 8 and 10 which respectively correspond to the cases I, II and III in Section 3.3 (and easily satisfy assumptions (iv), (iv)', (v) and (v)'), using normal density-based multiplicative L-th order bivariate kernel functions proposed in Robinson (1988b), $K(u_1, u_2) = K_L(u_1)K_L(u_2)$, where

$$K_L(u) = \sum_{j=0}^{(L-2)/2} c_j u^{2j}\phi(u) \ ,$$

such that

$$\sum_{j=0}^{(L-2)/2} c_j m_{2(i+j)} = \delta_{i0}, \ i = 0, 1, \cdots, (L-2)/2 \ , \qquad (5.1)$$

$$m_{2j} = \int u^{2j}\phi(u)\,du \ ,$$

and δ_{i0} is Kronecker's delta. The values of c_j calculated from these simultaneous equations are in Table 5.1. We chose $H(u_1, u_2) = \phi(u_1)\phi(u_2)$ in estimation of κ_1 in the empirical Edgeworth expansions. We considered inference on the two elements of $\bar{\mu}$ individually, but since the results for these are very similar we report them for the first only.

TABLE 5.1

L-th order kernel functions.

L	c_0	c_1	c_2	c_3	c_4
4	1.5	-0.5	-	-	-
8	2.185	-2.185	0.4375	-0.02083	-
10	1.924	-1.347	0.1230	0.00698	-0.000489

5.2 Edgeworth approximation

Figures 1-21 compare approximations to the distribution of the unstudentized statistic $(U_{(1)} - \bar{\mu}_{(1)})/\sigma$, where $U_{(1)}$ and $\bar{\mu}_{(1)}$ are the first elements of U, $\bar{\mu}$, and $\sigma^2 = 0.00887$. We used $h = 1$, 0.8, 0.6 and 0.4 for $n=100$, and $h=0.8$, 0.6 and 0.4 for $n=400$, with 600 replications, and we set $b=1.2h$ following the discussion in Section 2.1. U and $\tilde{\kappa}_i$ involved in Figures 1-7, 8-14, and 15-21 used respectively kernel functions of orders $L=4$, 8, and 10; see (5.1) and Table 5.1. The solid line is the empirical distribution function of Z, while the dotted, broken, and broken-and-dotted lines are the standard normal distribution function Φ, the empirical Edgeworth expansion (Theorem 2), and the theoretical Edgeworth expansion (Theorem 1) respectively. The empirical Edgeworth correction results from averaging $\tilde{\kappa}_i$ across 600 replications for each sample size, bandwidth choice and kernel order. The two empirical Edgeworth expansions in each Figure involve respectively all three correction terms (shorter broken line) and one correction term of order $n^{-1/2}$ (longer broken line) in (2.27), which corresponds to the feasible version of (3.46). We examine the "one-term" case because this is the one we would hope to be able to recommend, since it involves just the "parametric" $n^{-1/2}$ correction and, depending only on κ_3 and κ_4 but not on κ_1 and κ_2, is free of K.

We first compare the "three-term" empirical Edgeworth expansion (EE3) with the empirical distribution (ED) and the normal approximation (N), finding a range of n and h where EE3 well approximates ED, and better than N, for example, see Figures 1, 2, 3, and 4. It emerges that $h=1.0$ (Figure 1) is too large in that neither N nor EE3 performs well, but when $h=0.8$ or 0.6 (Figures 2, 3) EE3 is satisfactory, and better than N, whereas when $h=0.4$ (Figure 4), the opposite outcome is observed. It is not surprising that N sometimes outperforms EE3 since n is finite (see Hall (1992), p.45) and the $\tilde{\kappa}_i$ are subject to sampling error. We also considered, but have not included, the case $h=0.1$ with $n=100$, where the variance in the empirical distribution is very large, and both N and EE3 performed poorly. Neither N nor EE3 could be expected to work well for sufficiently large or small h. Comparing Figure 6 with Figures 2, 3, say, EE3 appears to improve with increasing n. Making broad comparisons across the three groups of figures, 1-7,8-14, and 15-21, we find that bias tends to vary inversely with L, keeping n and h fixed. This is consistent with the theoretical (asymptotic) bias-reducing properties motivating higher-order kernels, but Monte Carlo studies of semiparametric estimates employing such kernels (see e.g. Robinson (1988b)) have found that these properties are not necessarily mirrored in finite samples, so these results of ours are rather pleasing.

It might then come as something as a surprise that in most cases the figures reveal that EE3 approximates ED better than the "three-term" theoretical Edgeworth expansion (TE3). A possible explanation is as follows. The proof of Theorem 1 (see (2.15)) implies that an alternative theoretical Edgeworth approximation to (2.1) is

$$\Phi(z)-\phi(z)\left\{n^{1/2}\sigma_v^{-1}v^{\tau}(EU-\bar{\mu})+\frac{E(W_{12}^2)}{n}z+\frac{4E(v_1^3)+12E(W_{12}v_1v_2)}{3n^{1/2}}(z^2-1)\right\}. \tag{5.2}$$

The expectations are untidy, depending on n so the proof goes on to obtain the simpler and more elegant $\tilde{F}(z)$, involving the n-free κ_i. However, in comparing (5.2) with the $\tilde{\kappa}_i$ EE3 might seem to most directly estimate (5.2), which might be a more accurate approximation to ED than TE3, (2.1).

Among the three cases in Section 3.3, $L=10$, satisfying the condition $L>2(d+2)$, corresponds to the case III there, when EE1 is valid if $nh^L+n^{-1/2}h^{-d-2}=o(1)$. Comparing shorter broken and longer broken lines, the "one-term" empirical Edgeworth expansion (EE1) is better than EE3 when $(n,h,L)=(100,0.8,10)$ and $(400,0.8,10)$ (Figures 16 and 19). It is interesting that for other values of L, EE1 is slightly better for some values of h depending on n than EE3 in particular when $(n,h,L)=(100,0.6, 4)$ and $(400,0.4,4)$ (Figures 3 and 7).

These are the cases of relatively small h, so that the bias is small but $n^{-1}h^{-d-2}$ is relatively large, namely the $\tilde{\kappa}_1$ correction is negligible but the $\tilde{\kappa}_2$ one tends to be too large, having the effect of pushing the curve up and down around -1 and 1 respectively. It is clear from the discussion in Section 3.3 that one expects the choice of h to be especially crucial where "one-term" expansions are concerned.

Figures 22-45 compare approximations to the distribution of studentized statistics $n^{1/2}(U_{(1)} - \bar{\mu}_{(1)})/\hat{\sigma}$, where $\hat{\sigma}^2$ is the leading element of $\hat{\Sigma}$, based on Theorem 4. U, $\tilde{\kappa}_i$, and $\hat{\Sigma}$ involved in Figures 22-29, 30-37, and 38-45 used respectively kernel functions of orders $L=4$, 8, and 10 and we took $h=0.2$, 0.4, 0.6, 0.8 for each of $n=100$, 400 with 600 replications. Because the theoretical Edgeworth expansions (Theorem 3) performed less well than in the unstudentized cases featured in Figures 1-21, and because they are in any case of less practical interest than empirical expansions, we exclude the former cases from Figures 22-45 for ease of reading.

Generally in Figures 22-29, we observe that EE3 approximates ED very well except for largish h (see e.g. Figure 26), where N also performs poorly. Comparing Figures 22-29 with Figures 1-7 for the unstudentized statistic (with $L=4$ throughout), EE3 is seen to work better for the studentized statistic. The reason may be similar to the one we offered for the apparent superiority of EE3 over TE3 in the unstudentized case, namely, $Var(U_1)$ can better normalize U than Σ, and $\hat{\Sigma}$, in view of its construction, more directly estimate $Var(U_1)$.

We proposed optimal bandwidth choices which minimize the error of the normal approximation in Section 3.3. (3.49), (3.51) with $L=4$ and κ_i described above yield the optimal bandwidth as $h=0.485$ and 0.374 for $n=100$ and 400 respectively. ED with these values of h, as well as $h=0.2$ and 0.6, is compared in Figures 46 and 47 with N, which seems to best approximate ED with optimal h.

We next consider interval estimation. We have discussed in 1.3.1 that interval estimates based on normal approximation can be modified using Edgeworth expansions. A $100(1-\alpha)\%$ confidence interval based on N is, like (1.76),

$$\left(U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} z_{\frac{\alpha}{2}},\ U_{(1)} + \frac{\hat{\sigma}}{n^{1/2}} z_{\frac{\alpha}{2}} \right), \tag{5.3}$$

where z_γ satisfies $\int_{z_\gamma}^{\infty} \phi(z)\,dz = \gamma$. We can correct this interval using Theorem 4. Inverting the empirical Edgeworth expansion there, we have the Cornish-Fisher expansion (see (1.86)),

$$w_\gamma = z_\gamma + n^{1/2} h^L \tilde{\kappa}_1 - \frac{\tilde{\kappa}_2}{n h^{d+2}} z_\gamma - \frac{4}{3n^{1/2}} \{ (2z_\gamma^2 + 1)\tilde{\kappa}_3 + 3(z_\gamma^2 + 1)\tilde{\kappa}_4 \}$$
$$+ o(n^{1/2} h^L + n^{-1} h^{-d-2} + n^{-1/2})$$
$$= \tilde{w}_\gamma + o(n^{-1/2} + n^{1/2} h^L + n^{-1} h^{-d-2}),$$

where w_γ is the $100\gamma\%$ quantile of the sampling distribution. Then the corrected interval estimate is, similarly to (1.87),

$$\left(U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} \tilde{w}_{1-\frac{\alpha}{2}},\ U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} \tilde{w}_{\frac{\alpha}{2}} \right). \tag{5.4}$$

Note that $\tilde{w}_{1-\alpha/2} \neq -\tilde{w}_{\alpha/2}$ in general so that (5.4) is not symmetric around the point estimate $U_{(1)}$, unlike (5.3). According to our interval estimation in the current Tobit example, this correction is supported when the Edgeworth expansion approximates well the empirical distribution function, which is mostly the case for the studentized statistic. We report two typical cases where the correction appears effective. One is when N fails to well approximate ED due to the large bias of U, and the other is when $\hat{Z}$ has variance significantly less than unity. Figures 48-51 show the "true" 80% and 90% confidence intervals derived from ED (solid line), the corresponding interval estimates obtained from N, see (5.3) (dotted line) and from EE3, see (5.4) (broken line) for $(n,h,L)=(100,0.6,4)$, $(400,0.2,4)$. The vertical closely-spaced dotted line indicates the true parameter value $\bar{\mu}_{(1)} = -1/(8\pi)$. The "true" interval is derived like (5.3) or (5.4) as

$$\left(U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} t_{1-\frac{\alpha}{2}},\ U_{(1)} - \frac{\hat{\sigma}}{n^{1/2}} t_{\frac{\alpha}{2}} \right), \tag{5.5}$$

where t_γ denotes the $100\gamma\%$ quantile of ED. This is due to $P(t_{\frac{\alpha}{2}} \leq \hat{Z} \leq t_{1-\frac{\alpha}{2}}) = 1-\alpha$. Both estimates (5.3) and (5.4) include the true value in all four figures. In Figures 48 and 50, we observe that they are of similar length, though (5.3) is typically biased to the right and it does not cover the left part of the "true" interval, while (5.4) covers almost the whole true interval. In Figures 49 and 51, we observe that (5.3) clearly overestimates (5.5), while (5.4) performs satisfactorily. When $(n, h, L)=(100, 0.6, 4)$, N is biased to the left (Figure 23) and when $(n, h, L)=(400, 0.2, 4)$, it has larger variance than ED (Figure 29) so that (5.3) estimates the confidence interval as described. Our experiment demonstrates that the Cornish-Fisher expansion can produce better interval estimates than N.

As discussed in Section 3.3, below (3.47), Theorems 1 and 3 also imply that bias correction should have the greatest influence in improving the second order properties of U when the minimum MSE bandwidth h^* is used. In view of Theorem 2 and Lemma 11, $h^{*L}\tilde{\kappa}_1$ estimates the bias $\sigma_v^{-1}v^T(EU-\bar{\mu})$ consistently and so $\sigma_v^{-1}v^TU-h^{*L}\tilde{\kappa}_1$ is a bias-corrected estimate of $\sigma_v^{-1}v^T\bar{\mu}=-1/(8\pi\sigma)$. Table 5.2 shows the average estimates of $\sigma_v^{-1}v^T(U-\bar{\mu})$, $h^{*L}\tilde{\kappa}_1$, and $\sigma_v^{-1}v^T(U-\bar{\mu})-h^{*L}\tilde{\kappa}_1$ for each n from 600 replications when $L=4$ and the (infeasible) minimum MSE bandwidth choice of Powell and Stoker (1996) was used. The bandwidth was calculated by means of Monte Carlo simulation to be h^* = 0.9048, 0.8061 and 0.7128 for n=100, 200, and 400 respectively. We used $b=1.2h^*$ in estimating $\tilde{\kappa}_1$. Comparing the first and the third column of Table 5.2, the bias-corrected estimate is seen to perform much better than the uncorrected one, especially for n=400.

TABLE 5.2

The effect of bias correction.

n	$\sigma^{-1}(U_{(1)}-\bar{\mu}_{(1)})$	$h^{*L}\tilde{\kappa}_1$	$\sigma^{-1}(U_{(1)}-\bar{\mu}_{(1)})-h^{*L}\tilde{\kappa}_1$
100	0.0664	0.0496	0.0168
200	0.0520	0.0427	0.0093
400	0.0342	0.0338	0.0004

Powell and Stoker (1996) also proposed a feasible minimum-MSE bandwidth $\tilde{h}^*$, which depends on two user-specified parameters h_0 and τ (see (4.35), (4.38), and (4.40) of Powell and Stoker (1996)), but we did not use it. It is because, on the basis of our calculations, though both absolute bias $|E(\tilde{h}^* - h^*)|$ and MSE $E(\tilde{h}^* - h^*)^2$ were relatively insensitive to h_0 (while exhibiting some tendency to decrease in h_0), they were highly sensitive to τ and the results depend on too much on its choice.

5.3 Bootstrap approximation

We compare the Edgeworth and bootstrap approximations to ED based on Theorems 2 and 5 for the same Tobit model. We also examine the bootstrap distribution for the studentized statistics $\hat{Z}$ in (4.29) though we have not yet verified it. These Theorems and (4.30) imply that ED can be approximated by EE3 and bootstrap distribution (BD) equally well and better than N. Figures 52-72 and 73-96 show ED (solid line), N (dotted line), BD (dotted-and-broken line) and EE3 (broken line) for unstudentized statistic Z and studentized statistics $\hat{Z}$ respectively for each sample size, bandwidth and kernel order stated. We used the same combinations of (n,h,L) as we did in the previous subsection. ED and EE3 drawn there are exactly the same as those in the corresponding ones in Figures 1-45, for example, Figure 1 and Figure 52 are both for unstudentized statistics with $(n,h,L)=(100,1.0,4)$ and they share the same ED and EE3. Bootstrap distribution is a random function so we simulated it as follows.

1. We generate a Tobit sample of size n as in Section 5.2.
2. We draw a random sample of size n from the original sample in step 1, then calculate the density-weighted AD from the subsample for each (h,L).
3. We repeat step 2 600 times and obtain the density-weighted AD for each subsample.
4. We construct an empirical distribution from the 600 bootstrap estimates in

step 3.

The empirical distribution in step 4 is an estimate of bootstrap distribution.

Figures 52-58, 59-65, and 66-72, related to the unstudentized statistics, used kernel functions of orders 4, 8 and 10 respectively and we took $h=1.0$, 0.8, 0.6, and 0.4 for $n=100$ and $h=0.8$, 0.6, and 0.4 for $n=400$, while Figures 73-80, 81-88, and 89-96, related to the studentized statistics, used kernel functions of orders 4, 8 and 10 respectively and we took $h=0.8$, 0.6, 0.4, and 0.2 for $n=100$, 400. We first look at the unstudentized cases in Figures 52-72. BD should approximate ED asymptotically as good as EE3 by Theorems 2 and 5, however we find, generally, that BD and EE3 approximate ED equally well for medium values of h, but BD tends to outperform EE3 for smallish and largish h in the smaller sample $n=100$. For smallish h, see Figures 55, 62, and 69. EE3 have decreasing parts because of the correction term of order $n^{-1}h^{-d-2}$, but BD is always non-decreasing since it is a distribution function, thus BD outperforms EE3. For largish h, see Figures 52, 56, 59, 63, 66, and 70. We may explain this by the difference of bias adjustment way between BD and EE3. To clarify the effect of bias in ED, BD, and EE3, let us consider the distribution of the bias removed statistic

$$F_m(z) = P[\sigma^{-1}\lambda^T(U-EU) \le z] \ .$$

Bootstrap distribution and Edgeworth expansion corresponding to this are respectively

$$F_m^*(z) = P^*[\sigma^{*-1}\lambda^T(U^*-E^*U^*) \le z]$$

and

$$\tilde{F}_m(z) = \Phi(z) - \phi(z)\left\{\frac{\tilde{\kappa}_2}{nh^{d+2}}z + \frac{4(\tilde{\kappa}_3+3\tilde{\kappa}_4)}{n^{1/2}}(z^2-1)\right\} .$$

In view of the proofs of Theorems 2 and 5, it is easily seen that

$$sup\ |F_m(z) - F_m^*(z)| = o(n^{-1}h^{-d-2} + n^{-1/2}) \text{ a.s.}$$

and

$$sup\ |F_m(z) - \tilde{F}_m(z)| = o(n^{-1}h^{-d-2} + n^{-1/2}) \text{ a.s.}$$

The above equations imply that $F_m^*(z)$ and $\tilde{F}_m(z)$ approximate $F_m(z)$ asymptotically equally well. Now we consider the effects of inclusion of the bias term $EU-\bar{\mu}$. From the definitions of $F(z)$, $F^*(z)$ and $\tilde{F}(z)$, we have

$$F(z) = F_m(z - \sigma^{-1} v^T (EU - \bar{\mu})) \quad ,$$

$$F^*(z) = F_m^*(z - \sigma^{*-1} v^T (E^* U^* - \bar{\mu}^*)) \quad ,$$

and

$$\tilde{F}(z) = \tilde{F}_m(z) - \phi(z) n^{1/2} h^L \tilde{\kappa}_1 \quad .$$

Since $\sigma^{*-1} v^T (E^* U^* - \bar{\mu}^*) - \sigma^{-1} v^T (EU - \bar{\mu}) \to 0$ a.s. (see Lemma 34-(d) and proof of Theorem 2), it is likely that $F^*(z)$ is close to $F(z)$. Noting

$n^{1/2} h^L \tilde{\kappa}_1 - \sigma^{-1} v^T (EU - \bar{\mu}) \to 0$ a.s. (see proof of Theorem 2),

if we tried to approximate $F(z)$ similarly to $F^*(z)$, by

$$\tilde{G}(z) \equiv \tilde{F}_m(z - n^{1/2} h^L \tilde{\kappa}_1) \quad ,$$

not by $\tilde{F}(z)$, then the approximation should have been as good as that by $F^*(z)$. However, bias adjustment in $\tilde{F}(z)$ is a vertical shift of $\tilde{F}_m(z)$ by $\phi(z) n^{1/2} h^L \tilde{\kappa}_1$ so that it may approximate $F(z)$ worse than $F^*(z)$.

Figures 73-96 present ED and its approximants for the studentized statistic $\hat{Z}$. Generally we have similar observations as for unstudentized cases other than that BD does not necessarily outperforms EE3 for smallish h. The reason is that EE3 for studentized statistics do not have decreasing parts.

We found some cases when BD approximates ED amazingly well for both studentized and unstudentized cases (see e.g. Figures 57, 63, 78, 86), all of which are when $n=400$. However we also find some cases when BD shows poor approximation (see e.g. Figure 75, 76, 79, 83) compared with EE3. We found this occurs rather independently of h and L, but it occurs more when $n=100$ than when $n=400$. The reason seems to be that we sometimes obtain rather unbalanced samples in step 1 of generating the original sample especially when n is small so that bootstrap distributions estimated from these samples do not work well.

FIGURE 1 . n=100 . h=1.0 . L=4

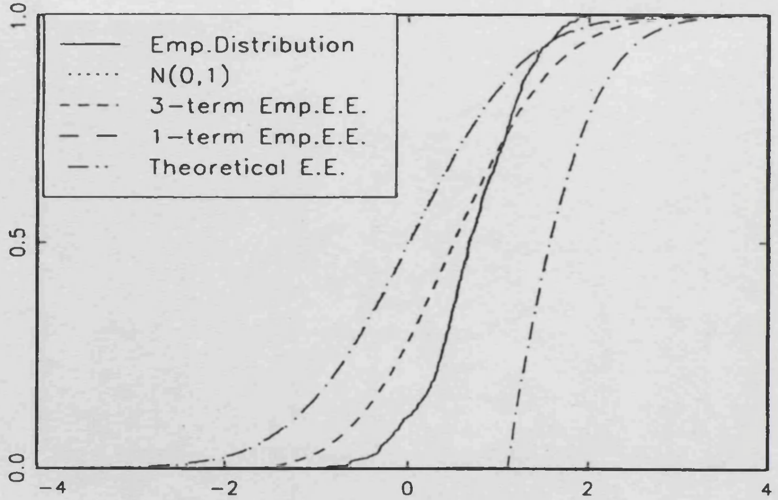


FIGURE 2 . n=100 . h=0.8 . L=4

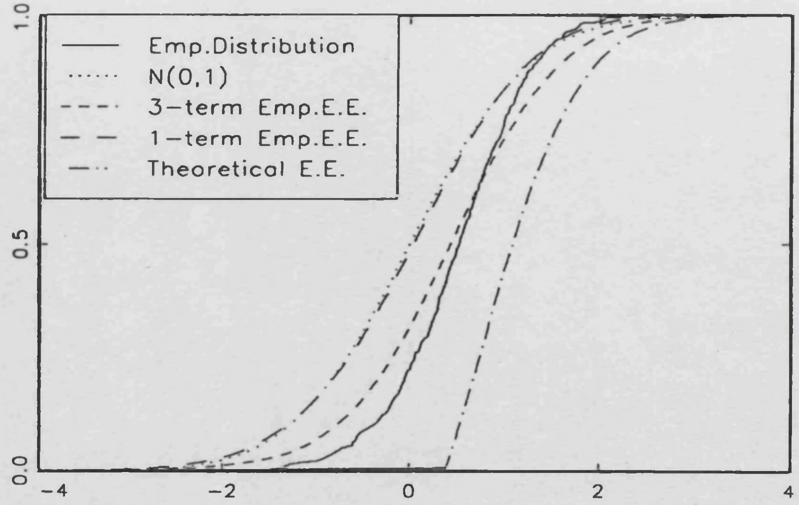


FIGURE 3 . n=100 . h=0.6 . L=4

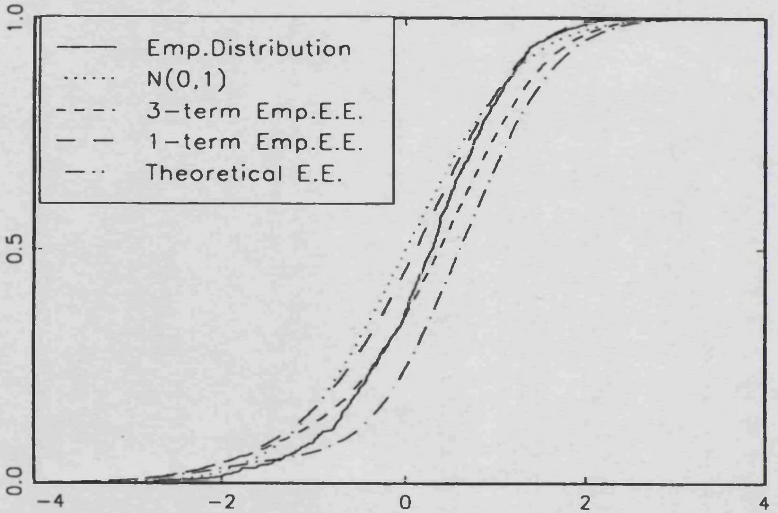


FIGURE 4 . n=100 . h=0.4 . L=4

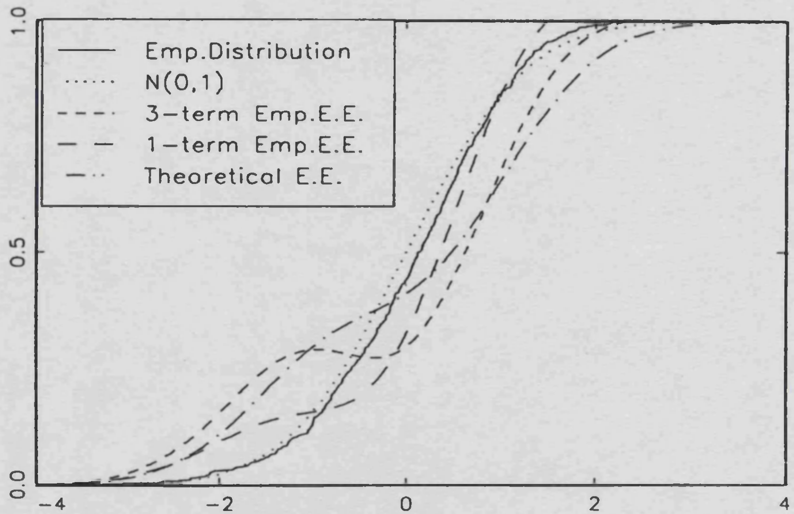


FIGURE 5 , n=400 , h=0.8 , L=4

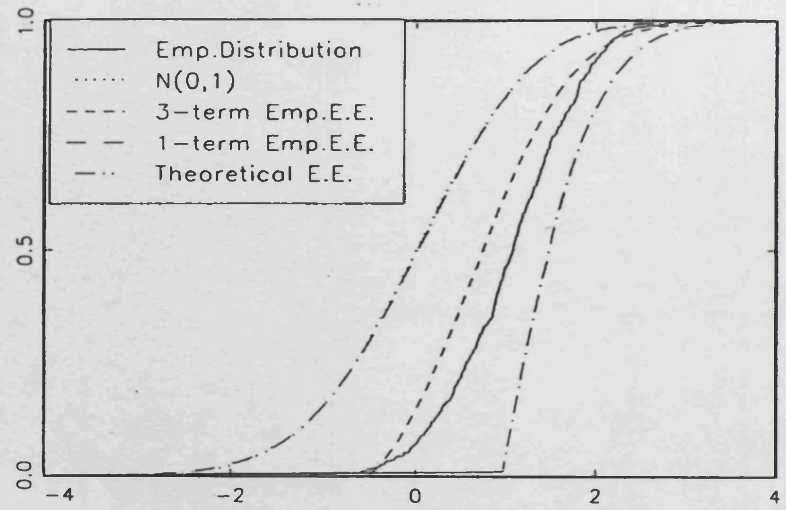


FIGURE 6 , n=400 , h=0.6 , L=4

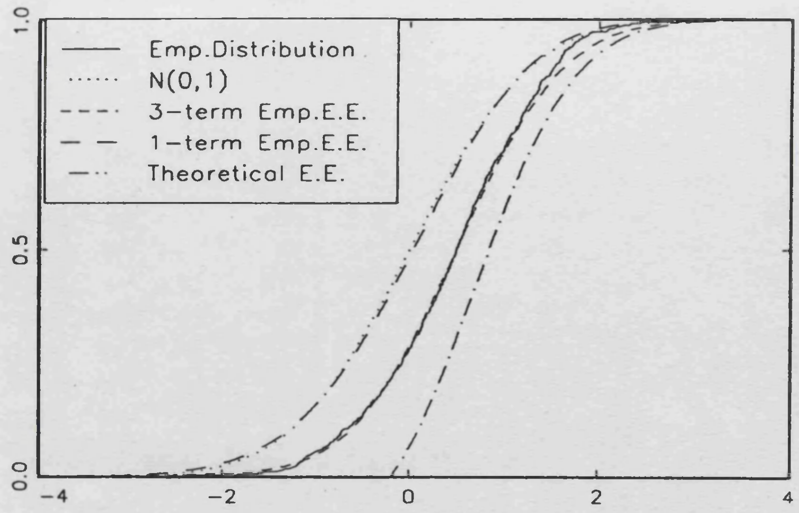


FIGURE 7 , n=400 , h=0.4 , L=4

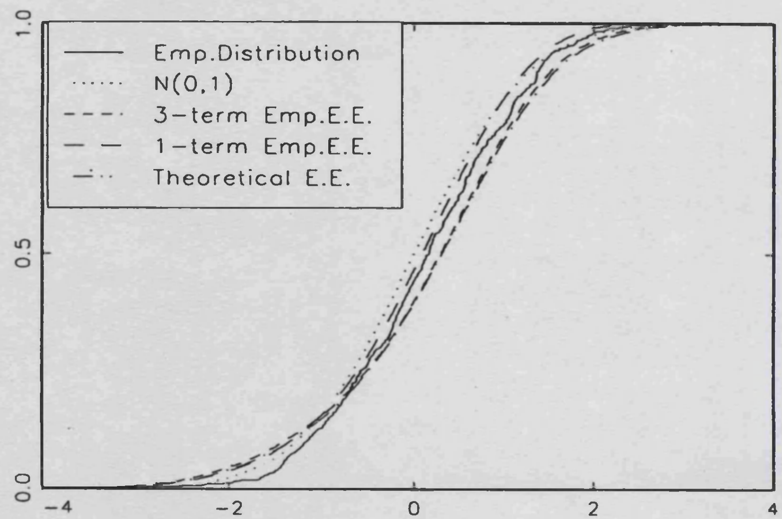


FIGURE 8 , n=100 , h=1.0 , L=8

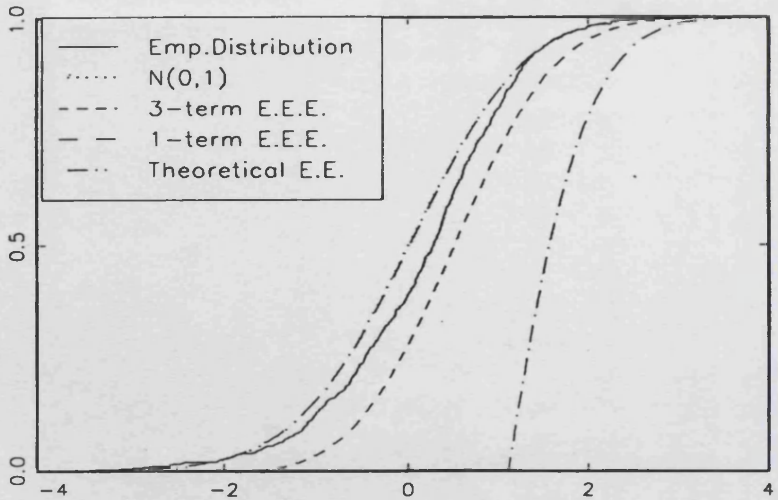


FIGURE 9 , n=100 , h=0.8 , L=8

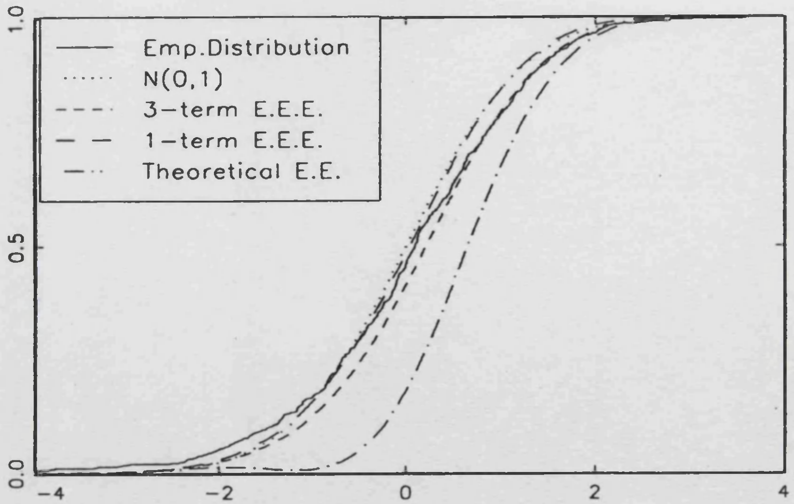


FIGURE 10 , n=100 , h=0.6 , L=8

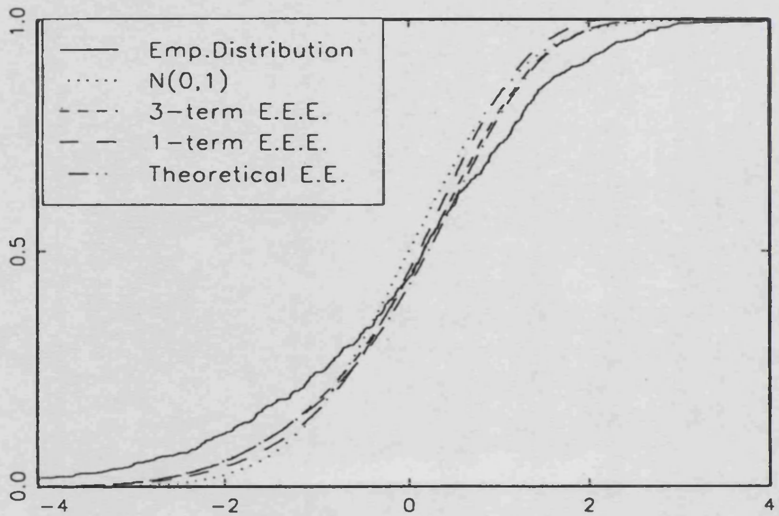


FIGURE 11 , n=100 , h=0.4 , L=8

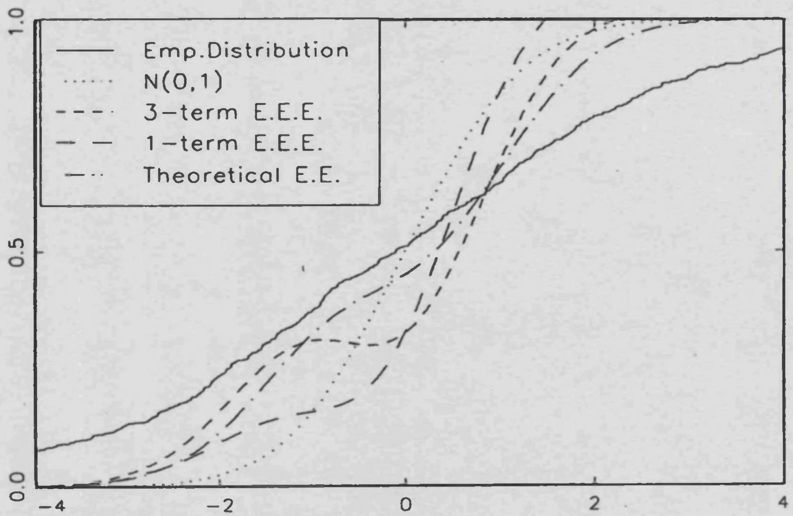


FIGURE 12 , n=400 , h=0.8 , L=8

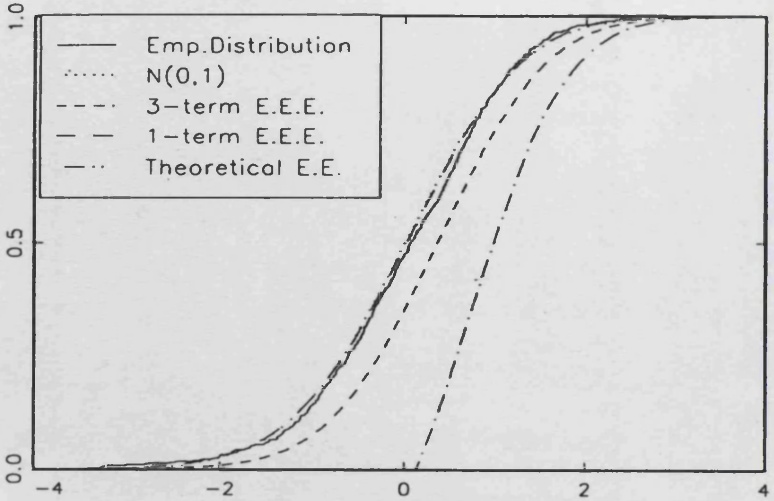


FIGURE 13 , n=400 , h=0.6 , L=8

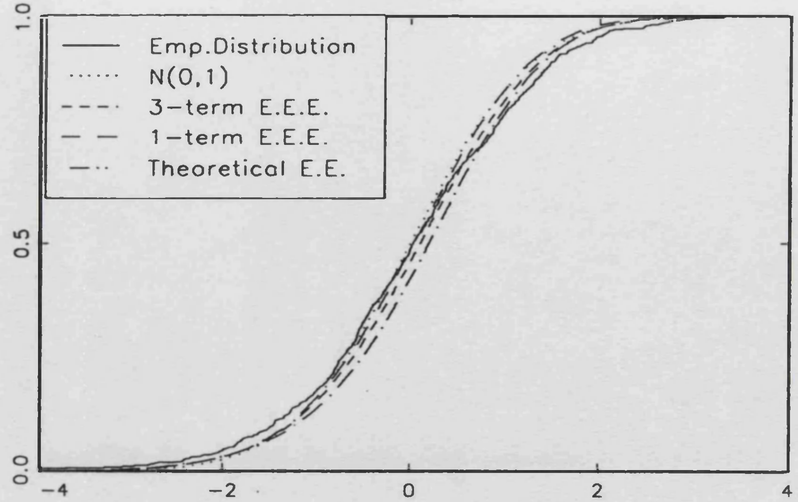


FIGURE 14 , n=400 , h=0.4 , L=8

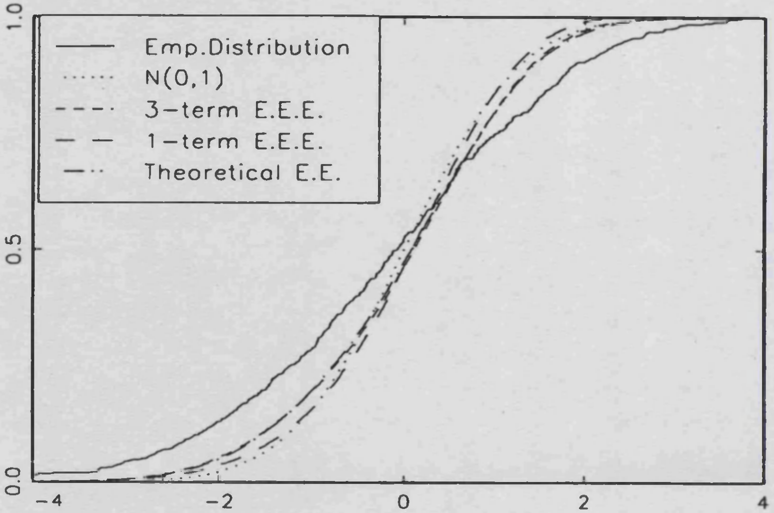


FIGURE 15 , n=100 , h=1.0 , L=10

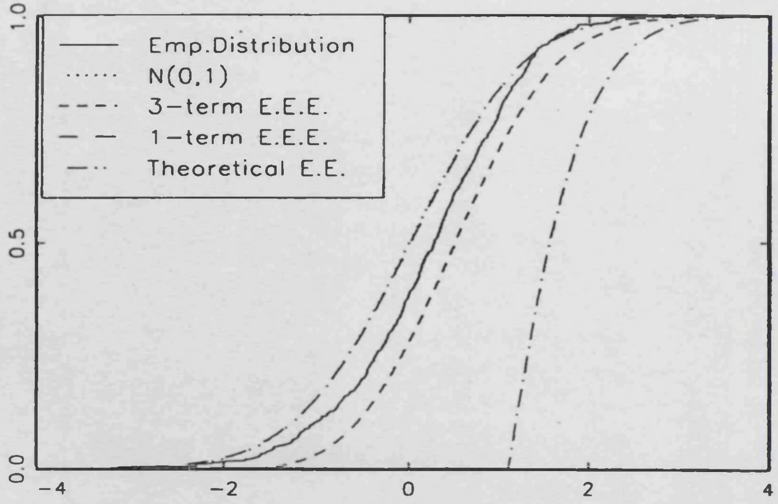


FIGURE 16 , n=100 , h=0.8 , L=10

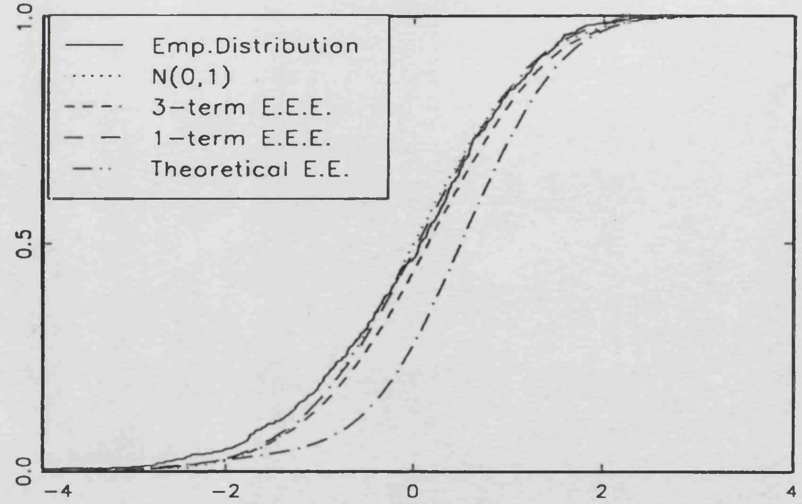


FIGURE 17 , n=100 , h=0.6 , L=10

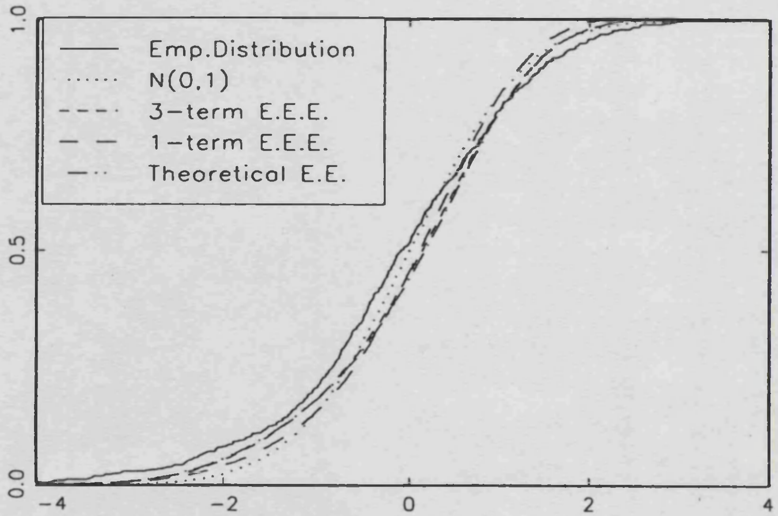


FIGURE 18 , n=100 , h=0.4 , L=10

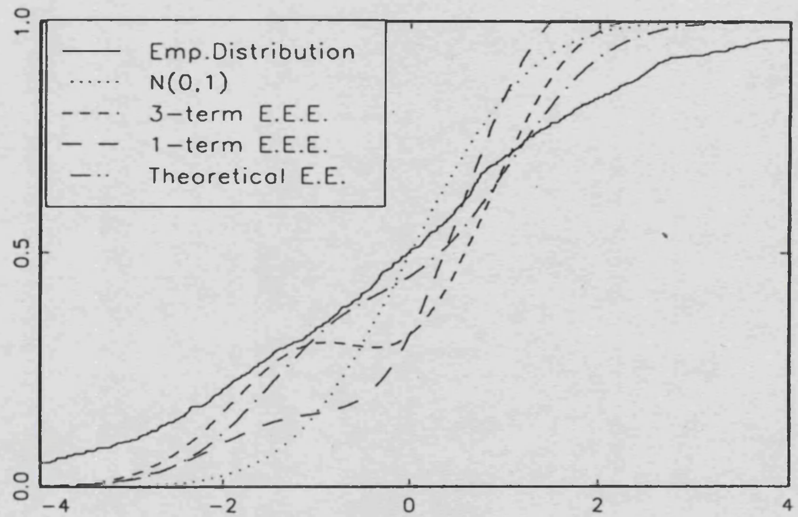


FIGURE 19 , n=400 , h=0.8 , L=10

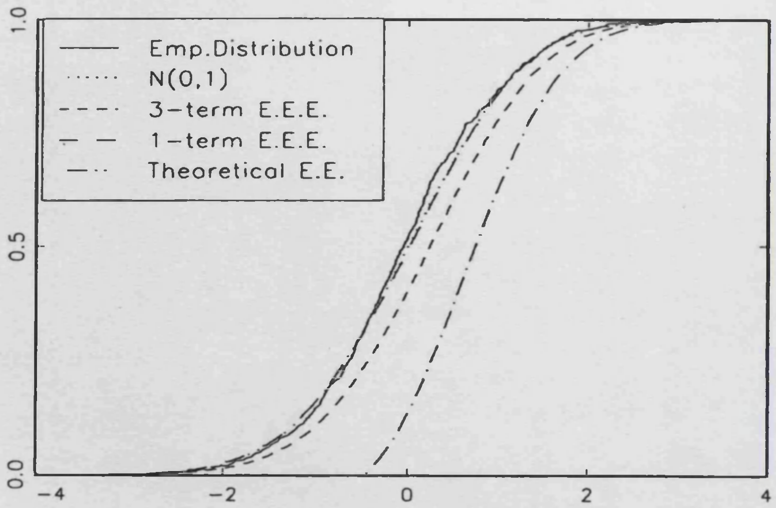


FIGURE 20 , n=400 , h=0.6 , L=10

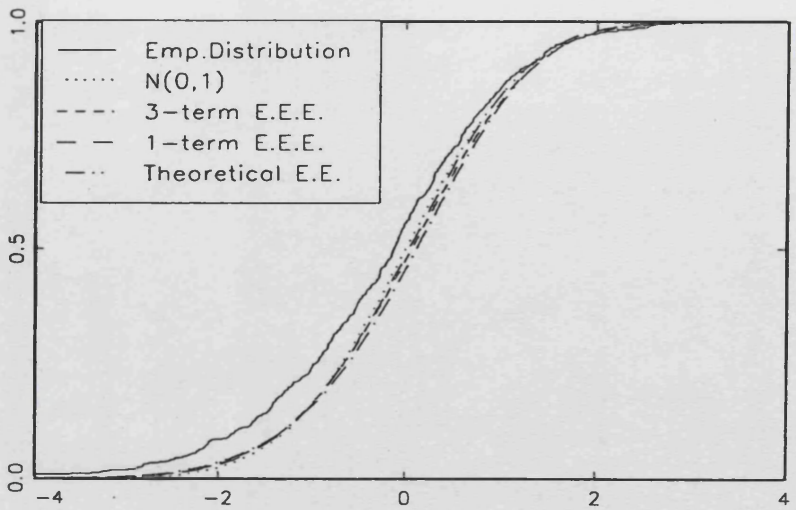


FIGURE 21 , n=400 , h=0.4 , L=10

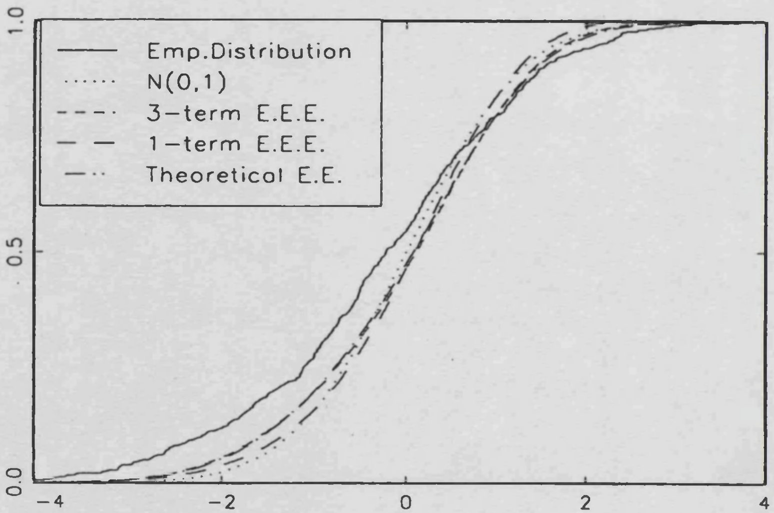


FIGURE 22. n=100 , h=0.8 , L=4 (Studentized case)

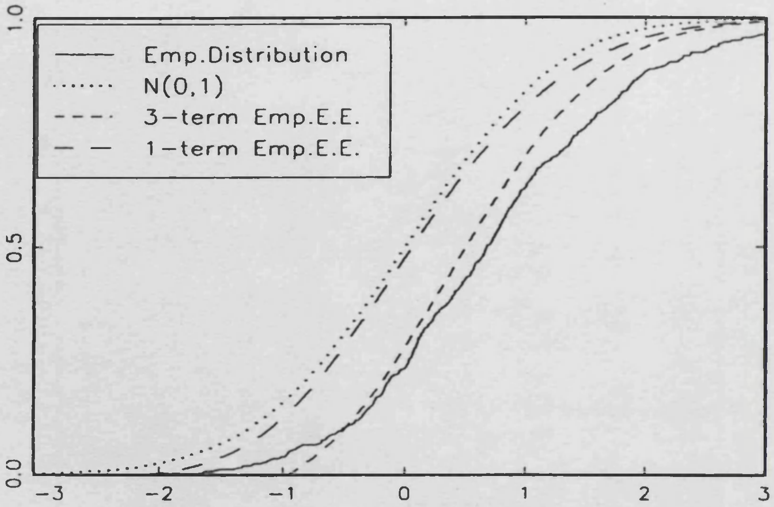


FIGURE 23. n=100 , h=0.6 , L=4 (Studentized case)

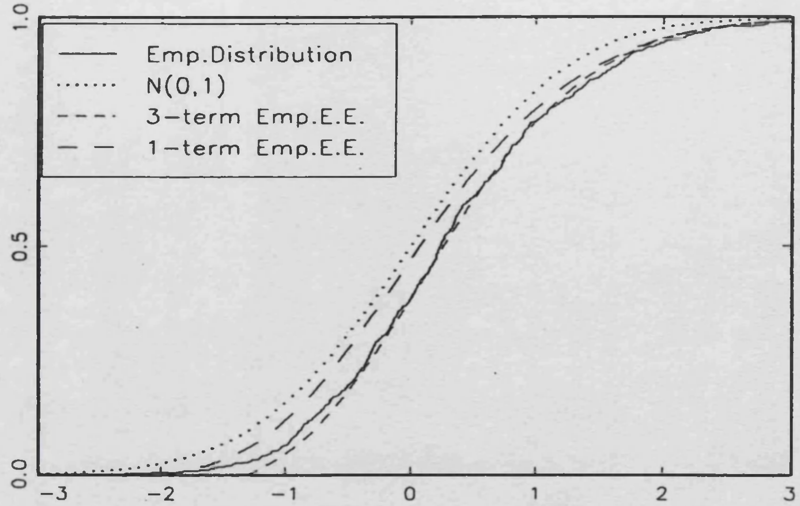


FIGURE 24. n=100 , h=0.4 , L=4 (Studentized case)

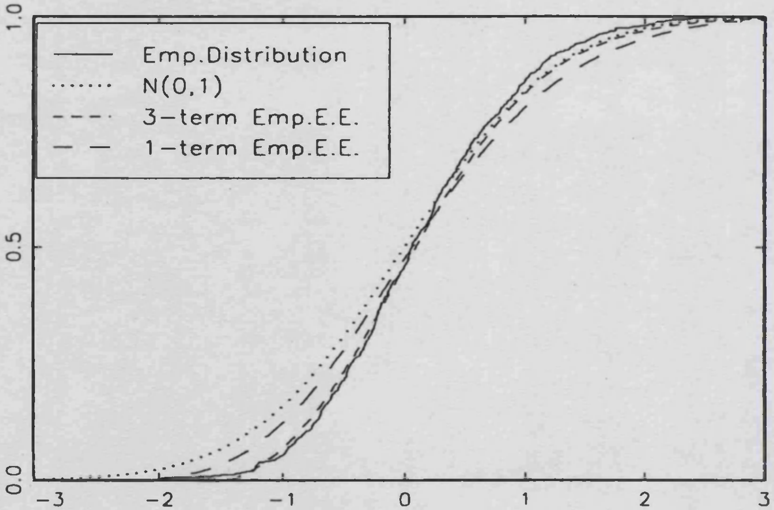


FIGURE 25. n=100 , h=0.2 , L=4 (Studentized case)

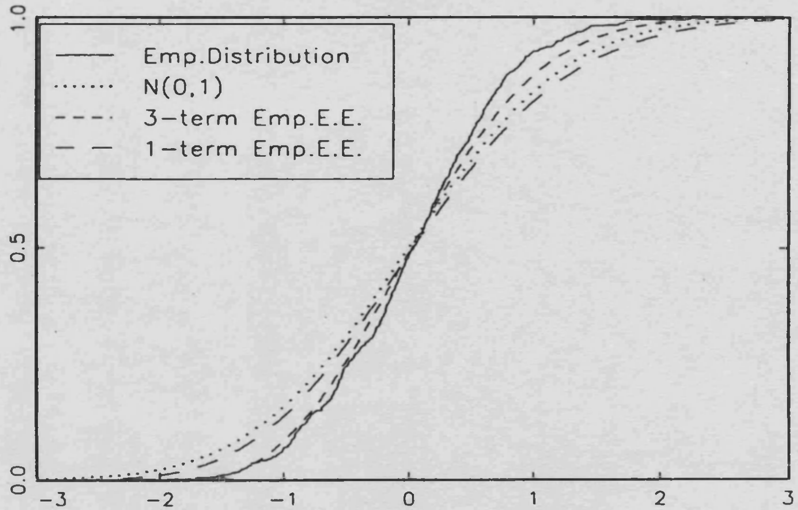


FIGURE 26. n=400 , h=0.8 , L=4 (Studentized case)

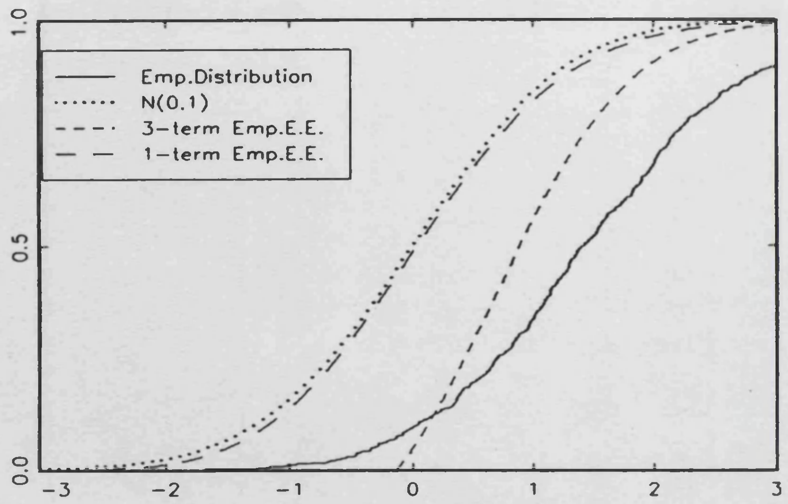


FIGURE 27. n=400 , h=0.6 , L=4 (Studentized case)

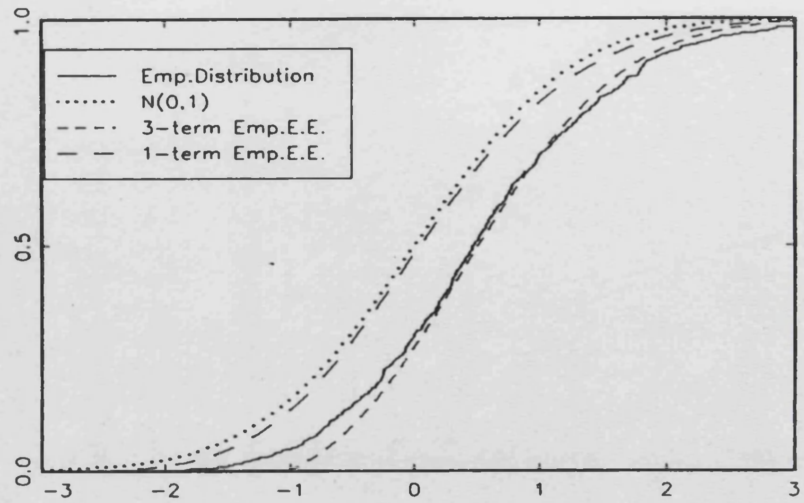


FIGURE 28. n=400 , h=0.4 , L=4 (Studentized case)

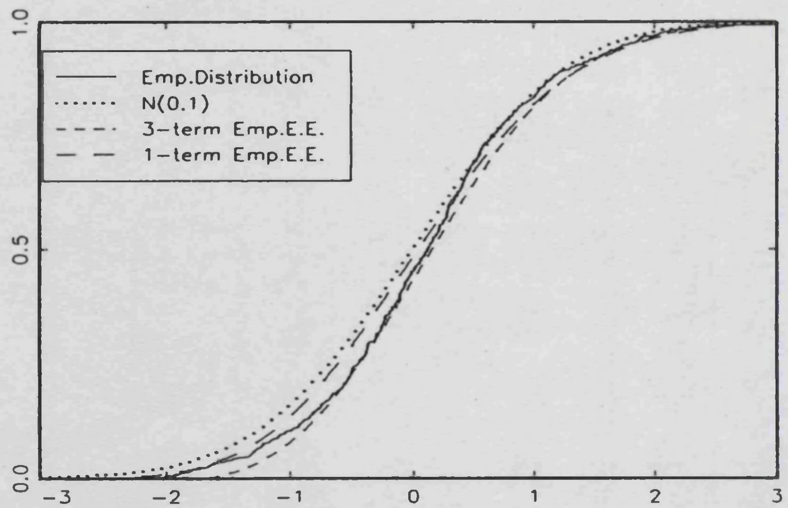


FIGURE 29. n=400 , h=0.2 , L=4 (Studentized case)

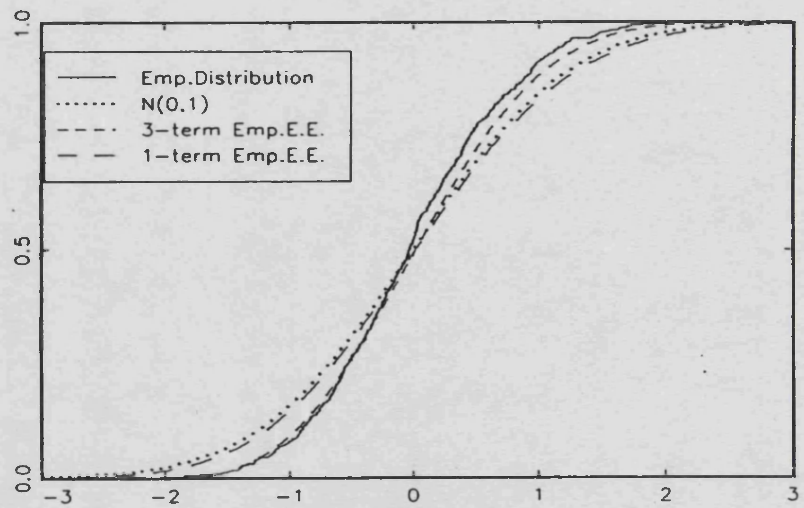


FIGURE 30. n=100 , h=0.8 , L=8 (Studentized case)

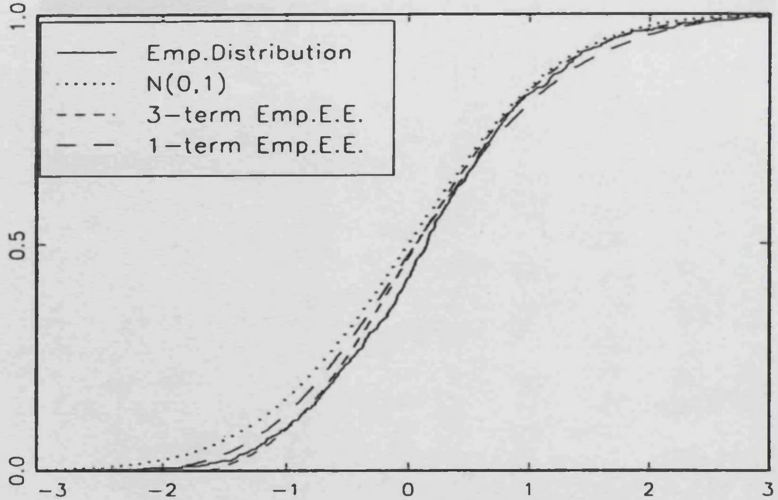


FIGURE 31. n=100 , h=0.6 , L=8 (Studentized case)

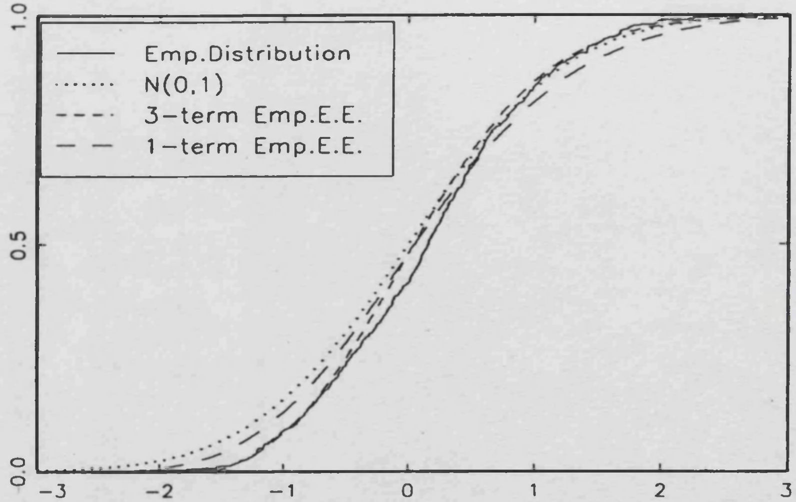


FIGURE 32. n=100 , h=0.4 , L=8 (Studentized case)

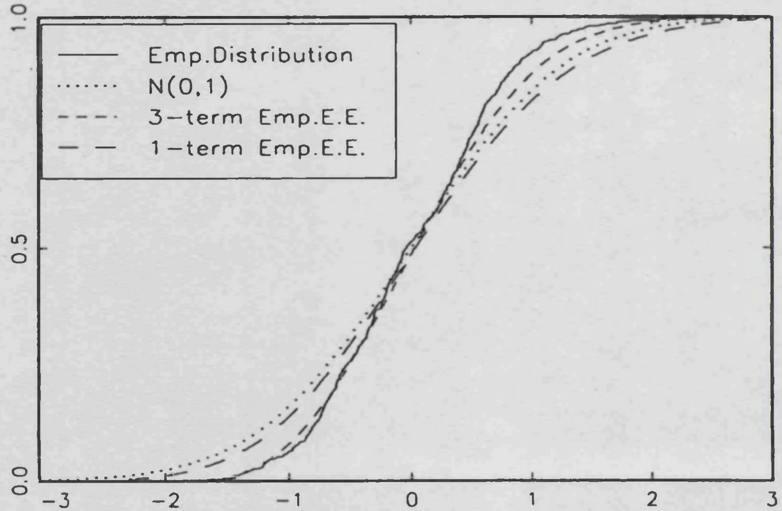


FIGURE 33. n=100 , h=0.2 , L=8 (Studentized case)

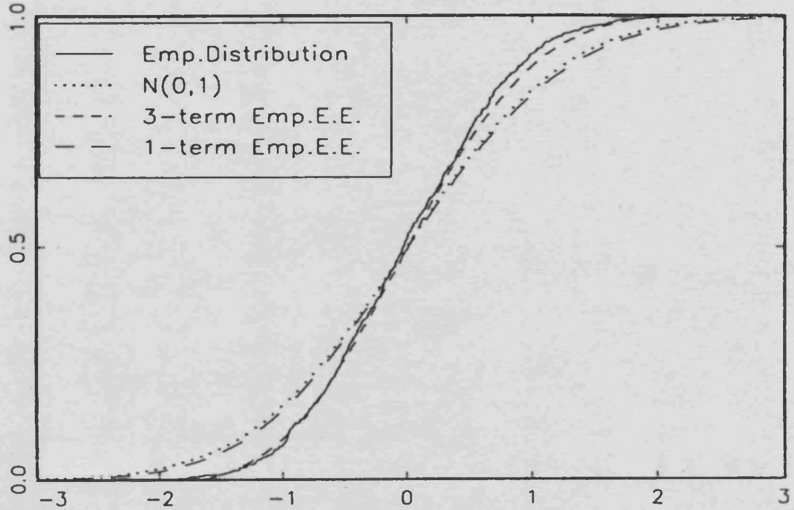


FIGURE 34. n=400 , h=0.8 , L=8 (Studentized case)

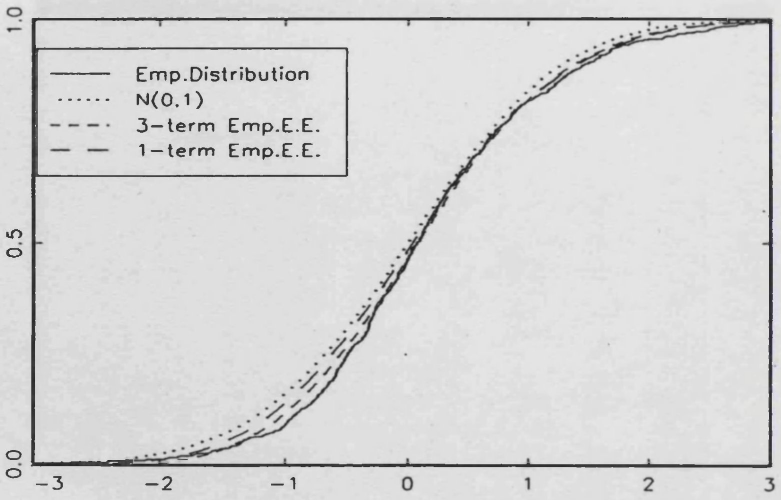


FIGURE 35. n=400 , h=0.6 , L=8 (Studentized case)

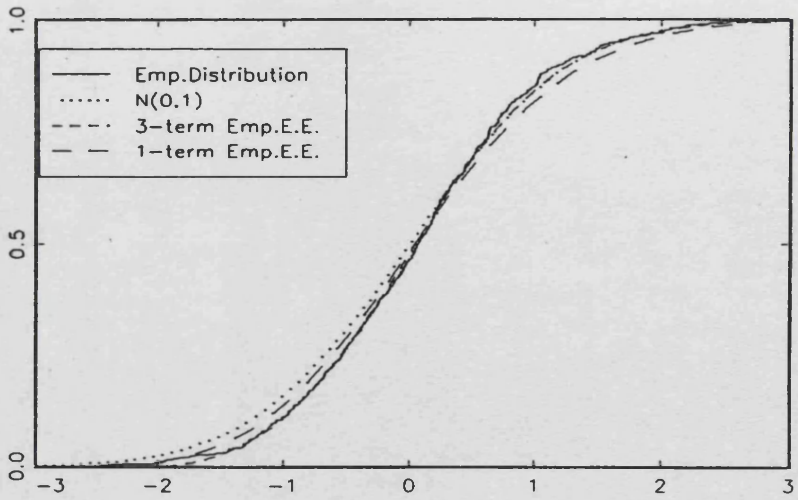


FIGURE 36. n=400 , h=0.4 , L=8 (Studentized case)

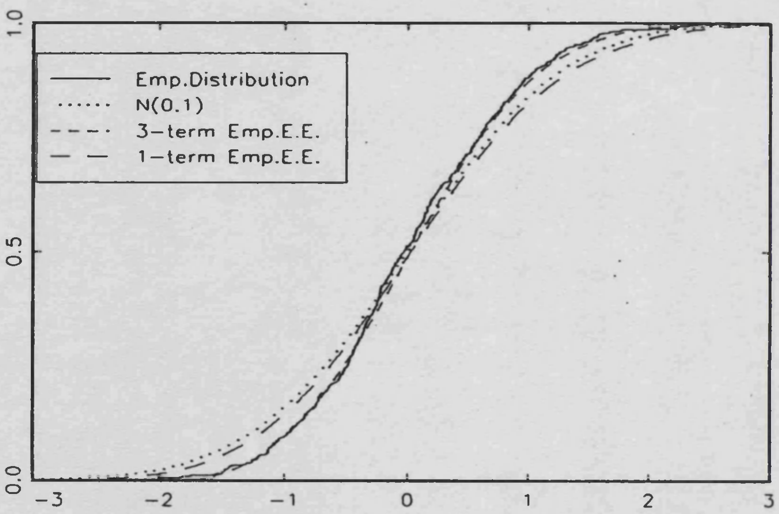


FIGURE 37. n=400 , h=0.2 , L=8 (Studentized case)

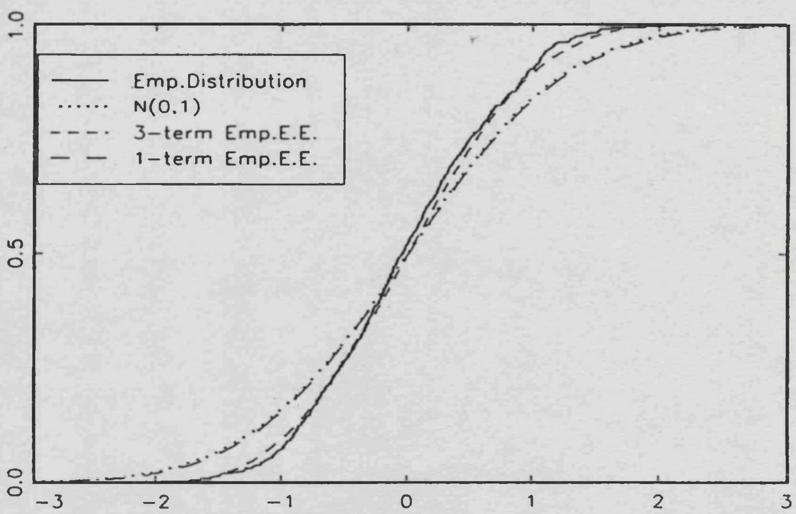


FIGURE 38. n=100 , h=0.8 , L=10 (Studentized case)

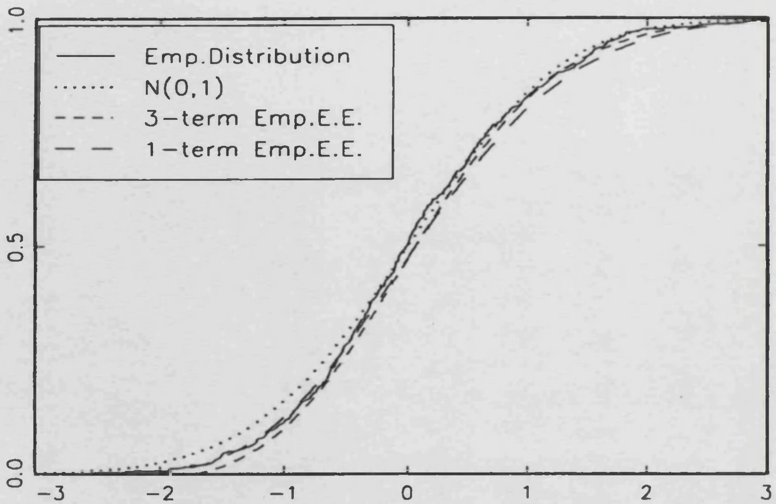


FIGURE 39. n=100 , h=0.6 , L=10 (Studentized case)

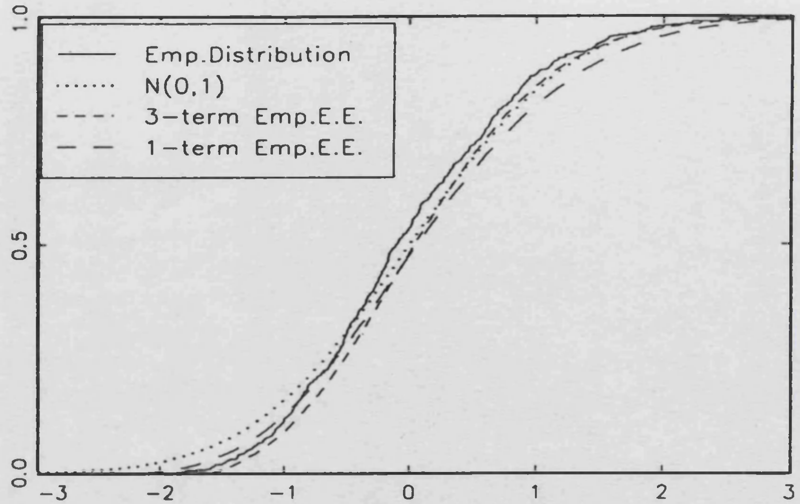


FIGURE 40. n=100 , h=0.4 , L=10 (Studentized case)

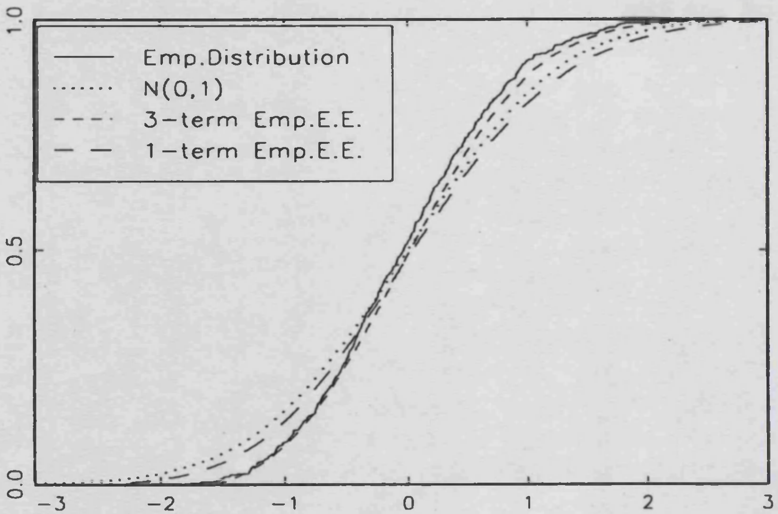


FIGURE 41. n=100 , h=0.2 , L=10 (Studentized case)

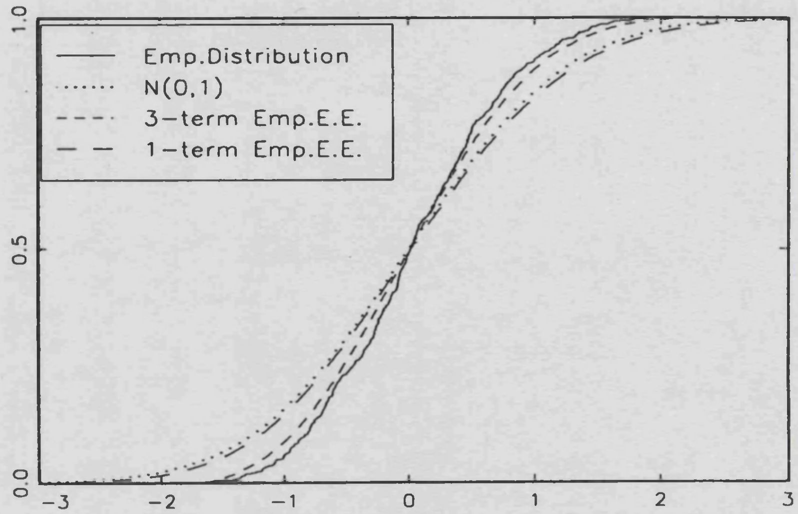


FIGURE 42. n=400 . h=0.8 . L=10 (Studentized case)

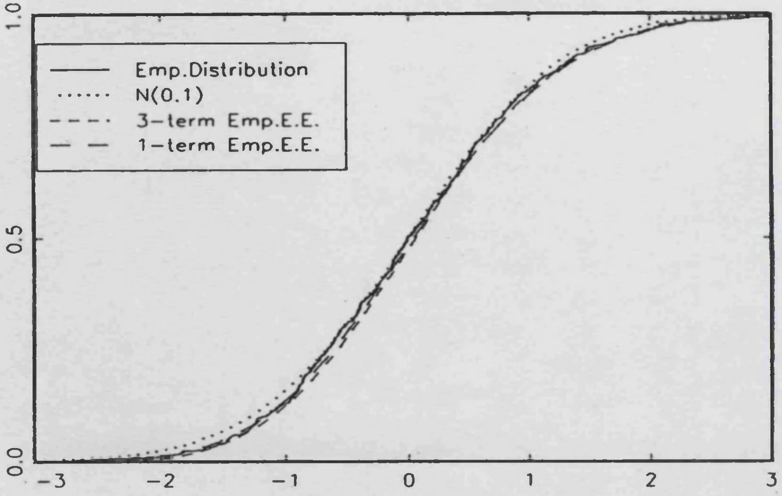


FIGURE 43. n=400 . h=0.6 . L=10 (Studentized case)

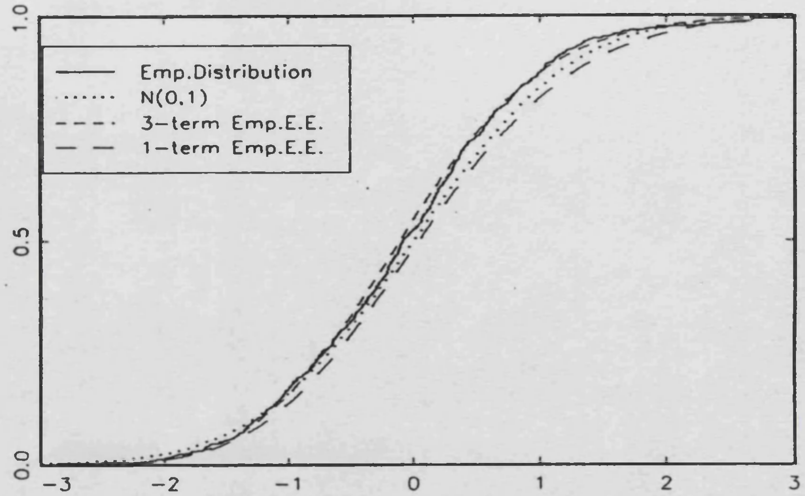


FIGURE 44. n=400 . h=0.4 . L=10 (Studentized case)

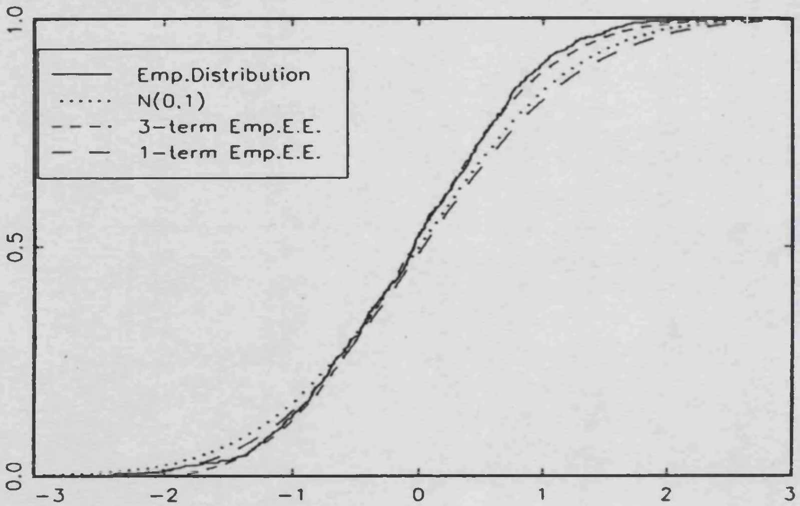


FIGURE 45. n=400 . h=0.2 . L=10 (Studentized case)

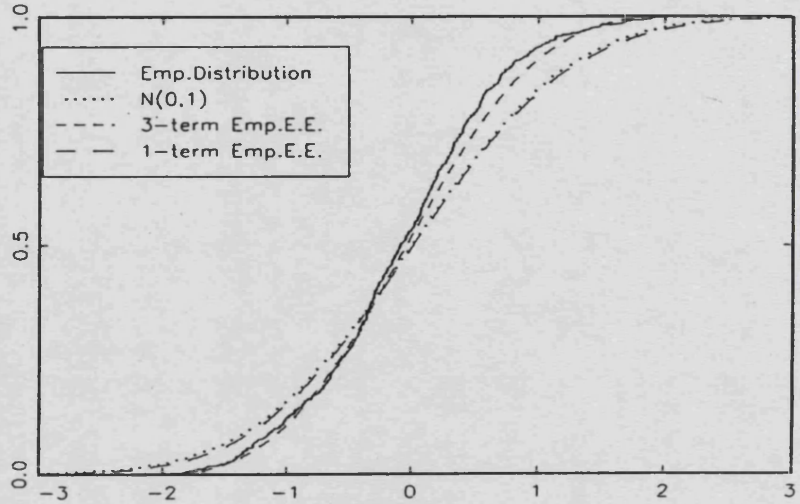


FIGURE 46. Empirical distributions for n=100 , L=4
Optimal bandwidth = 0.445

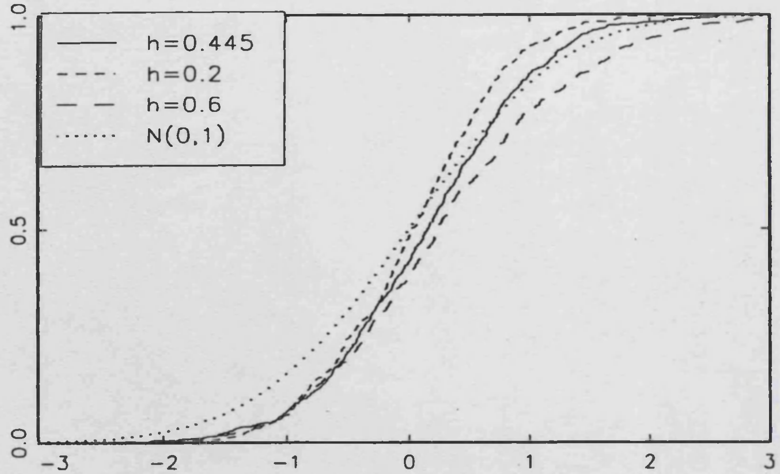


FIGURE 47. Empirical distributions for n=400 , L=4
Optimal bandwidth = 0.343

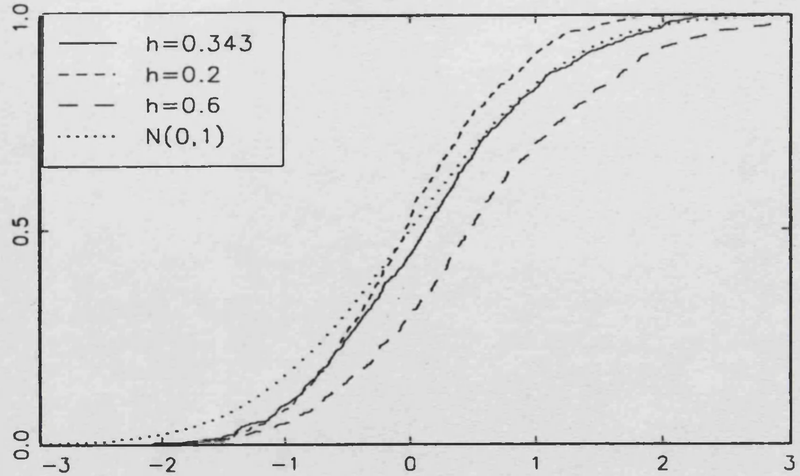


FIGURE 48. 80% C.I. n=100 , h=0.6 , L=4

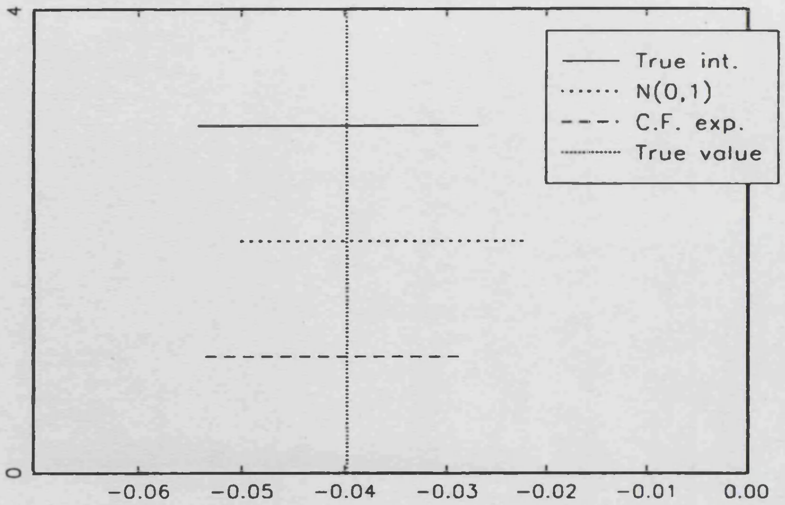


FIGURE 49. 80% C.I. n=400 , h=0.2 , L=4

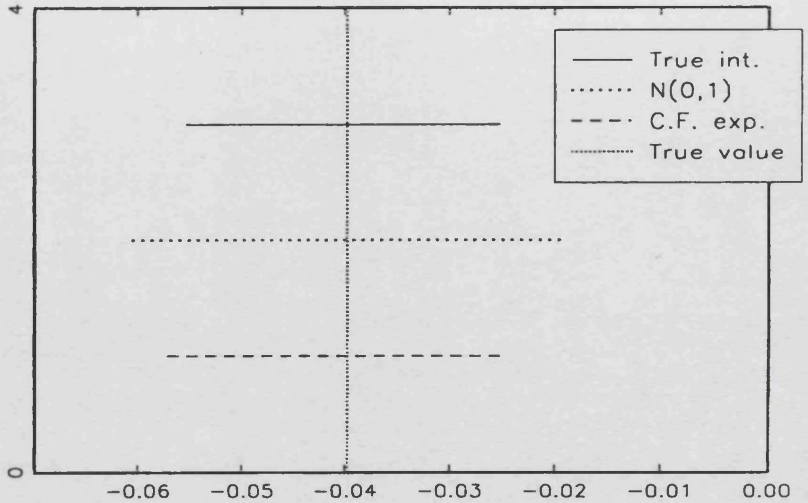


FIGURE 50. 90% C.I. n=100 , h=0.6 , L=4

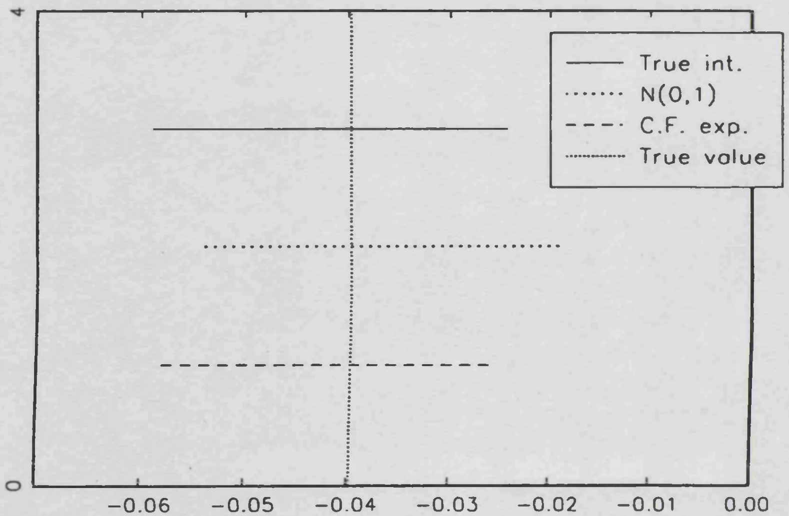


FIGURE 51. 90% C.I. n=400 , h=0.2 , L=4

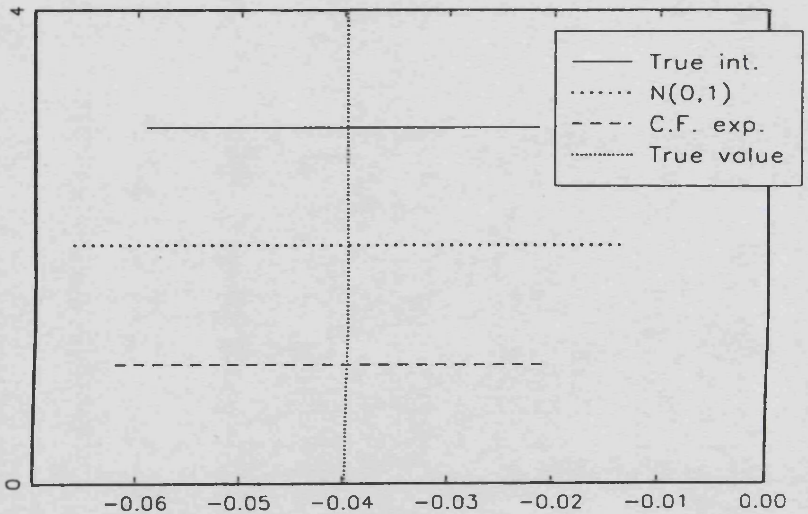


FIGURE 52. n=100 , h=1.0 , L=4

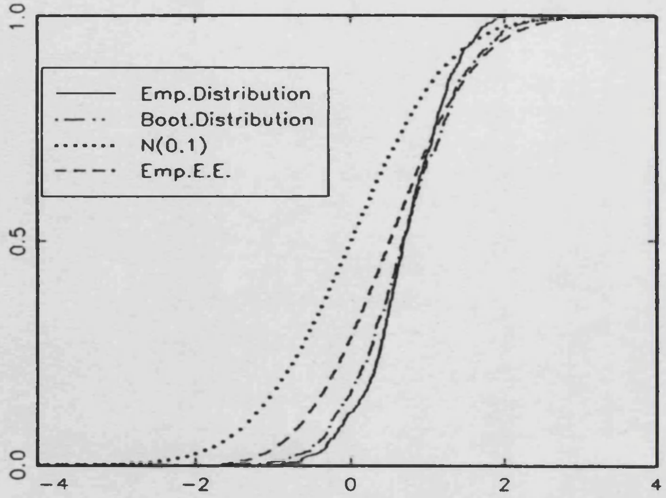


FIGURE 53. n=100 , h=0.8 , L=4

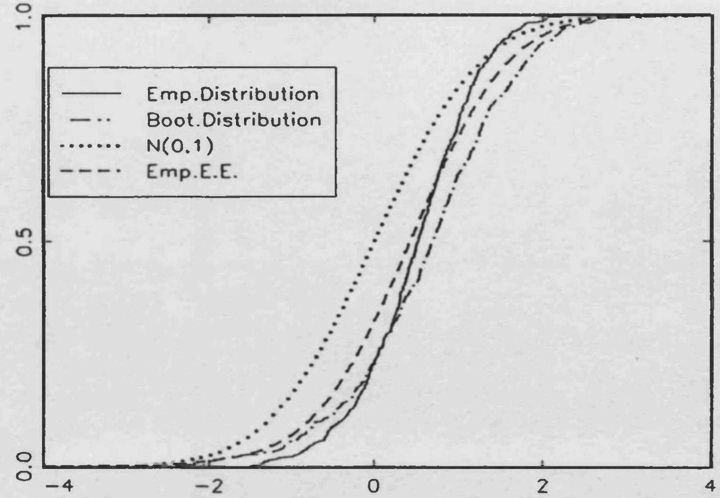


FIGURE 54. n=100 , h=0.6 , L=4

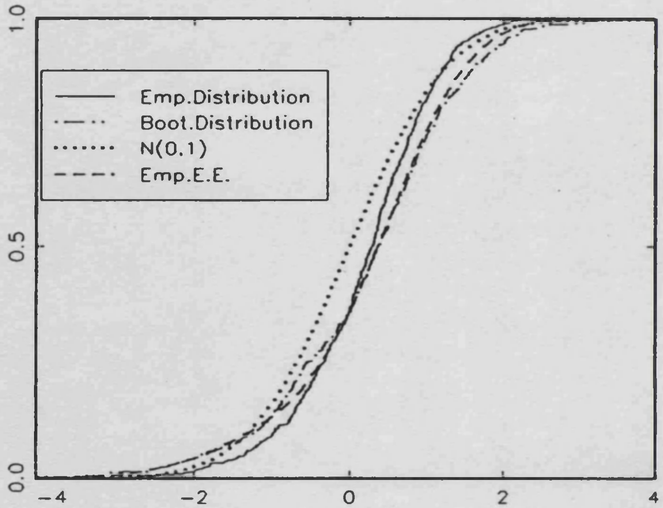


FIGURE 55. n=100 , h=0.4 , L=4

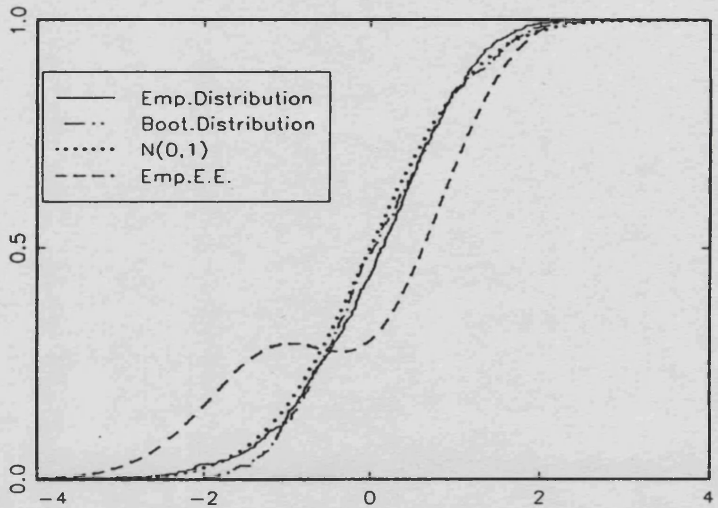


FIGURE 56. n=400 , h=0.8 , L=4

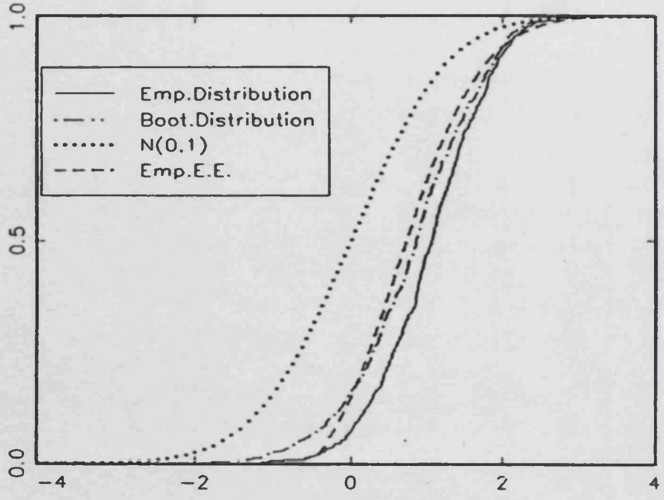


FIGURE 57. n=400 , h=0.6 , L=4

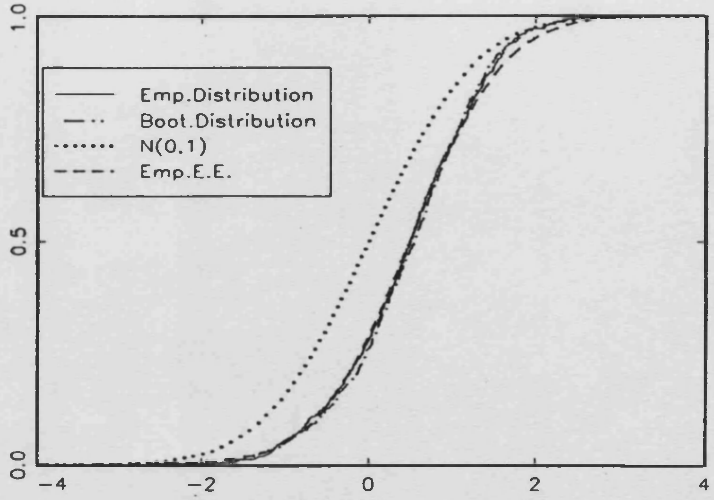


FIGURE 58. n=400 , h=0.4 , L=4

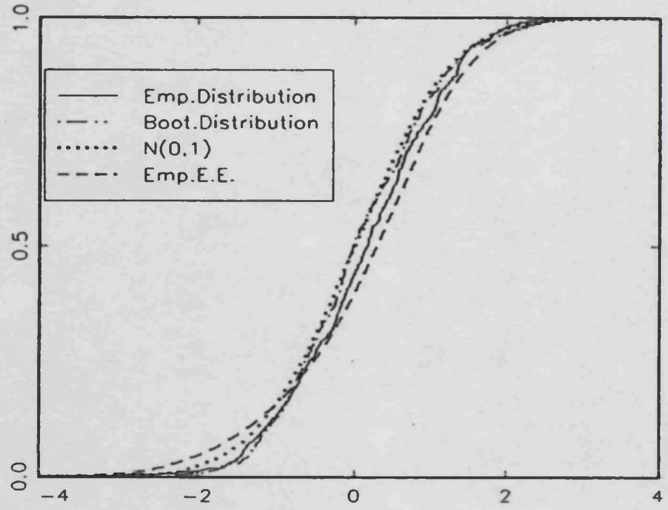


FIGURE 59. n=100 , h=1.0 , L=8

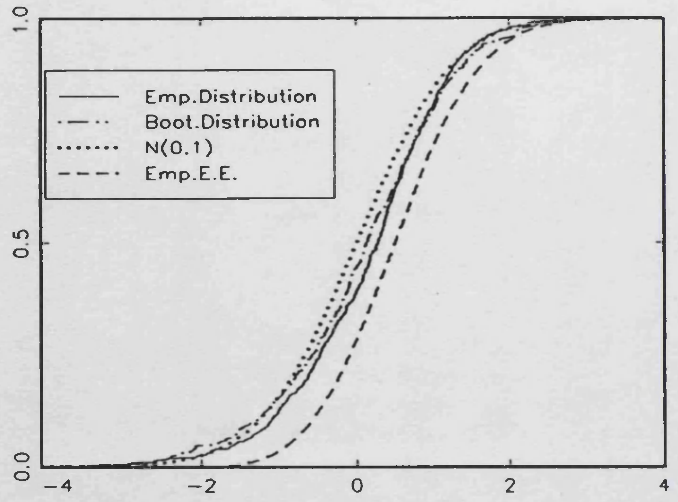


FIGURE 60. n=100 , h=0.8 , L=8

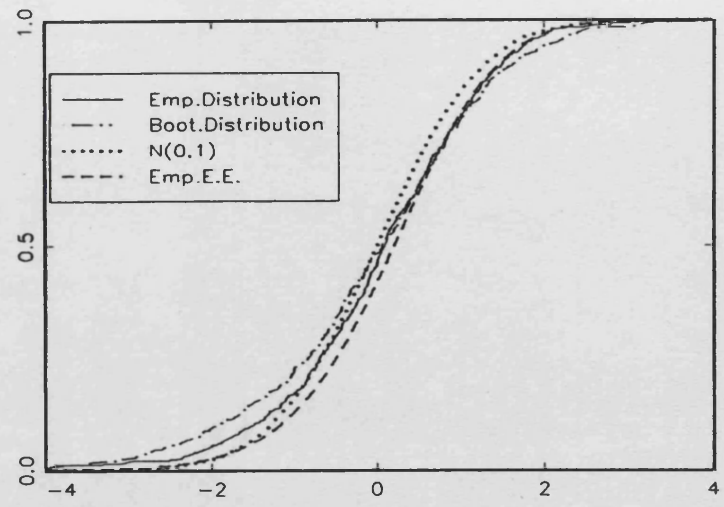


FIGURE 61. n=100 , h=0.6 , L=8

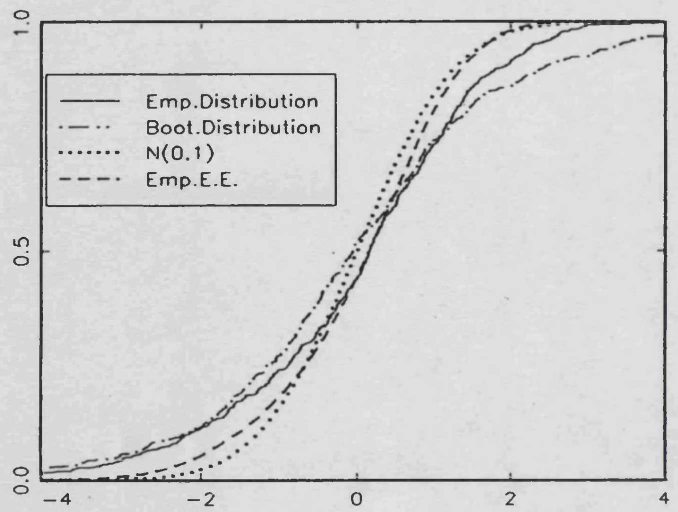


FIGURE 62. n=100 , h=0.4 , L=8

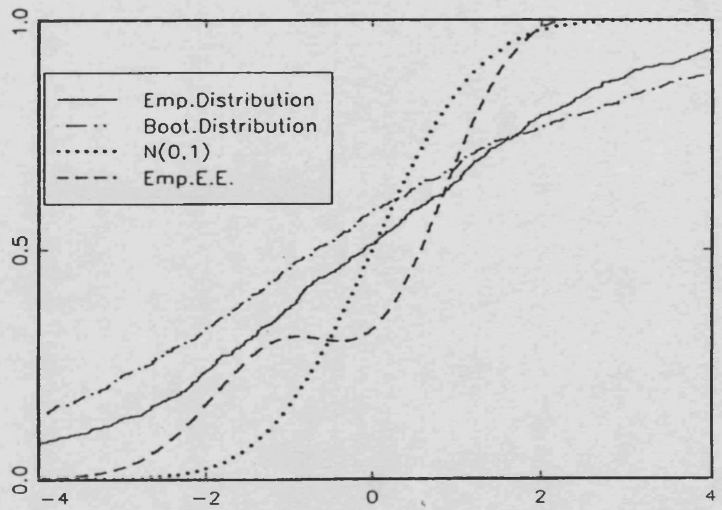


FIGURE 63. n=400 , h=0.8 , L=8

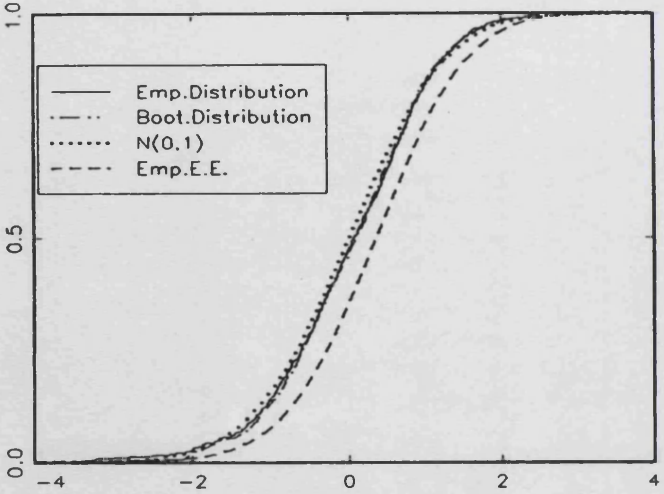


FIGURE 64. n=400 , h=0.6 , L=8

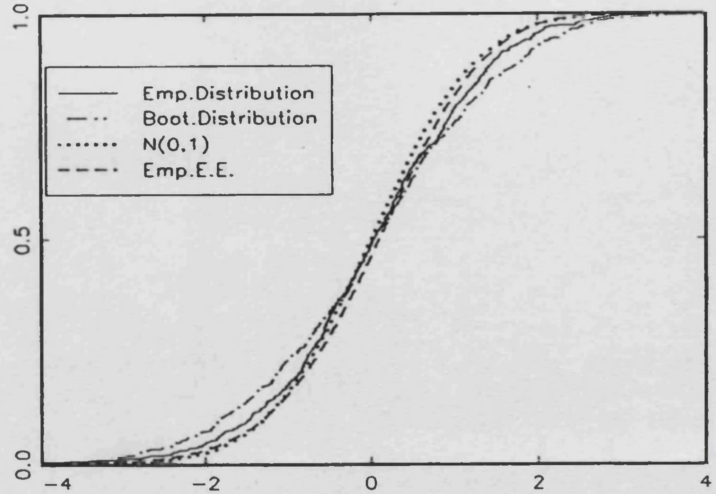


FIGURE 65. n=400 , h=0.4 , L=8

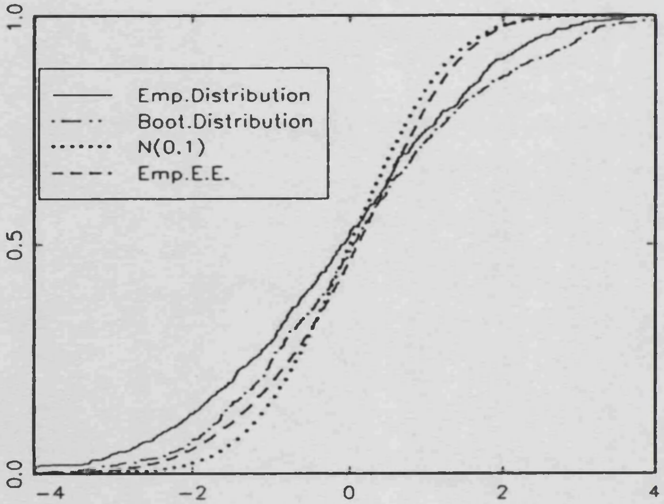


FIGURE 66. n=100 , h=1.0 , L=10

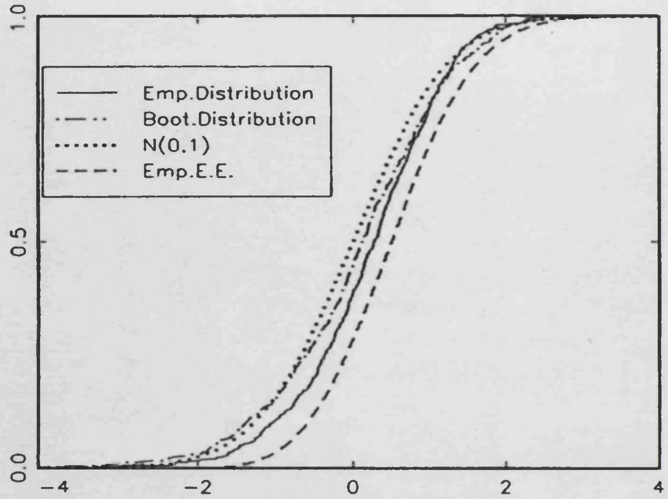


FIGURE 67. n=100 , h=0.8 , L=10

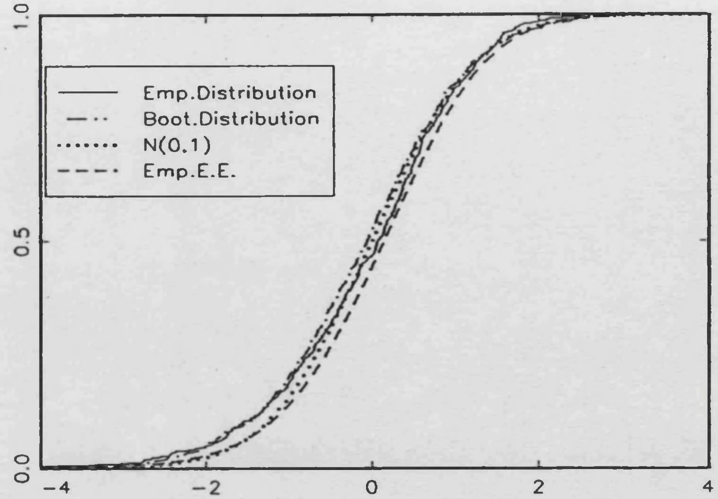


FIGURE 68. n=100 , h=0.6 , L=10

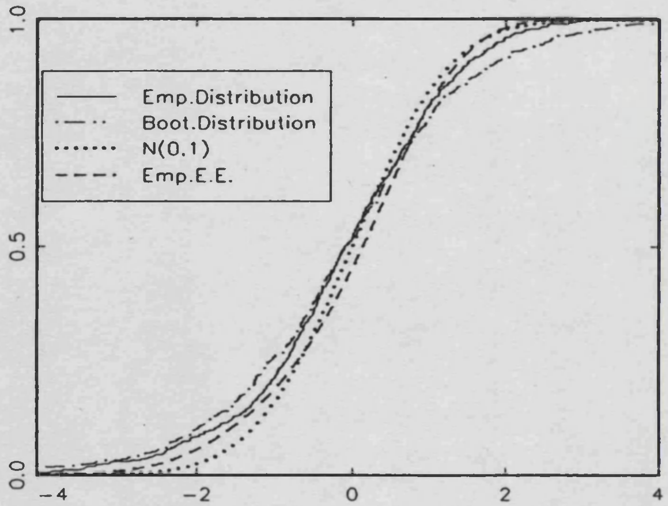


FIGURE 69. n=100 , h=0.4 , L=10

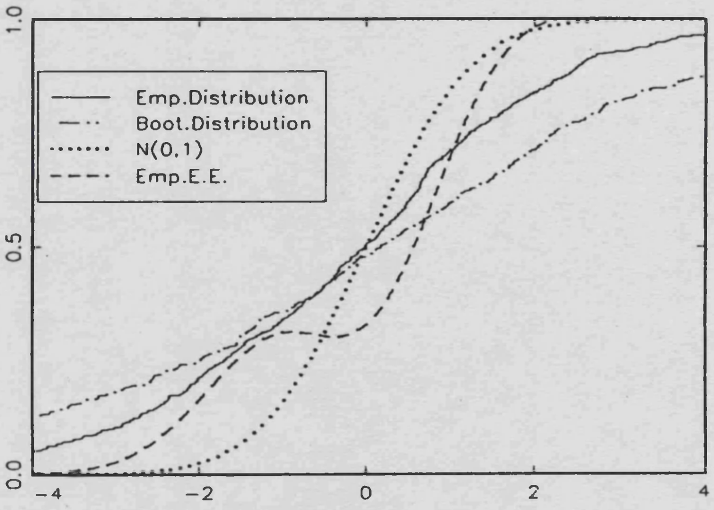


FIGURE 70. n=400 , h=0.8 , L=10

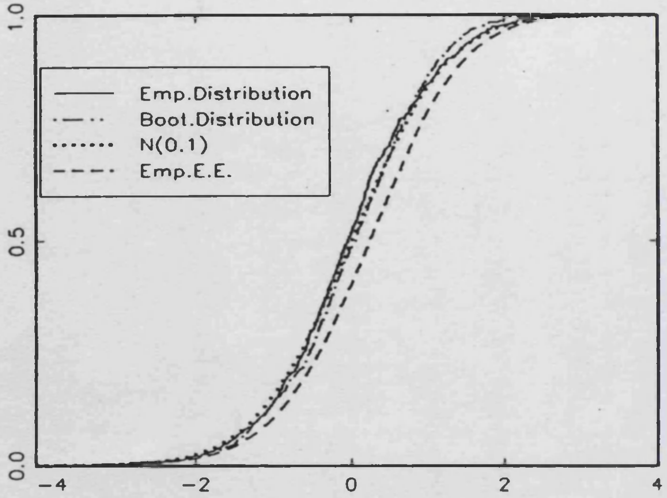


FIGURE 71. n=400 , h=0.6 , L=10

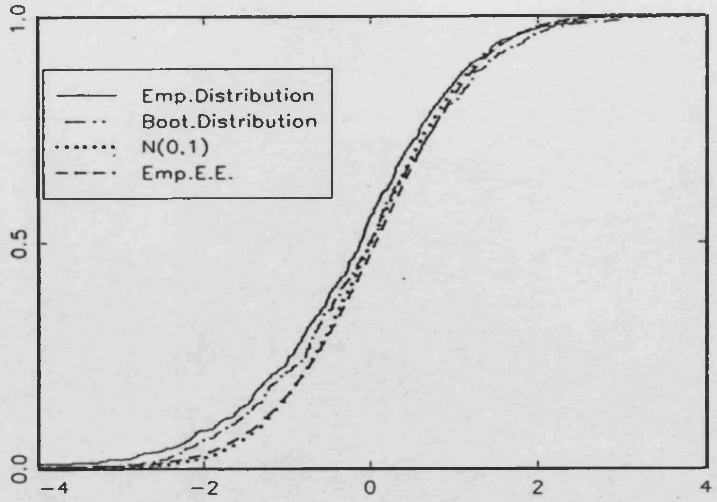


FIGURE 72. n=400 , h=0.4 , L=10

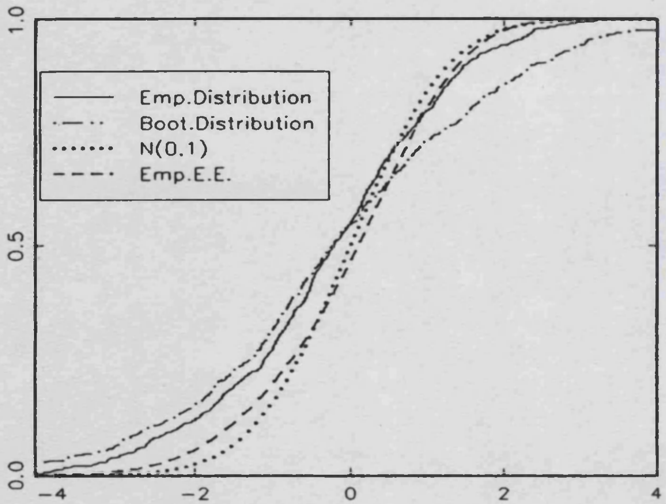


FIGURE 73. n=100 , h=0.8 , L=4 (Studentized case)

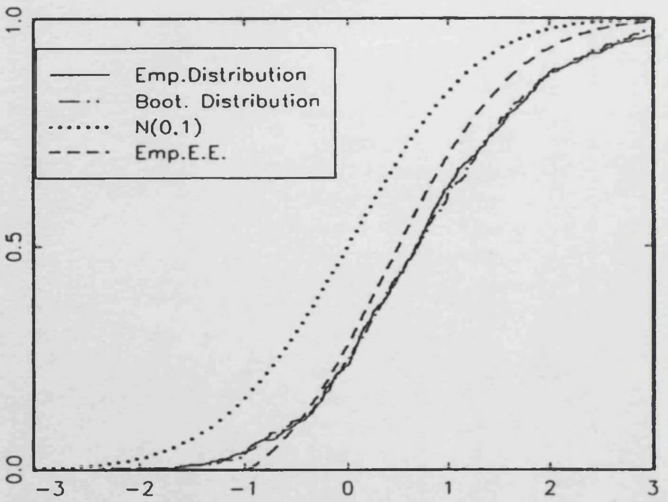


FIGURE 74. n=100 , h=0.6 , L=4 (Studentized case)

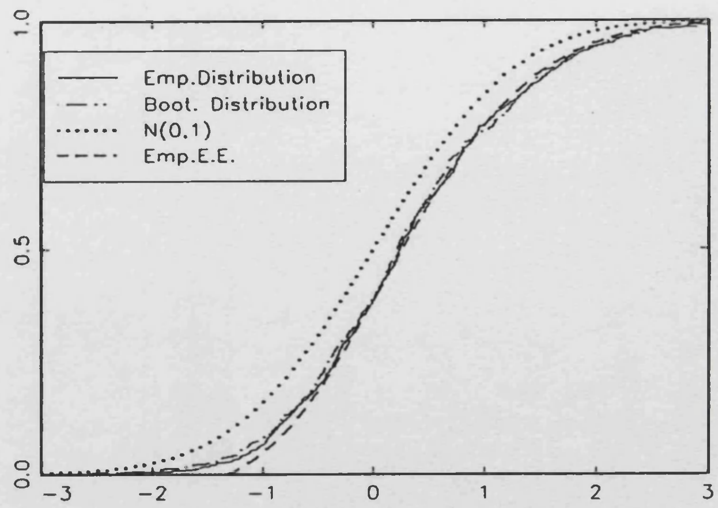


FIGURE 75. n=100 , h=0.4 , L=4 (Studentized case)

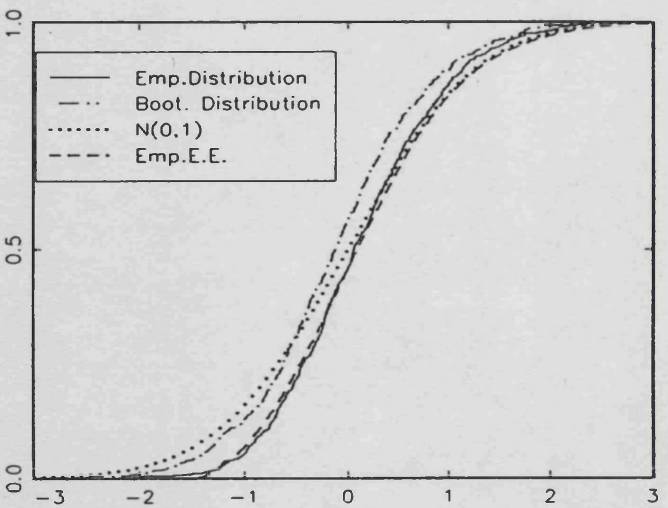


FIGURE 76. n=100 , h=0.2 , L=4 (Studentized case)

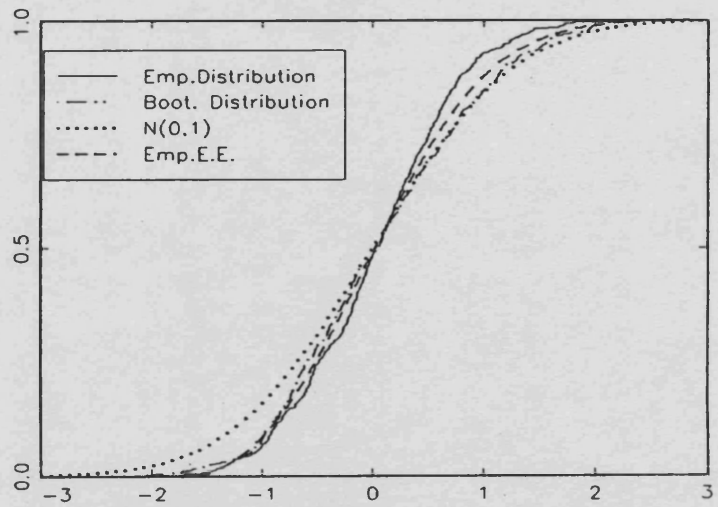


FIGURE 77. n=400 , h=0.8 , L=4 (Studentized case)

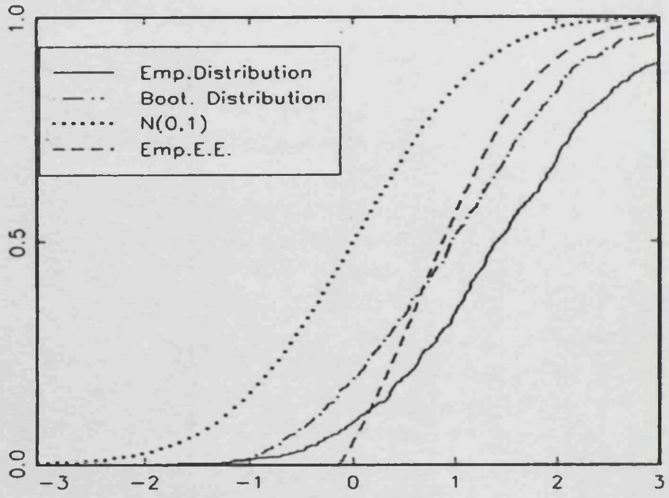


FIGURE 78. n=400 , h=0.6 , L=4 (Studentized case)

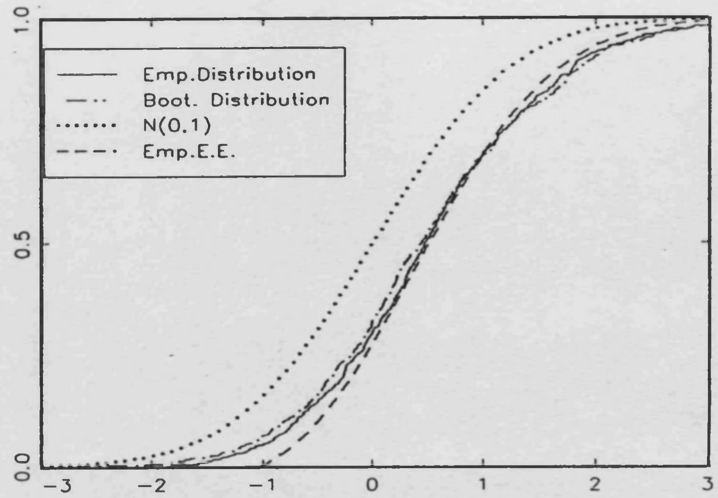


FIGURE 79. n=400 , h=0.4 , L=4 (Studentized case)

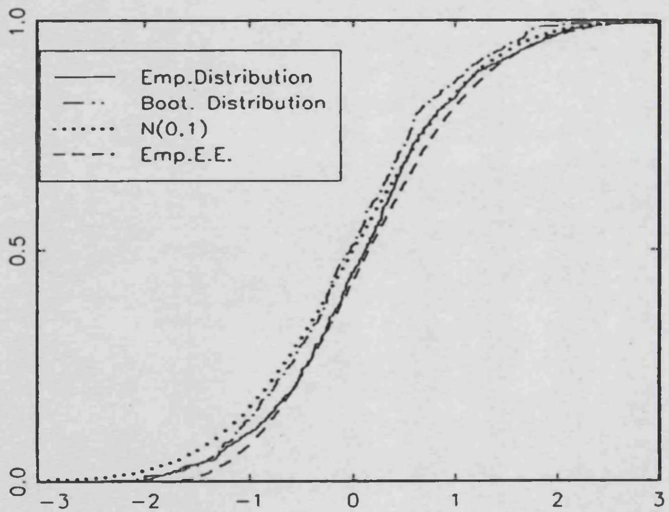


FIGURE 80. n=400 , h=0.2 , L=4 (Studentized case)

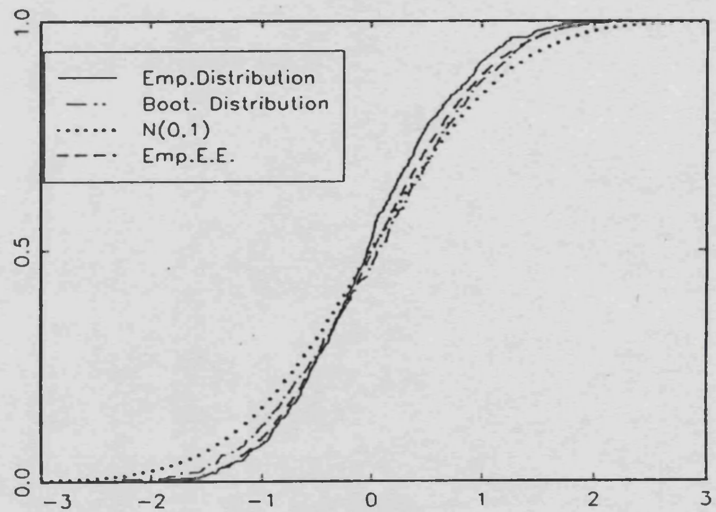


FIGURE 81. n=100 , h=0.8 , L=8 (Studentized case)

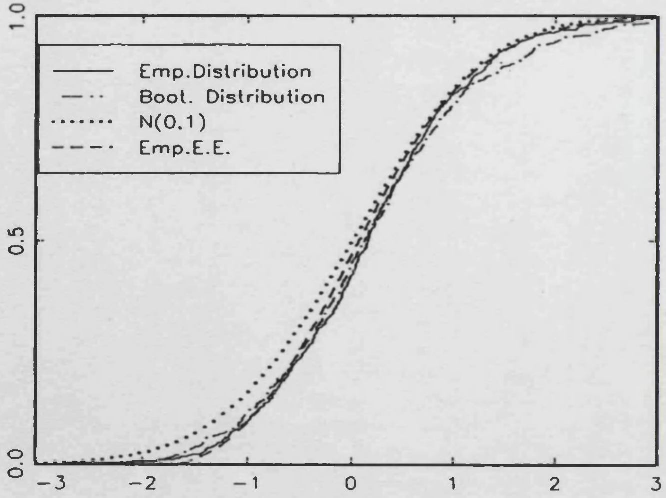


FIGURE 82. n=100 , h=0.6 , L=8 (Studentized case)

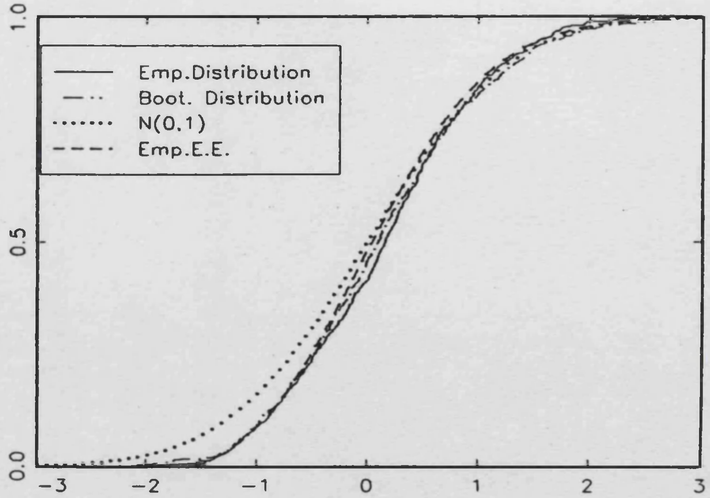


FIGURE 83. n=100 , h=0.4 , L=8 (Studentized case)

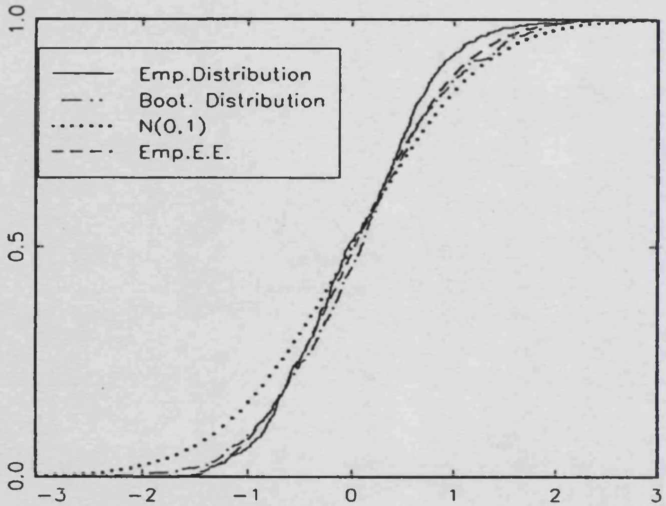


FIGURE 84. n=100 , h=0.2 , L=8 (Studentized case)

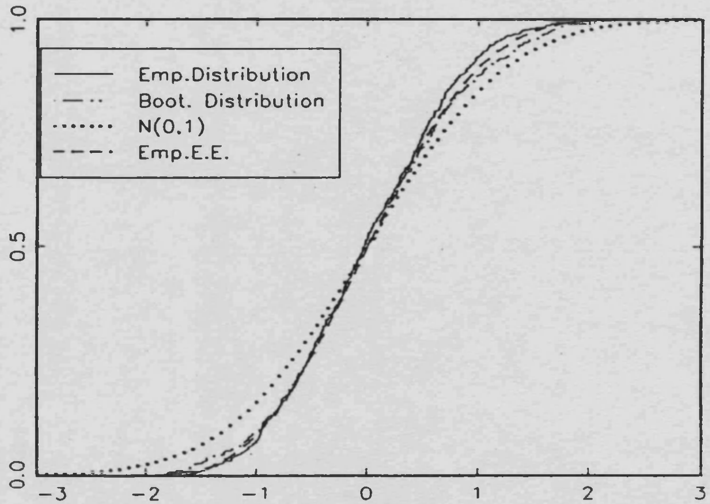


FIGURE 85. n=400 , h=0.8 , L=8 (Studentized case)

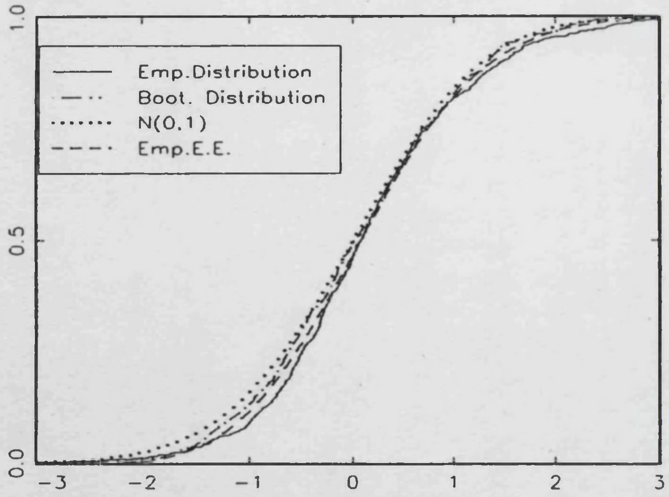


FIGURE 86. n=400 , h=0.6 , L=8 (Studentized case)

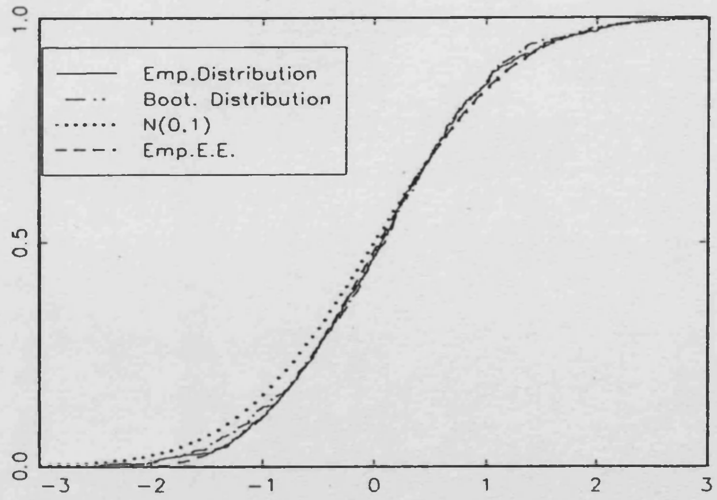


FIGURE 87. n=400 , h=0.4 , L=8 (Studentized case)

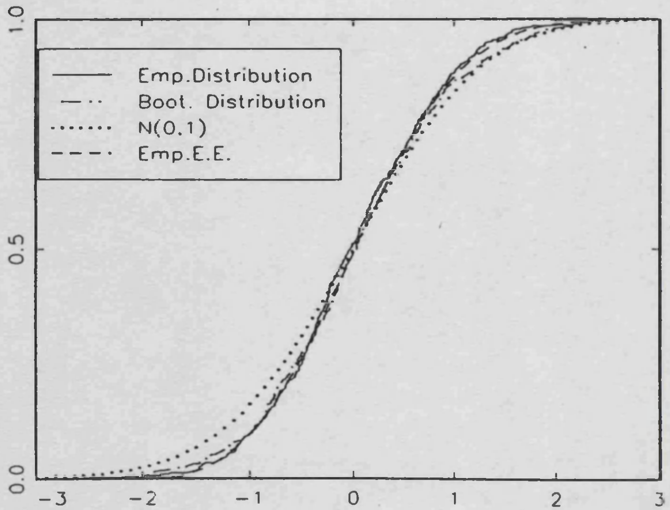


FIGURE 88. n=400 , h=0.2 , L=8 (Studentized case)

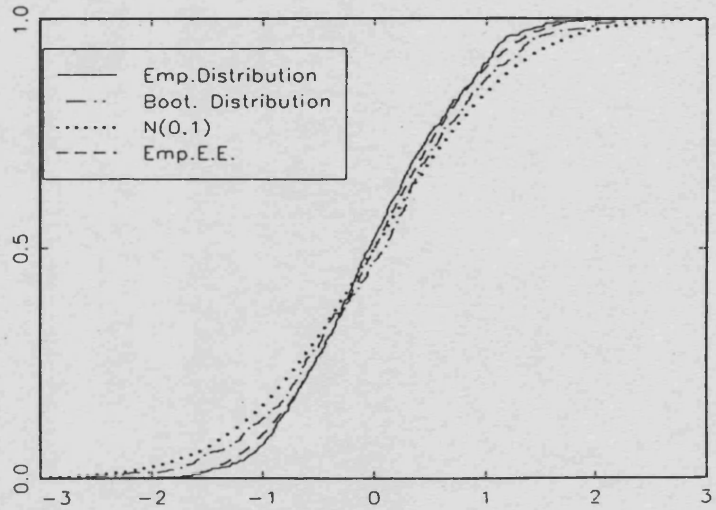


FIGURE 89. n=100 , h=0.8 , L=10 (Studentized case)

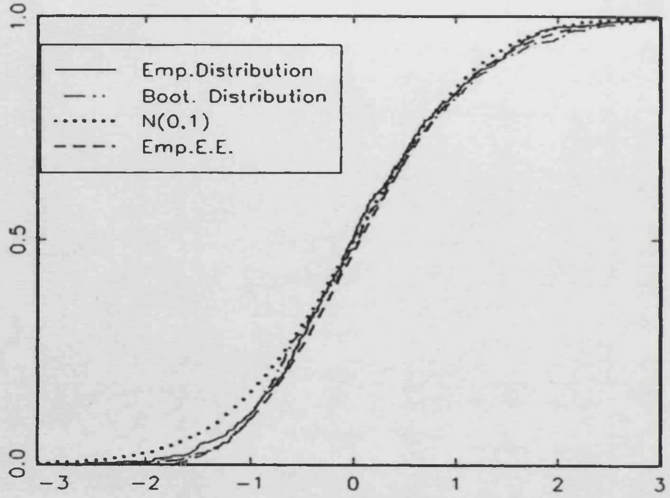


FIGURE 90. n=100 , h=0.6 , L=10 (Studentized case)

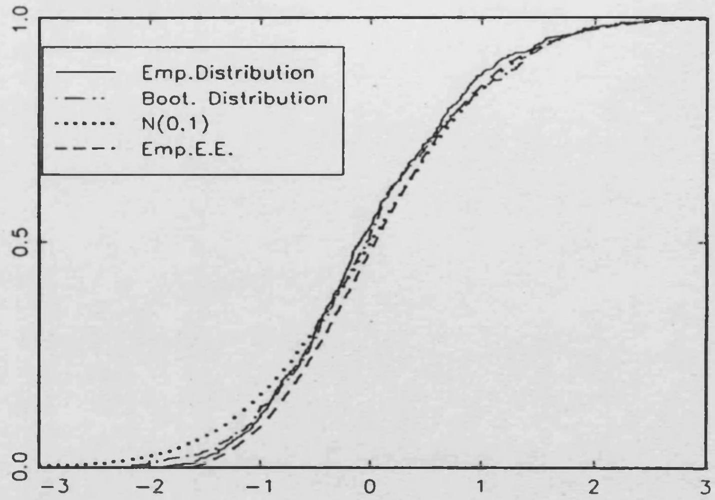


FIGURE 91. n=100 , h=0.4 , L=10 (Studentized case)

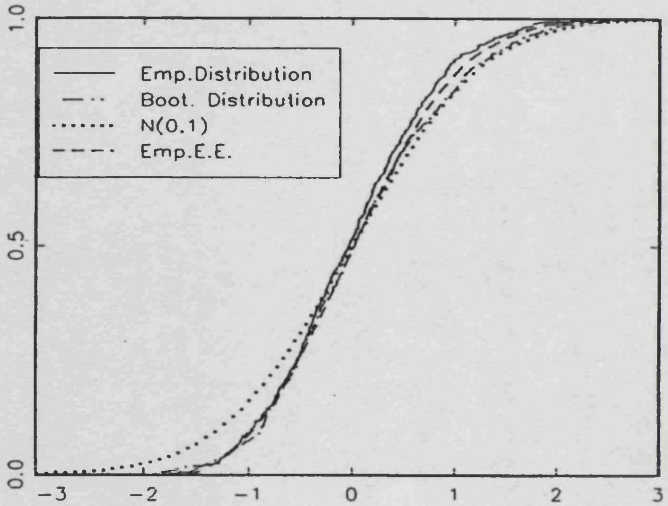


FIGURE 92. n=100 , h=0.2 , L=10 (Studentized case)

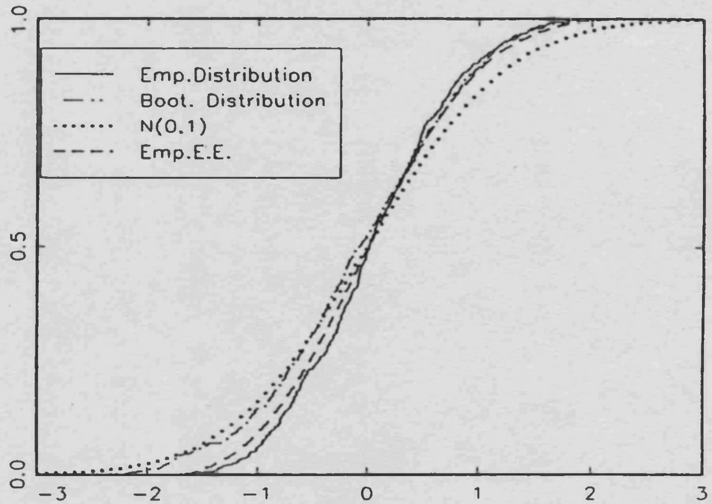


FIGURE 93. n=400 , h=0.8 , L=10 (Studentized case)

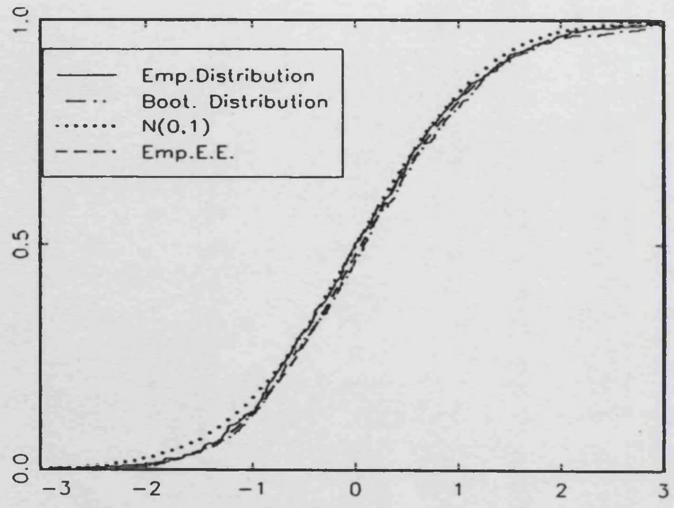


FIGURE 94. n=400 , h=0.6 , L=10 (Studentized case)

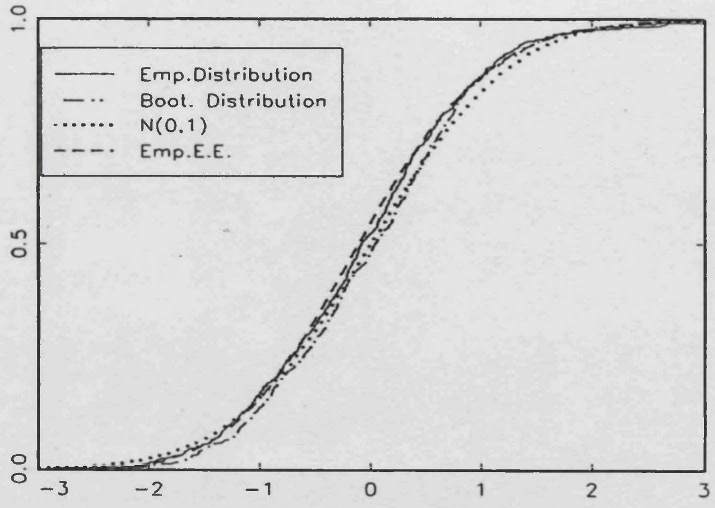


FIGURE 95. n=400 , h=0.4 , L=10 (Studentized case)

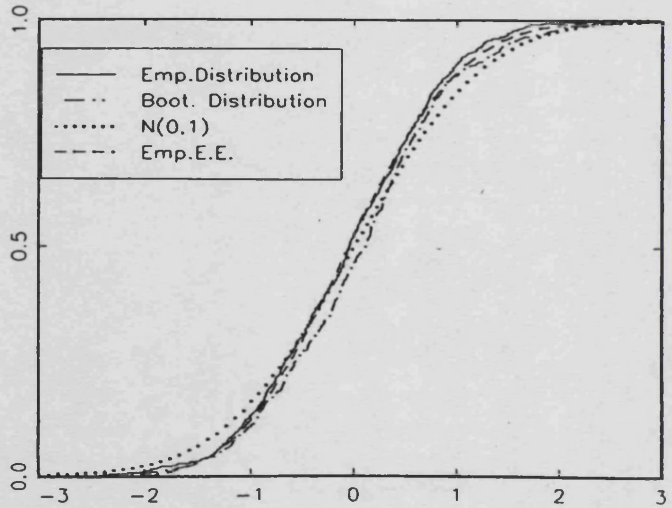
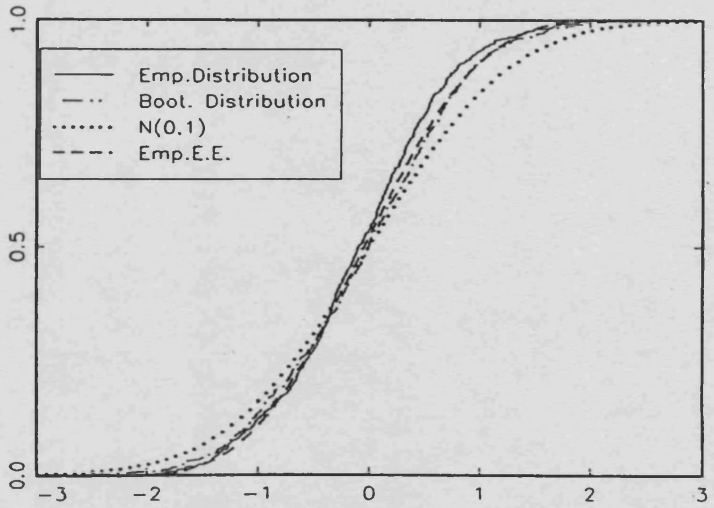


FIGURE 96. n=400 , h=0.2 , L=10 (Studentized case)



Appendix A : Established results in Robinson (1995a)

We reproduce some established results proved in Robinson (1995a), which are frequently used in the main proofs of Theorems and also Lemmas in Appendix B.

(a) Under (i), (iv), (vii) and (viii),

$$E(U) - \bar{\mu} = O(h^L) \quad . \tag{A.1}$$

(b) Under (i), (iv), (v), (vii) and (viii),

$$S - \Sigma = O(h^L) \quad . \tag{A.2}$$

(c) Under (i), (iv), (v), (vi), (vii) and (viii),

$$|E(V_2 W_{12} \,|1)\,| \le |E(V_2 \lambda^{\tau} U_{12} \,|1)\,| \le C(\,|Y_1| + 1) \quad \text{a.s.} \tag{A.3}$$

(d) Under (i), (iv), (v), (vi), (vii) and (viii),

$$|E(V_1 V_2 W_{12})\,| \le C \quad . \tag{A.4}$$

(e) Assuming $E|Y|^r < \infty$ for integer $r \ge 1$, (iv) and (viii),

$$E|W_{12}|^r \le C h^{-(r-1)d-r} \quad . \tag{A.5}$$

(f) Under (i), (iii), (iv), (vi), (viii) and (ix),

$$E|\sum_{j=2}^{n} W_{1j}\,|^r = O((nh^{-d-2})^{\frac{r}{2}}) \quad . \tag{A.6}$$

(g) Under (i), (iii), (iv), (vi), (viii) and (ix),

$$E|\bar{W}(m)\,|^r = O\left((\frac{m}{n^2 h^{d+2}})^{\frac{r}{2}} \right) ,$$

$$\text{especially } E|\bar{W}|^r = E|\bar{W}(n-1)\,|^r = O\left((\frac{1}{nh^{d+2}})^{\frac{r}{2}} \right) . \tag{A.7}$$

(a), (b), (d), (e) and (f) correspond to Lemmas 1, 2, 3, 4 and 6 of Robinson (1995a) respectively, while (c) was shown in the proof of Lemma 3 and (g) corresponds to equation (14) noting b_{3m} in Robinson (1995a) is equivalent to $\bar{W}(m)$ here.

Appendix B : Technical Lemmas

LEMMA 1: *Under* (i), (ii), (iii), (iv), (v), (vii), *and* (viii),

(a) $E(\nu_1) = 0, \quad Var(2\nu_1) = 1$,

(b) $E(V_1) = E(W_{12}) = 0$,

(c) $E(W_{12} \mid 1) = E(W_{12} \mid 2) = 0 \ a.s.$,

(d) $|\nu_1|^r + |V_1|^r \le C(|Y_1|^r + 1) \ a.s. \ for \ r \ge 0$.

PROOF: The proof for (a)-(c) is straightforward from the definitions. To prove (d) writing

$$\nu_1 = \sigma^{-1} \mathrm{v}^\tau \{ Y_1 f'(X_1) - e'(X_1) - \bar{\mu} \} ,$$
$$V_1 = \sigma^{-1} \mathrm{v}^\tau \left\{ Y_1 \int f'(X_1 - hu) K(u)\, du - \int e'(X_1 - hu) K(u)\, du - EU \right\} ,$$

since $\sigma^{-1} < C$ due to (ii), we have

$$|\nu_1|^r \le C\{ |Y_1|^r |\mathrm{v}^\tau f'(X_1)|^r + |\mathrm{v}^\tau e'(X_1)|^r + |\mathrm{v}^\tau \bar{\mu}|^r \} ,$$
$$|V_1|^r \le C\Big\{ |Y_1|^r \int |\mathrm{v}^\tau f'(X_1 - hu)|^r |K(u)|^r du$$
$$+ \int |\mathrm{v}^\tau e'(X_1 - hu)|^r |K(u)|^r du + |\mathrm{v}^\tau EU|^r \Big\} .$$

Apply (iv), (v), (viii), and (A.1). □

LEMMA 2: *Under* (iii), $\bar{\nu} - \bar{\nu}(m)$, $\bar{V} - \bar{V}(m)$, *and* $\bar{W} - \bar{W}(m)$ *are independent of* $(X_1^\tau, Y_1), \cdots, (X_m^\tau, Y_m)$.

PROOF: Straightforward. □

LEMMA 3: *Under* $E|Y_1|^r < \infty$, (ii), (iv), (v), (vii), *and* (viii),

$$E|V_1 - \nu_1|^r = O(h^{rL}) .$$

PROOF: Writing $|V_1 - \nu_1|^r \le C\{ |\mathrm{v}^\tau (U_1 - \mu_1)|^r + |\mathrm{v}^\tau (EU - \bar{\mu})|^r \}$ due to (ii), we have from (A.1) that the second term on the right is $O(h^{rL})$. Writing

$$\mathrm{v}^\tau (U_1 - \mu_1) = Y_1 \int \mathrm{v}^\tau \{ f'(X_1 - hu) - f'(X_1) \} K(u)\, du$$
$$- \int \mathrm{v}^\tau \{ e'(X_1 - hu) - e'(X_1) \} K(u)\, du ,$$

both integrals on the right are $O(h^L)$ a.s. by Taylor expansion and (viii) (see the proof of Lemma 2 of Robinson (1995a)) so that $|\nu^{\tau}(U_1-\mu_1)|^r \leq C(|Y_1|^r+1)h^{rL}$. Therefore

$$|V_1-v_1|^r \leq C(|Y_1|^r+1)h^{rL} \text{ a.s.} \tag{B.1}$$

Then apply $E|Y_1|^r<\infty$. □

LEMMA 4: *Under* (ii), (iii), (iv), (vi), *and* (viii),

$$E(|W_{12}|^r\,|1) \leq C(|Y_1|^r+1)h^{-(r-1)d-r} \quad a.s. \quad for \;\; 1\leq r\leq 3 .$$

PROOF: Using (ii) and an elementary inequality, write

$$E(|W_{12}|^r\,|1) \leq C\{E(|\nu^{\tau}U_{12}|^r\,|1)+|\nu^{\tau}EU|^r+|V_1|^r+E|V_2|^r\} .$$

(A.1) and Lemma 1-(d) give $|\nu^{\tau}EU|^r+|V_1|^r+E|V_2|^r \leq C(|Y_1|^r+1)$. The remaining term is bounded by

$$\frac{C}{h^{r(d+1)}}\{|Y_1|^r\int|\nu^{\tau}K'(\tfrac{X_1-x}{h})|^r f(x)\,dx$$

$$+\int E(|Y_2|^r\,|X_2=x)\,|\nu^{\tau}K'(\tfrac{X_1-x}{h})|^r f(x)\,dx\}$$

$$\leq \frac{Ch^d}{h^{r(d+1)}}\{|Y_1|^r\int|\nu^{\tau}K'(u)|^r dx+\int|\nu^{\tau}K'(u)|^r du\}$$

$$\leq \frac{C}{h^{(r-1)d+r}}(|Y_1|^r+1) ,$$

where the first inequality uses (iv) and (vi) and the second inequality uses (viii). □

LEMMA 5: *Under* (i), (ii), (iii), (iv), (v), (vi), (vii), *and* (viii),

$$E|(V_1-v_1)W_{12}| = O(h^{L-1}) .$$

PROOF: Using (B.1), Lemma 4 and (i),

$$E|(V_1-v_1)W_{12}|=E\{|V_1-v_1|E(|W_{12}|\,|1)\}\leq Ch^{L-1}E(|Y_1|^2+1)=O(h^{L-1}) .\ \square$$

LEMMA 6: *Under* (i), (ii), (iii), (iv), (vi), (vii), *and* (viii),

$$E|E(W_{13}W_{23}\,|1,2)|^r = O(h^{-(r-1)d-2r}) \quad for \;\; 1\leq r\leq 3 .$$

PROOF: $E(W_{13}W_{23}\,|1,2) = \sigma^{-2}E(\nu^{\tau}U_{13}\nu^{\tau}U_{23}\,|1,2)-\sigma^{-1}E(V_3\nu^{\tau}U_{13}\,|1)$

$$-\sigma^{-1}E(V_3\nu^{\tau}U_{23}\,|2)-\sigma^{-1}\nu^{\tau}EU(V_1+V_2)-V_1V_2+E(V_3^2)-(\sigma^{-1}\nu^{\tau}EU)^2 .$$

(ii) and (A.3) gives

$$|\sigma^{-1}E(V_3 v^\tau U_{13}|1)|^r + |\sigma^{-1}E(V_3 v^\tau U_{23}|2)|^r \le C(|Y_1|^r + |Y_2|^r + 1) \quad \text{a.s.},$$

and (A.1), (ii) and Lemma 1-(d) give

$$|\sigma^{-1}v^\tau EU(V_1+V_2) - V_1V_2|^r \le C(|Y_1|^r+1)(|Y_2|^r+1) \quad \text{a.s.},$$

so that their expectations are bounded under (i) and (iii). By Lemma 1-(d), (i), (ii) and (A.1), $|E(V_3^2) - (\sigma^{-1}v^\tau EU)^2| < \infty$. Finally, since

$$\begin{aligned}
&E[\, v^\tau U_{13} v^\tau U_{23}|1,2] \\
&= E\left[\frac{Y_1-Y_3}{h^{d+1}} v^\tau K'(\tfrac{X_1-X_3}{h}) \frac{Y_2-Y_3}{h^{d+1}} v^\tau K'(\tfrac{X_2-X_3}{h})\,|1,2\right] \\
&= h^{-2d-2}E\left[E\left\{(Y_3-Y_1)(Y_3-Y_2)\, v^\tau K'(\tfrac{X_1-X_3}{h})\, v^\tau K'(\tfrac{X_2-X_3}{h})\,|1,2,X_3\right\}\right] \\
&= h^{-2d-2}E\left[\{q(x)-(Y_1+Y_2)g(x)+Y_1Y_2\}\, v^\tau K'(\tfrac{X_1-X_3}{h})\, v^\tau K'(\tfrac{X_2-X_3}{h})\,|1,2\right] \\
&= h^{-d-2}\int\{q(u-hX_2)-(Y_1+Y_2)g(u-hX_2)+Y_1Y_2\} \\
&\qquad\qquad \times v^\tau K'(u-\tfrac{X_2-X_1}{h})\, v^\tau K'(u) f(u-hX_2)\,du \quad ,
\end{aligned}$$

using $\{q(u-hX_2) + |g(u-hX_2)|+1\}f(u-hX_2) \le C$ a.s. due to (iv) and (vi), we have

$$\begin{aligned}
|E(v^\tau U_{13} v^\tau U_{23}|1,2)|^r &\le \frac{C}{h^{r(d+2)}}(|Y_1|^r+1)(|Y_2|^r+1) \\
&\qquad \times\int\left|v^\tau K'(u-\tfrac{X_2-X_1}{h})\, v^\tau K'(u)\right|^r du \quad .
\end{aligned}$$

Taking expectation on both sides,

$$\begin{aligned}
E|E(v^\tau U_{13} v^\tau U_{23}|1,2)|^r &\le \frac{C}{h^{r(d+2)}}E\Big[(|Y_1|^r+1)\int\{E(|Y_2|^r|X_2=x)+1\} \\
&\qquad \times\int\left|v^\tau K'(u-\tfrac{x-X_1}{h})\, v^\tau K'(u)\right|^r du f(x)\,dx\Big] \\
&\le \frac{C}{h^{r(d+2)}}E\left[(|Y_1|^r+1)\iint|v^\tau K'(u-\tfrac{x-X_1}{h})|^r|v^\tau K'(u)|^r du\,dx\right] \\
&\le \frac{Ch^r}{h^{r(d+2)}}E\left[(|Y_1|^r+1)\iint|v^\tau K'(u-t)|^r|v^\tau K'(u)|^r du\,dt\right] \\
&= O(h^{-(r-1)d-2r}) \quad ,
\end{aligned}$$

where the first inequality uses $E(|Y_2|^r|X_2=x)f(x) \le C$ for $1 \le r \le 3$ due to (vi) and the last inequality uses (i) and (viii). □

LEMMA 7: *Under assumptions* (i), (ii), (iii), (iv), (v), (vii), *and* (viii),

$$E|\bar{\omega}(m)|^3 = O\left((\frac{m}{n})^{3/2} h^{3L} \right) \text{ for } m=1,\dots,n.$$

In particular, $E|\bar{\omega}| = E|\bar{\omega}(n)| = O(h^L)$.

PROOF: Since $(V_i - v_i)$ are iid by (iii) and $E(V_i - v_i) = 0$ by Lemma 1-(a),(b), applying the Theorem of DFJ, we have

$$E|\bar{\omega}(m)|^3 \le C\left(\frac{m}{n}\right)^{\frac{3}{2}} \left\{ \frac{1}{m}\sum_{i=1}^{m} E|V_i - v_i|^3 \right\} = \left(\frac{m}{n}\right)^{\frac{3}{2}} O(h^{3L}) \quad ,$$

by Lemma 3. Then $E|\bar{\omega}| = O(h^L)$ by Hölder's inequality. □

LEMMA 8: *Under* (i), (ii), (iii), (iv), (v), *and* (viii),

$$E[W_{12} w_1(t) w_2(t)] = \frac{4(it)^2}{n} E(W_{12} v_1 v_2) + O\left((\frac{t^4}{n^2} + \frac{|t|^3}{n^{3/2}}) h^{-1} \right) .$$

PROOF: Writing $\zeta_i = 2n^{-1/2} v_i$, by Lemma 1-(c) and (2.6) the left side is

$$\begin{aligned} &E[W_{12}\{w_1(t) - 1\}\{w_2(t) - 1\}] \\ &= E\Big[W_{12}\big\{(e^{it\zeta_1} - 1 - it\zeta_1)(e^{it\zeta_2} - 1 - it\zeta_2) + it\zeta_1(e^{it\zeta_2} - 1 - it\zeta_2) \\ &\qquad + it\zeta_2(e^{it\zeta_1} - 1 - it\zeta_1)\big\}\Big] + E[W_{12}(it\zeta_1)(it\zeta_2)] \\ &= \frac{4(it)^2}{n} E(W_{12} v_1 v_2) + O\left(\frac{t^4}{n^2} E(|W_{12}| v_1^2 v_2^2) + \frac{|t|^3}{n^{3/2}} E(|W_{12}| v_1^2 |v_2|) \right) . \end{aligned}$$

By an elementary inequality,

$$E|W_{12} v_1^2 v_2^2| \le \sigma^{-1} E|\nu^{\tau} U_{12} v_1^2 v_2^2| + E|V_1 v_1^2 v_2^2| + E|V_2 v_1^2 v_2^2| + \sigma^{-1} E|\nu^{\tau} EU v_1^2 v_2^2| .$$

The second, third and fourth terms on the right are bounded due to Lemma 1-(d) and (A.1). Using (ii) and an elementary inequality, the remaining term is bounded by

$$\begin{aligned} &\frac{C}{h^{d+2}} E\left\{ |\nu^{\tau} K'(\tfrac{X_1 - X_2}{h})| (|Y_1| + |Y_2|)(Y_1^2 + 1)(Y_2^2 + 1) \right\} \\ &\le \frac{C}{h^{d+2}} E\left\{ |\nu^{\tau} K'(\tfrac{X_1 - X_2}{h})| (|Y_1|^3 + |Y_1|^2 |Y_2| + |Y_1|) \right\} \\ &\le \frac{C}{h^{d+2}} E\int |\nu^{\tau} K'(\tfrac{X_1 - x}{h})| (|Y_1|^3 + |Y_1|^2 E(|Y_2| \, | X_2 = x) + |Y_1|) f(x)\, dx \\ &\le \frac{C}{h^{d+2}} E(|Y_1|^3 + |Y_1|^2 + |Y_1|) \int |\nu^{\tau} K'(\tfrac{X_1 - x}{h})| dx \\ &\le \frac{Ch^d}{h^{d+2}} E(|Y_1|^3 + |Y_1|^2 + |Y_1|) \int |\nu^{\tau} K'(u)| du \\ &= O(h^{-1}) \quad , \end{aligned} \tag{B.2}$$

where the third inequality uses the boundedness of f and $E(\,|Y|\,|X)f$ due to (iv) and (vi). Therefore

$$E|W_{12}v_1^2v_2^2| = O(h^{-1}) \tag{B.3}$$

Similarly, $E(\,|W_{12}\,|v_1^2|v_2|) \leq Ch^{-1}$. □

LEMMA 9: *Under* (i), (ii), (iii), (iv), (v), *and* (viii),

$$E\left[W_{12}^2 w_1(t)\,w_2(t)\right] = E(W_{12}^2) + O\left(\frac{|t|}{n^{1/2}h^{d+2}}\right) .$$

PROOF: The left side is

$$\begin{aligned} E(W_{12}^2) + E\left[W_{12}^2\{e^{it(\zeta_1+\zeta_2)}-1\}\right] &= E(W_{12}^2) + O(|t|\,E|W_{12}^2\zeta_1|) \\ &= E(W_{12}^2) + O\left(\frac{|t|}{n^{1/2}}E|W_{12}^2 v_1|\right) \end{aligned}$$

using (2.6), and by Lemmas 1-(d) and 4 and (i),

$$E|W_{12}^2 V_1| = E\{\,|v_1|E(\,|W_{12}|^2|1)\,\} \leq Ch^{-d-2}E(\,|Y_1|^3+1) = O(h^{-d-2}) .\ \square$$

LEMMA 10: *Under* (i), (ii), (iii), (iv), (v), *and* (viii),

$$E\left[W_{12}W_{13}w_1(t)\,w_2(t)\,w_3(t)\right] = O\left(\frac{|t|^3}{n^{3/2}h^2}\right) .$$

PROOF: By Lemma 1-(c) the left side is

$$\begin{aligned} &E\left[W_{12}W_{13}(e^{it\zeta_1}-1)(e^{it\zeta_2}-1)(e^{it\zeta_3}-1)\right] \\ &= O(|t|^3E(\,|W_{12}||W_{13}||\zeta_1||\zeta_2||\zeta_3|)) \\ &= O\left(\frac{|t|^3}{n^{3/2}}E(\,|W_{12}||W_{13}||v_1||v_2||v_3|)\right) \end{aligned}$$

using (2.6). Writing

$$E(\,|W_{12}||W_{13}||v_1||v_2||v_3|) = E\{\,|W_{12}||v_1||v_2|E(\,|W_{13}v_3|\,|1)\,\} , \tag{B.4}$$

similarly to the proof of Lemma 8,

$$\begin{aligned} E(\,|W_{13}v_3|\,|1) &\leq \frac{C}{h^{d+1}}E\left\{\,|v^\tau K'(\tfrac{X_1-X_3}{h})\,|(\,|Y_1|+|Y_3|)(\,|Y_3|+1)\,\,|1\right\}+C \\ &\leq \frac{C}{h^{d+1}}|Y_1|E\left\{\,|v^\tau K'(\tfrac{X_1-X_3}{h})\,|(\,|Y_3|+1)\,\,|1\right\} \\ &\quad + \frac{C}{h^{d+1}}E\left\{\,|v^\tau K'(\tfrac{X_1-X_3}{h})\,|(Y_3^2+|Y_3|)\,\,|1\right\}+C \\ &\leq \frac{C}{h}(\,|Y_1|+1) , \end{aligned}$$

where the third inequality uses the same method as (B.2), so that (B.4) is smaller than

$$\frac{C}{h}E\{\,|W_{12}||v_1||v_2|(\,|Y_1|+1)\,\}$$

and it is $O(h^{-2})$ as in (B.3). □

LEMMA 11: *Under* (i), (ii), (iii), (iv), (vii), *and* (viii),

$$\frac{\sigma^{-1}\nu^{\tau}(EU-\bar{\mu})}{h^{L}} = \kappa_1 + o(1) \quad as \ n \to \infty .$$

PROOF: Writing

$$\begin{aligned}\sigma^{-1}\nu^{\tau}EU &= h^{-d-1}\sigma^{-1}E\{\nu^{\tau}K'(\tfrac{X_1-X_2}{h})(Y_1-Y_2)\} = 2h^{-d-1}\sigma^{-1}E\{\nu^{\tau}K'(\tfrac{X_1-X_2}{h})Y_1\}\\ &= 2h^{-d-1}\sigma^{-1}E\left\{Y_1\int \nu^{\tau}K'(\frac{X_1-x}{h})f(x)\,dx\right\}\\ &= 2h^{-1}\sigma^{-1}E\left\{g(X_1)\int \nu^{\tau}K'(u)f(X_1-hu)\,du\right\}\\ &= 2\sigma^{-1}E\left\{g(X_1)\int \nu^{\tau}f'(X_1-hu)K(u)\,du\right\},\end{aligned}$$

where the second equality uses (iii) and that K is even, and the fifth equality uses (vii) and (viii), Young's form of Taylor's Theorem (see e.g. Serfling, 1980, p45) gives

$$\sigma^{-1}\nu^{\tau}(EU-\bar{\mu}) = b_1h + \cdots + b_{L-1}h^{L-1} + b_Lh^{L} + o(h^{L})$$

where

$$b_l = \frac{2(-1)^{l}\sigma^{-1}}{l\,!}\sum_{\substack{0\le l_1,\cdots,l_d\le l\\ l_1+\cdots+l_d=l}}\cdots\sum\left\{\int\prod_{i=1}^{d}u_i^{l_i}K(u)\,du\right\}\nu^{\tau}E[\{\Delta^{(l_1,\cdots,l_d)}f\}g],$$

$l=1,\cdots,L$. But $b_1=\cdots=b_{L-1}=0$ by (viii) and $b_L=\kappa_1$. □

LEMMA 12: *Under* (i), (ii), (iii), (iv), (v), (vi), (vii), *and* (viii),

$$h^{d+2}E(W_{12}^2) = \kappa_2 + o(1) \quad as \ n \to \infty .$$

PROOF: From Lemma 1-(b),(c),

$$E(W_{12}^2) = E(\sigma^{-1}\nu^{\tau}U_{12})^2 - 2E(V_1^2) - (\sigma^{-1}\nu^{\tau}EU)^2 .$$

By (i), Lemma 1-(d), and (A.1), $E(V_1^2) + |\sigma^{-1}\nu^{\tau}EU| \le C$ for sufficiently large n. Also

$$\begin{aligned}E(\sigma^{-1}\nu^{\tau}U_{12})^2 &= E[E\{(\sigma^{-1}\nu^{\tau}U_{12})^2|1\}]\\ &= h^{-2d-2}\sigma^{-2}E\left[\int\{\nu^{\tau}K'(\tfrac{X_1-x}{h})\}^2\{Y_1^2-2Y_1g(x)+q(x)\}f(x)\,dx\right]\\ &= h^{-d-2}\sigma^{-2}E\left[\int\{\nu^{\tau}K'(u)\}^2f(X-hu)\{q-2gg(X-hu)+q(X-hu)\,du\right]\end{aligned}$$

and the expectation converges, as $n\to\infty$ (i.e. as $h\to 0$), to κ_2 by (iv) - (vi). □

LEMMA 13: *Under* (i), (ii), (iii), (iv), (v), (vii) *and* (viii),

(a) $E(v_1^3) = \kappa_3$,

(b) $E(V_1^3) = \kappa_3 + o(1)$ *as* $n \to \infty$.

PROOF: To prove (a), putting $b = \mathrm{v}^\tau f'$, $c = \mathrm{v}^\tau a$, routine calculation gives

$$E(v_1^3) = \sigma^{-3}[E\{(Y-g)^3 b^3\} - 3E\{(Y-g)^2 b^2 c\} + 3E\{(Y-g) bc^2\} - Ec^3]$$
$$= \sigma^{-3}E[\{r - 3(q-g^2)g - g^3\}b^3 - 3(q-g^2)b^2c - c^3]$$
$$= \sigma^{-3}E[(r - 3qg + 2g^3)b^3 - 3(q-g^2)b^2c - c^3]$$

because by straight forward calculation,

$$E[(Y-g)^3 b^3] = E[\{r - 3(q-g^2)g - g^3\}b^3] ,$$
$$E[(Y-g)^2 b^2 c] = E\{(q-g^2)b^2c\} , \quad E[(Y-g)bc^2] = 0 .$$

To verify (b), it suffices to show $E(V_1^3 - v_1^3) = o(1)$. Lemma 1-(d) and (B.1) give

$$E|V_1^3 - v_1^3| \le CE|V_1 - v_1|(V_1^2 + v_2^2) \le Ch^L E(|Y_1|^3 + 1) = O(h^L) . \quad \square$$

LEMMA 14: *Under* (i), (ii), (iii), (iv), (v), (vi), (vii), *and* (viii),

(a) $E(W_{12} v_1 v_2) = \kappa_4 + o(1)$ *as* $n \to \infty$,

(b) $E(W_{12} V_1 V_2) = \kappa_4 + o(1)$ *as* $n \to \infty$.

PROOF: Using Lemma 1-(a), putting $a_i = a(X_i)$, $K'_{ij} = K'(\frac{X_i - X_j}{h})$,

$$E(W_{12} v_1 v_2) = \sigma^{-1}E[\mathrm{v}^\tau(U_{12} - EU)v_1 v_2] - E[V_1 v_1 v_2] - E[V_2 v_1 v_2]$$
$$= \sigma^{-1}E(\mathrm{v}^\tau U_{12} v_1 v_2)$$
$$= h^{-d-1}\sigma^{-3}E\left[\mathrm{v}^\tau K'_{12}(Y_1 - Y_2)\,\mathrm{v}^\tau\{(Y_1 - g_1)f'_1 - a_1\}\mathrm{v}^\tau\{(Y_2 - g_2)f'_2 - a_2\}\right]$$
$$= h^{-d-1}\sigma^{-3}E\Big(\mathrm{v}^\tau K'_{12}\,\mathrm{v}^\tau\{(Y_1 - g_1)f'_1 - a_1\}$$
$$\times\big[Y_1 \mathrm{v}^\tau\{(Y_2 - g_2)f'_2 - a_2\} - Y_2 \mathrm{v}^\tau\{(Y_2 - g_2)f'_2 - a_2\}\big]\Big) .$$

Taking expectations with respect to Y_2 given X_1, X_2, Y_1 , it equals

$$h^{-d-1}\sigma^{-3}E\left[\mathrm{v}^\tau K'_{12}\,\mathrm{v}^\tau\{(Y_1 - g_1)f'_1 - a_1\}\{-Y_1 \mathrm{v}^\tau a_2 - \mathrm{v}^\tau\{(q_2 - g_2^2)f'_2 - a_2 g_2\}\}\right]$$
$$= h^{-1}\sigma^{-3}E\int \mathrm{v}^\tau K'(u)\,\mathrm{v}^\tau\{(Y-g)f' - a\}[-Y\mathrm{v}^\tau a(X-hu)$$
$$-\{q(X-hu) - g(X-hu)^2\}\mathrm{v}^\tau f'(X-hu) + \mathrm{v}^\tau a(X-hu)g(X-hu)]f(X-hu)\,du .$$

Since $g'f$ vanishes on the boundary of its support by assumption (vii), integration by parts gives the above quantity is equal to

$$\sigma^{-3}E\int K(u)\,\mathrm{v}^{\tau}\{(Y-g)f'-a\}$$
$$\times\Big\{[-Y\mathrm{v}^{\tau}a'(X-hu)\,\mathrm{v}-\mathrm{v}^{\tau}\{q'(X-hu)-2g(X-hu)g'(X-hu)\}f'(X-hu)^{\tau}\mathrm{v}$$
$$-\{q(X-hu)-g(X-hu)^2\}\mathrm{v}^{\tau}f''(X-hu)\,\mathrm{v}$$
$$+\mathrm{v}^{\tau}a'(X-hu)\,\mathrm{v}g(X-hu)+\mathrm{v}^{\tau}a(X-hu)g'(X-hu)^{\tau}\mathrm{v}]f(X-hu)$$
$$-[Y\mathrm{v}^{\tau}a(X-hu)+\{q(X-hu)-g(X-hu)^2\}\mathrm{v}^{\tau}f'(X-hu)-\mathrm{v}^{\tau}a(X-hu)g(X-hu)]$$
$$\times\mathrm{v}^{\tau}f'(X-hu)\Big\}du\ .$$

Using assumptions (iv),(v) and (vi), which imply the continuity of f, f', f'', g, g', q, q', a and a', as $n\to\infty$, i.e., $h\to 0$,

$$E(W_{12}v_1v_2)$$
$$\to\sigma^{-3}E\Big[\{b(Y-g)-c\}\{-[(\mathrm{v}^{\tau}a'\mathrm{v})(Y-g)+\{\mathrm{v}^{\tau}(q'-2gg')\}b$$
$$+(q-g^2)(\mathrm{v}^{\tau}f''\mathrm{v})-c(\mathrm{v}^{\tau}g')]f-[c(Y-g)+(q-g^2)b]b\}\Big]$$
$$=\sigma^{-3}E\Big[\{b(Y-g)-c\}\{-(f\mathrm{v}^{\tau}a'\mathrm{v}+bc)(Y-g)-fb\{\mathrm{v}^{\tau}(q'-2gg')\}$$
$$-f(q-g^2)(\mathrm{v}^{\tau}f''\mathrm{v})+fb(\mathrm{v}^{\tau}g')-(q-g^2)b^2\}\Big]$$
$$=-\sigma^{-3}E\Big[f(q-g^2)b(\mathrm{v}^{\tau}a'\mathrm{v})-fb\{\mathrm{v}^{\tau}(q'-2gg')\}c$$
$$-f(q-g^2)c(\mathrm{v}^{\tau}f''\mathrm{v})+f(\mathrm{v}^{\tau}g')c^2\Big]\ .$$

To prove (b), it suffices to show $E(W_{12}V_1V_2)=E(W_{12}v_1v_2)+o(1)$. Writing

$$|E(W_{12}V_1V_2)-E(W_{12}v_1v_2)|\le|E\{W_{12}(V_1-v_1)V_2\}|+|E\{W_{12}v_1(V_2-v_1)\}|\ ,$$

the first term on the right is equal to

$$|E\{(V_1-v_1)E(W_{12}V_2|1)\}|\le E\{|V_1-v_1||E(W_{12}V_2|1)|\}$$
$$\le Ch^LE(|Y_1|^2+1)=O(h^L)\ ,$$

by (B.1), (A.3) and (i). We can handle the second term similarly and show it is also $O(h^L)$. □

LEMMA 15: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|V_1W_{12}|^r=E|V_2W_{12}|^r=O(h^{-(r-1)d-r})\ \ for\ 1\le r\le 3\ .$$

PROOF. Using Lemmas 1-(d) and 4,

$$E|V_1W_{12}|^r\le E\{|V_1|^rE(|W_{12}|^r|1)\}$$
$$\le CE\{(|Y_1|^r+1)^2\}h^{-(r-1)d-r}$$

$$\leq Ch^{-(r-1)d-r} \text{ for } 1 \leq r \leq 3 \text{ under (i)'.}$$

$E|V_1 W_{12}|^r = E|V_2 W_{12}|^r$ is obvious by symmetry of W_{12} and (iii). □

LEMMA 16: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|\tilde{V}_1|^r = O(1) \quad for \ 1 \leq r \leq 6 \ .$$

PROOF. By (A.3),

$$|V_1|^r = |E(V_2 W_{12}|1)|^r \leq C(|Y_1|^r + 1) \quad \text{a.s.} \tag{B.5}$$

so (i)' immediately produces the conclusion.
so (i)' immediately produces the conclusion. □

LEMMA 17. *Under* (i)', (ii), (iii), (iv), (v), (vi) *and* (viii),

(a) $E|\tilde{W}_{11}|^r = O(h^{-r(d+2)}) \quad for \ 1 \leq r \leq 3$,

(b) $E|\tilde{W}_{12}|^r = O(h^{-(r-1)d-2r}) \quad for \ 1 \leq r \leq 3$.

PROOF.

(a). $\tilde{W}_{11} = E(W_{12}^2|1) \leq C(|Y_1|^2 + 1)h^{-d-2}$ a.s. by Lemma 4 so again application of (i)' completes the proof.

(b). Apply Lemma 6. □

LEMMA 18: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii), *for* d_1 *given in* (3.36),

(a) $E|d_1 V_2|^r = O(1) \quad for \ 1 \leq r \leq 3$,

(b) $E|d_1 V_1|^r = O(1) \quad for \ 1 \leq r \leq 2$.

PROOF.

(a) By (iii), $E|d_1 V_2|^r = E|d_1|^r E|V_2|^r$, where the second factor is bounded due to Lemma 1-(d). From Lemma 1-(d) and (B.5),

$$|d_1|^r \leq C\left(|V_1^2|^r + |\tilde{V}_1|^r + 1\right) \leq C(|Y_1|^{2r} + 1) \ , \tag{B.6}$$

then apply (i)'. □

(b) By an elementary inequality and (3.8),

$$E|d_1V_1|^r \le C(E|V_1^3|^r + E|\tilde{V}_1V_1|^r + E|V_1|^r) \ .$$

By Lemma 1-(d) and (B.5), $E|V_1^3|^r + E|V_1|^r = O(1)$ for $1 \le r \le 2$, and

$$E|\tilde{V}_1V_1|^r \le CE(|Y_1|^r + 1)^2 = O(1) \tag{B.7}$$

for $1 \le r \le 3$ by (i)'. □

LEMMA 19: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

(a) $E|W_{12}V_1V_3|^r = E|W_{12}V_2V_3|^r = O(h^{-(r-1)d-r})$ *for* $1 \le r \le 3$,

(b) $E|W_{12}V_1^2|^r = E|W_{12}V_2^2|^r = O(h^{-(r-1)d-r})$ *for* $1 \le r \le 2$.

PROOF.

(a). Using (iii), Lemma 1-(d) and Lemma 15, for $1 \le r \le 3$,

$$E|W_{12}V_1V_3|^r = E|W_{12}V_1|^r E|V_3|^r = O(h^{-(r-1)d-r}) \ .$$

$E|W_{12}V_1V_3|^r = E|W_{12}V_2V_3|^r$ is straightforward by (iii) and symmetry of W_{12} .

(b). By Lemmas 1-(d) and 4, the left side is

$$E\left\{|V_1|^{2r}E(|W_{12}|^r|1)\right\} \le E\left\{C(|Y_1|^{3r}+1)h^{-(r-1)d-r}\right\} = O(h^{-(r-1)d-r}) \ .$$

$E|W_{12}V_1^2|^r = E|W_{12}V_2^2|^r$ is straightforward by (iii) and symmetry of W_{12} . □

LEMMA 20: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii), *with* e_{12} *given in* (3.36),

(a) $E|e_{12}V_3|^r = O(h^{-(r-1)d-2r})$ *for* $1 \le r \le 3$,

(b) $E|e_{12}V_1|^r = E|e_{12}V_2|^r = O(h^{-(r-1)d-2r})$ *for* $1 \le r \le 2$.

PROOF.

(a). By (iii) and Lemma 1-(d), write

$$E|e_{12}V_3|^r = E|e_{12}|^r E|V_3|^r \le C(E|V_1W_{12}|^r + E|\tilde{V}_1|^r + E|\tilde{W}_{12}|^r) \ .$$

Then apply Lemmas 15, 16 and 17-(b).

(b). An elementary inequality gives

$E|e_{12}V_1|^r$

$$\leq C\left(E|V_1V_2W_{12}|^r + E|V_1^2W_{12}|^r + E|\tilde{V}_1V_1|^r + E|\tilde{V}_1V_2|^r + E|\tilde{W}_{12}V_1|^r \right) . \quad (B.8)$$

Writing due to (ii)

$$E|V_1V_2W_{12}|^r \leq C(E|V_1V_2\nu^\tau U_{12}|^r + E|V_1^2V_2| + E|V_1V_2^2| + E|V_1V_2||\nu^\tau EU|) ,$$

the last three terms on the right are bounded due to Lemma 1-(d) and (A.1). Using symmetry of $\nu^\tau K'(\frac{X_1-X_2}{h})$, Lemma 1-(d) and an elementary inequality,

$$E|V_1V_2\nu^\tau U_{12}|^r$$
$$\leq \frac{C}{h^{r(d+2)}} E\left\{ (|Y_1|^r + 1) (|Y_2|^r + 1) (|Y_1|^r + |Y_2|^r) |\nu^\tau K'(\frac{X_1-X_2}{h}) |^r \right\}$$
$$= \frac{2C}{h^{r(d+2)}} E\left\{ (|Y_1|^{2r} |Y_2|^r + |Y_1|^r |Y_2|^r + |Y_1|^{2r} + |Y_1|^r) |\nu^\tau K'(\frac{X_1-X_2}{h}) |^r \right\} .$$

Expanding the right hand side, we can bound the first term by

$$E\left\{ |Y_1|^{2r} |Y_2|^r |\nu^\tau K'(\frac{X_1-X_2}{h}) |^r \right\}$$
$$= E\left\{ |Y_1|^{2r} \int E(|Y_2|^r |X_2=x) |\nu^\tau K'(\frac{X_1-x}{h}) |^r f(x) dx \right\}$$
$$\leq CE\left\{ |Y_1|^{2r} \int |\nu^\tau K'(\frac{X_1-x}{h}) |^r dx \right\} \leq Ch^d E\left\{ |Y_1|^{2r} \int |\nu^\tau K'(u) |^r du \right\}$$
$$\leq Ch^d ,$$

where the first inequality uses (vi) and the last inequality uses (i) and (viii). The other three terms are also $O(h^d)$ similarly. Thus, for $1 \leq r \leq 3$,

$$E|V_1V_2W_{12}|^r = O(h^{-(r-1)d-r}) . \quad (B.9)$$

The second term in (B.8) has the same order bound as (B.9) by Lemma 19-(b) for $1 \leq r \leq 2$. The third term in (B.8) is bounded due to (B.7), while the fourth term is bounded due to Lemma 1-(d) and Lemma 16. We handle the last term in (B.8) similarly to Lemma 6,

$$E|\tilde{W}_{12}V_1|^r = E|E(W_{13}W_{23} |1, 2) V_1|^r \leq CE\{ |E(\nu^\tau U_{13} \nu^\tau U_{23} |1, 2) |^r (|Y_1|^r + 1) \}$$
$$= O(h^{-(r-1)d-2r}) . \quad (B.10)$$

$E|e_{12}V_1|^r = E|e_{12}V_2|^r$ is straightforward by (iii) and symmetry of e_{12} . □

LEMMA 21: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

(a) $E|d_1W_{23}|^r = O(h^{-(r-1)d-r})$ *for* $1 \leq r \leq 3$,

(b) $E|d_1 W_{12}|^r = E|d_2 W_{12}|^r = O(h^{-(r-1)d-r})$ *for* $1 \le r \le 2$.

PROOF.

(a). Using (B.6) and (A.5),

$$E|d_1 W_{23}|^r = E|d_1|^r E|W_{23}|^r = O(h^{-(r-1)d-r})$$

for $1 \le r \le 3$ under (i)'.

(b). Using (B.6) and Lemma 4 the left side is

$$E\left\{|d_1|^r E(|W_{12}|^r|1)\right\} \le E\{(|Y_1|^{2r}+1)\, C(|Y_1|^r+1)\, h^{-(r-1)d-r}\}$$
$$\le CE(|Y_1|^{3r}+1)\, h^{-(r-1)d-r} = O(h^{-(r-1)d-r})$$

for $1 \le r \le 2$ under (i)'. $E|d_1 W_{12}|^r = E|d_2 W_{12}|^r$ is straightforward by (iii) and symmetry of W_{12} . □

LEMMA 22: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

(a) $E|\tilde{W}_{12} W_{12}|^r = O(h^{-(2r-1)d-3r})$ *for* $1 \le r \le 3$,

(b) $E|\tilde{W}_{12} W_{13}|^r = O(h^{-2(r-1)d-3r})$ *for* $1 \le r \le 3$,

(c) $E|\tilde{W}_{12} W_{23}|^r = O(h^{-2(r-1)d-3r})$ *for* $1 \le r \le 3$,

(d) $E|\tilde{W}_{12} W_{34}|^r = O(h^{-2(r-1)d-3r})$ *for* $1 \le r \le 6$.

PROOF.

(a). In view of the proof of Lemma 6,

$$E|\tilde{W}_{12} W_{12}|^r = E|\tilde{W}_{12}|^r |W_{12}|^r$$
$$\le h^{-r(d+2)} CE(1+|Y_1|^r+|Y_2|^r+|Y_1|^r|Y_2|^r)\,|W_{12}|^r$$
$$\le Ch^{-r(d+2)}\left(E|W_{12}|^r + E|Y_1 W_{12}|^r + E|Y_2 W_{12}|^r + E|Y_1 Y_2 W_{12}|^r\right) .$$

The first term in parentheses is $O(h^{-(r-1)d-r})$ by (A.5). From inspecting their proofs Lemma 15 and (B.9) still hold with V_1 and V_2 replaced by Y_1 and Y_2 so that the other terms are $O(h^{-(r-1)d-r})$ for $1 \le r \le 3$.

(b). Using Lemma 4, for $1 \le r \le 3$,

$$E|\tilde{W}_{12} W_{13}|^r = E\left\{|\tilde{W}_{12}|^r E(|W_{13}|^r|1, 2)\right\}$$

$$\leq E\left\{|\tilde{W}_{12}|^r C(|Y_1|^r + 1) h^{-(r-1)d-r}\right\}$$

$$= Ch^{-(r-1)d-r}\left(E|\tilde{W}_{12}Y_1|^r + E|\tilde{W}_{12}|^r\right) .$$

We may replace V_1 by Y_1 in (B.10), so that using also Lemma 17-(b),

$$E|\tilde{W}_{12}W_{13}|^r = O(h^{-2(r-1)d-3r}) \text{ for } 1 \leq r \leq 3 .$$

(c). The proof is as in (b).

(d). Writing $E|\tilde{W}_{12}W_{34}|^r = E|\tilde{W}_{12}|^r E|W_{12}|^r$ by (iii), the proof is straightforward by (A.5) and Lemma 17-(b). □

LEMMA 23: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|\bar{V}|^r = O(1) \text{ for } 2 \leq r \leq 6 .$$

PROOF. Since V_i , $i=1,\ldots,n$ is an iid sequence, the result follows straightforwardly by DFJ and Lemma 1-(d). □

LEMMA 24: *Under* (i)', (ii), (v), (vi), (vii) *and* (viii)

$$|T_1|^r = O(n^{-r}h^{-r(d+2)}) \text{ for } r > 0 .$$

PROOF. Using (A.5) and $|\delta| < C$ due to (3.8),

$$|T_1|^r \leq \frac{C}{n^r}|E(W_{12}^2)|^r = O(n^{-r}h^{-r(d+2)}) .$$ □

LEMMA 25: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|T_2|^r = O(n^{-\frac{r}{2}}) \text{ for } 2 \leq r \leq 3 .$$

PROOF. Using (3.8), write

$$E|T_2|^r \leq \frac{C}{n^r}E|\sum_{i=1}^n (4V_i^2 - s^2)|^r + \frac{C}{n^r}E|\sum_{i=1}^n 8\tilde{V}_i|^r .$$

Since $E(4V_i^2) = s^2$ and $E(\tilde{V}_i) = 0$, by (iii) both the $4V_i^2 - s^2$ and $\tilde{V}_i$ are martingale differences and thus the theorem of DFJ applies to yield

$$E|\sum_{i=1}^n (4V_i^2 - s^2)|^r \leq Cn^{\frac{r}{2}}E|4V_1^2 - s^2|^r = O(n^{\frac{r}{2}})$$

for $2 \le r \le 3$ by (3.8) and Lemma 1-(d) and

$$E|\sum_{i=1}^{n} 8\tilde{V}_i|^r \le Cn^{\frac{r}{2}}E|\tilde{V}_1|^r = O(n^{\frac{r}{2}})$$

by Lemma 16. □

LEMMA 26: *Under* (i)', (ii), (iii), (iv), (v), (vi) *and* (viii)

$$E|T_3|^r = O(n^{-r}h^{-(r-1)d-2r}) \quad \textit{for} \quad 2 \le r \le 6 \ .$$

PROOF. Using (3.8), write

$$E|T_3|^r \le Cn^{-2r}E|\sum_{k=1}^{n-1} Z_k|^r \tag{B.11}$$

where $Z_k = \sum_{m=k+1}^{n} \tilde{W}_{km}$, for $k=1, \ldots, n$-1. Since

$$E(\tilde{W}_{12}|2) = E(\tilde{W}_{12}|1) = E\{E(W_{13}W_{23}|1,2)|1\} = E(W_{13}W_{23}|1) = 0 \ ,$$

Z_k , $k = n$-1, ... ,1 is a martingale difference sequence. Thus we apply DFJ to bound (B.11) by $Cn^{-2r}(n-1)^{\frac{r}{2}-1}\sum_{k=1}^{n-1}E|Z_k|^r$. Since $E(\tilde{W}_{km}|m) = 0$ for $m = k+1,\ldots,n$, $\tilde{W}_{km}$ are martingale differences. We use DFJ again and get by Lemma 17-(b),

$$E|Z_k|^r \le C(n-k)^{\frac{r}{2}-1}\sum_{m=k+1}^{n} E|\tilde{W}_{km}|^r \le C(n-k)^{\frac{r}{2}-1}(n-k)\, h^{-(r-1)d-2r}$$

$$= O(n^{\frac{r}{2}}h^{-(r-1)d-2r})$$

so that $(n-1)^{\frac{r}{2}-1}\sum_{k=1}^{n-1}E|Z_k|^r = O(n^r h^{-(r-1)d-2r})$. Therefore (B.11) is $O(n^{-r}h^{-(r-1)d-2r})$. □

LEMMA 27: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|Q_1|^r = O(n^{-r}h^{-(r-1)d-r}) \quad \textit{for} \quad 2 \le r \le 3 \ .$$

PROOF. Write $P_{ij} = (V_i + V_j)W_{ij} - \tilde{V}_i - \tilde{V}_j$. Then $\sum_{j=i+1}^{n} P_{ij}$ is a martingale difference sequence for $i = n$-1, ... ,1. We can proceed by replacing $\tilde{W}_{km}$ in Lemma 26 by P_{ij} due to the property $E(P_{ij}|j) = 0$ for $i \ne j$. Applying DFJ and (3.8),

$$E|Q_1|^r \le C\binom{n}{2}^{-r}E\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P_{ij}\right|^r \le C\binom{n}{2}^{-r}n^{\frac{r}{2}-1}\sum_{i=1}^{n-1}E|\sum_{j=i+1}^{n} P_{ij}|^r \ .$$

Since P_{ij} , $j = n, \ldots, i+1$ is a martingale difference for fixed i, we can apply the theorem of DFJ again and obtain $E|\sum_{j=i+1}^{n} P_{ij}|^r \le C(n-i)^{\frac{r}{2}-1} \sum_{j=i+1}^{n} E|P_{ij}|^r$. By Lemmas 15 and 16,

$$E|P_{ij}|^r \le C[E|\tilde{V}_i|^r + E|V_i W_{ij}|^r] = O(h^{-(r-1)d-r}) \quad \text{for } 1 \le r \le 3 \ . \qquad \square$$

LEMMA 28: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|Q_2|^r = O(n^{-r} h^{-(r-1)d-r}) \quad \textit{for} \ \ 2 \le r \le 6 \ .$$

PROOF. By an elementary inequality and (3.8),

$$E|Q_2|^r \le \frac{C}{n^r}\binom{n-1}{2}^{-r} E|\sum_{i=1}^{n} \sum_{k=1}^{n-1\,(i)} \sum_{m=k+1}^{n\,(i)} V_i W_{km}|^r$$

$$= \frac{C}{n^r}\binom{n-1}{2}^{-r} n^{r-1} \sum_{i=1}^{n} E| \sum_{k=1}^{n-1\,(i)} \sum_{m=k+1}^{n\,(i)} V_i W_{km}|^r \ .$$

$V_i W_{km}$, $m=k+1,\ldots,n$ is a martingale difference for fixed i, k, $k \ne i$ and $m \ne i$, and $\sum_{m=k+1}^{n\,(i)} V_i W_{km}$, $k=n-1,\ldots,1$ is also a martingale difference for fixed i and $k \ne i$ so that we apply DFJ repeatedly as in the proof of the previous Lemma and get

$$\sum_{i=1}^{n} E| \sum_{k=1}^{n-1\,(i)} \sum_{m=k+1}^{n\,(i)} V_i W_{km}|^r \le C \sum_{i=1}^{n} (n-2)^{\frac{r}{2}-1} \sum_{k=1}^{n-1\,(i)} E| \sum_{m=k+1}^{n\,(i)} V_i W_{ij}|^r$$

$$\le C(n-1)^{\frac{r}{2}-1} \sum_{i=1}^{n} \sum_{k=1}^{n-1\,(i)} E(n-k)^{\frac{r}{2}-1} \sum_{m=k+1}^{n\,(i)} E|V_i W_{km}|^r$$

$$\le C n^{r+1} h^{-(r-1)d-r}$$

for $2 \le r \le 6$ by (iii), Lemma 1-(d) and (A.5). $\square$

LEMMA 29: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|R_1|^r = O(n^{-r}) \ \textit{for} \ 2 \le r \le 6 \ .$$

PROOF. Writing $E|R_1|^r \le C\binom{n}{2}^{-r} E|\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} V_i V_j|^r$ due to (3.8), as in Lemma 26 or 27, $V_i V_j$, $i=1,\ldots,j-1$ is a martingale difference sequence for fixed j as well as $\sum_{j=i+1}^{n} V_i V_j$, $i= n-1,\ldots,1$. We use DFJ repeatedly again and (i)', (iii) and Lemma 1-(d) to obtain

$$E\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} V_i V_j\right|^r \le C(n-1)^{\frac{r}{2}-1}\sum_{i=1}^{n-1} E\left|\sum_{j=i+1}^{n} V_i V_j\right|^r$$
$$\le C(n-1)^{\frac{r}{2}-1}\sum_{i=1}^{n-1}(n-i)^{\frac{r}{2}-1}\sum_{j=i+1}^{n} E|V_i V_j|^r = O(n^r) \quad . \qquad \square$$

LEMMA 30: *Under* (i)', (ii), (iii), (iv), (v), (vi) *and* (viii),

$$E|R_2|^r = O(n^{-3(r-1)}h^{-2(r-1)d-2r}) \quad for \ 2 \le r \le 3 \ .$$

PROOF. Using (3.8), write

$$E|R_2|^r = \frac{C}{n^r}\binom{n-1}{2}^{-r} E\left|\sum_{i=1}^{n}\sum_{k=1}^{n-1 (i)}\sum_{m=k+1}^{n (i)} (W_{ik}W_{im} - \tilde{W}_{km})\right|^r \ .$$

Since R_2 has the same martingale structure as Q_2 , the same method of proof as in Lemma 28 applies. The difference is in the moment bounds of the two summands, i.e.

$$E|V_i W_{km}|^r = O(h^{-(r-1)d-r}) \quad \text{for } 1 \le r \le 6, \ i \ne k \ne m,$$

and

$$E|W_{ik}W_{im} - \tilde{W}_{km}|^r = O(h^{-2(r-1)d-2r}) \quad , \quad i \ne k \ne m$$

by (i)', Lemmas 1-(d), 4 and Lemma 17-(b) for $1 \le r \le 3$. $\square$

LEMMA 31: *Under* (i)', (ii), (iii), (iv), (v), (vi) *and* (viii)

$$E|R_3|^r = O(n^{-2r}h^{-(2r-1)d-2r}) \quad for \ 2 \le r \le 3 \ .$$

PROOF. Write $E|R_3|^r \le \frac{C}{n^r}\binom{n}{2}^{-r} E\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \{W_{ij}^2 - \tilde{W}_{ii} - \tilde{W}_{jj} + E(W_{12}^2)\}\right|^r$ using (3.8). Since

$$E\{W_{ij}^2 - \tilde{W}_{ii} - \tilde{W}_{jj} + E(W_{12}^2) \mid j\} = E\{W_{ij}^2 - \tilde{W}_{ii} - \tilde{W}_{jj} + E(W_{12}^2) \mid i\} = 0$$

for $j > i$, R_3 has the same martingale structure as T_3 . Therefore, we apply DFJ to obtain

$$E\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\{W_{ij}^2-\tilde{W}_{ii}-\tilde{W}_{jj}+E(W_{12}^2)\}\right|^r$$

$$\le C(n-1)^{\frac{r}{2}-1}\sum_{i=1}^{n-1}E\left|\sum_{j=i+1}^{n}\{W_{ij}^2-\tilde{W}_{ii}-\tilde{W}_{ii}\,E(W_{12}^2)\}\right|^r$$

$$\le C(n-1)^{\frac{r}{2}-1}\sum_{i=1}^{n-1}(n-i)^{\frac{r}{2}-1}\sum_{j=i+1}^{n}E|\{W_{ij}^2-\tilde{W}_{ii}-\tilde{W}_{jj}+E(W_{12}^2)\}|^r$$

$$=O(n^r h^{-(2r-1)d-2r})$$

by (A.5) and Lemma 17-(a). □

LEMMA 32: *Under* (i)', (ii), (iv), (v), (vi) *and* (viii)

$$E|R_4|^r = O(n^{-\frac{3}{2}r}h^{-r(d+2)}) \quad \textit{for} \quad 1\le r\le 3 \ .$$

PROOF. Write $E|R_4|^r \le \frac{C}{n^{2r}}E|\sum_{i=1}^{n}\{\tilde{W}_{ii}-E(W_{12}^2)\}|^r$ using (3.8). Since $\tilde{W}_{ii}-E(W_{12}^2)$ is a martingale difference by (iii), DFJ and Lemma 17-(a) give

$$E\left|\sum_{i=1}^{n}\{\tilde{W}_{ii}-E(W_{12}^2)\}\right|^r \le Cn^{\frac{r}{2}-1}\sum_{i=1}^{n}E|\tilde{W}_{ii}-E(W_{12}^2)|^r = O(n^{\frac{r}{2}}h^{-r(d+2)}) \ .$$

□

LEMMA 33: *Under* (i)', (ii), (iii), (iv), (v), (vi), (vii) *and* (viii),

$$E|R_5|^r = O(n^{-2r}h^{-r(d+2)}) \quad \textit{for} \quad 1\le r\le 3 \ .$$

PROOF. Using (3.8), DFJ and (A.6),

$$E|R_5|^r \le \frac{C}{n^{4r}}E\left|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}W_{ij}\right|^{2r} \le \frac{C}{n^{4r}}(n-1)^{r-1}\sum_{i=1}^{n-1}E\left|\sum_{j=i+1}^{n}W_{ij}\right|^{2r}$$

$$= O(n^{-2r}h^{-r(d+2)}) \ .$$

□

LEMMA 34:

(a) $E^*(V_1^*)=0, \quad E^*(2V_1^*)^2=1\,, \quad E^*(W_{12}^*)=0.$

(b) $E^*(W_{12}^*\,|\,1^*)=E^*(W_{12}^*\,|\,2^*)=0.$

Under (i), (iii), (iv), (v), (vii), (viii) *and* (ix),

(c) $\nu^{\tau}(E^*U^*-EU)=o((\log n)^{-3}) \ .$

Under (i)", (iii), (iv)', (v)', (vi)', (vii) *and* (viii),

(d) $\sigma^{*2} = \sigma^2 + o\,((\log n)^{-3})$, $\dfrac{1}{\sigma^{*r}} = \dfrac{1}{\sigma^r} + o\,((\log n)^{-3})$, $r = 1, 2, 3$.

(e) $E^*(V_1^{*4}) = O(1)$.

PROOF. The proof for (a) and (b) is straightforward from the definitions. To prove (c), write $(\log n)^3 \mathrm{v}^\tau (E^*U^* - EU) = (\log n)^3 \mathrm{v}^\tau (U - EU) - \dfrac{(\log n)^3}{n} \mathrm{v}^\tau U$, then the second term converges to zero a.s. because of Proposition 1. The first term is

$$
\begin{aligned}
&(\log n)^3 \mathrm{v}^\tau (U - EU) \\
&= \frac{2(\log n)^3}{n} \sum_{i=1}^n v_i + \frac{2(\log n)^3}{n} \sum_{i=1}^n (V_i - v_i) + (\log n)^3 \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n W_{ij}
\end{aligned}
$$

Since v_i and $V_i - v_i$ are iid sequences with means zero, DFJ, Lemmas 1-(d), 3 and (i) give

$$
\begin{aligned}
E\left| \frac{(\log n)^3}{n} \sum_{i=1}^n v_i \right|^3 &\le \frac{C(\log n)^9}{n^{3/2}} \\
E\left| \frac{(\log n)^3}{n} \sum_{i=1}^n (V_i - v_i) \right|^2 &\le \frac{C(\log n)^6 h^{2L}}{n} .
\end{aligned}
$$

Therefore, Borel-Cantelli lemma and (ix)' give

$$
\begin{aligned}
&\frac{2(\log n)^3}{n} \sum_{i=1}^n v_i \to 0 \text{ a.s.} \\
&\frac{2(\log n)^3}{n} \sum_{i=1}^n (V_i - v_i) \to 0 \text{ a.s.}
\end{aligned}
$$

Lemmas 1-(b), (c) and (A.5) give

$$
E\left|(\log n)^3 \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n W_{ij} \right|^2 \le \frac{C(\log n)^6}{n^2 h^{d+2}} ,
$$

so that Borel-Cantelli lemma gives

$$
(\log n)^3 \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n W_{ij} \to 0 \text{ a.s.},
$$

which completes the proof of (c). To prove (d), write

$$
\begin{aligned}
\sigma^{*2} &= \sigma^2 E^*[\,2\sigma^{-1} \mathrm{v}^\tau (U_i^* - E^*U^*)\,]^2 = \frac{4\sigma^2}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=1}^n \sigma^{-1} \mathrm{v}^\tau (U_{ij} - E^*U^*) \right)^2 \\
&= \frac{4\sigma^2}{n} \sum_{i=1}^n \left[\frac{1}{n} \sum_{j=1}^n (W_{ij} + V_j) + V_i + \sigma^{-1} \mathrm{v}^\tau (EU - E^*U^*) \right]^2
\end{aligned}
$$

$$= \frac{4\sigma^2}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}(W_{ij}+V_j)\right]^2 + \frac{4\sigma^2}{n}\sum_{i=1}^{n}V_i^2 + 4[\,\mathrm{v}^\tau(EU-E^*U^*)\,]^2$$

$$+ \frac{8\sigma^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(W_{ij}+V_j)V_i + \frac{8\sigma\mathrm{v}^\tau(EU-E^*U^*)}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(W_{ij}+V_j)$$

$$+ \frac{8\sigma\mathrm{v}^\tau(EU-E^*U^*)}{n}\sum_{i=1}^{n}V_i\,. \tag{B.12}$$

The second term on the right of (B.12) is $\sigma^2 + o((\log n)^{-3})$ since

$$\frac{4}{n}\sum_{i=1}^{n}V_i^2 = \left[\frac{1}{n}\sum_{i=1}^{n}\{(2V_i)^2-(2v_i)^2\}\right] + \left[\frac{1}{n}\sum_{i=1}^{n}\{(2v_i)^2 - E(2v_1)^2\}\right] + E(2v_1)^2$$

$$= \quad o((\log n)^{-3}) \quad + \quad o((\log n)^{-3}) \quad + \quad 1\,,$$

due to

$$E\left|\frac{(\log n)^3}{n}\sum_{j=1}^{n}\{(2V_j)^2-(2v_j)^2\}\right|^3 \le \frac{C(\log n)^9}{n^{3/2}}E\{|V_1-v_1|(|V_1|+|v_1|)\}^3$$

$$\le \frac{C(\log n)^9 h^L}{n^{3/2}}(E|Y_j|+1)^6$$

by Lemma 1-(d) and (B.1) for the first term,

$$\left|\frac{(\log n)^3}{n}\sum_{i=1}^{n}\{(2v_i)^2-E(2v_1)^2\}\right|^3 \le \frac{C(\log n)^9}{n^{3/2}}$$

by (i)" and Lemma 1-(d) for the second term, and Lemma 1-(a) for the last term. The third term on the right of (B.12), apart from the square, is $o((\log n)^{-3})$ due to Lemma 34-(c). We shall show the other terms on the right of (B.12) are all of $o((\log n)^{-3})$ so that we consider each of the quantities multiplied by $(\log n)^3$, dropping the constants. As terms with $i=j$ are of smaller order and negligible, typical terms of the first term on the right of (B.12) multiplied by $(\log n)^3$ are constant times

$$\frac{(\log n)^3}{n^3}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(W_{ij}+V_j)^2 \tag{B.13}$$

and

$$\frac{(\log n)^3}{n^3}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}(W_{ij}+V_j)(W_{ik}+V_k)\ . \tag{B.14}$$

Using (A.5) and Lemma 1-(d), $E(W_{12}+V_2)^2 = O(h^{-d-2})$ so that (B.13) is expressed as

$$\frac{(\log n)^3}{n^3}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[(W_{ij}+V_j)^2 - E\{(W_{ij}+V_j)^2|j\}\right]$$

$$+\frac{(\log n)^3}{n^3}\sum_{j=2}^{n}(j-1)\left[E\{(W_{ij}+V_j)^2|j\}-E(W_{12}+V_2)^2\right]+O(\frac{(\log n)^3}{n^2h^{d+2}}) \ .$$

The expectations of the first two terms above are zero. Since

$$E\left([(W_{ij}+V_j)^2-E\{(W_{ij}+V_j)^2|j\}][(W_{kl}+V_l)^2-E\{(W_{kl}+V_l)^2|l\}]\right)=0$$

for $i\neq k$ and any j, l, and the summand of the second term is an independent sequence, their variances are bounded by

$$\frac{C(\log n)^6}{n^6}[n^2E(W_{12}^4+V_2^4)+n^3E(W_{12}^2+V_2^2)(W_{13}^2+V_3^2)]$$
$$=O(\frac{(\log n)^6}{n^4h^{3d+4}}+\frac{(\log n)^6}{n^3h^{2d+4}})$$

and

$$\frac{C(\log n)^6}{n^6}[n^3E\{E(W_{12}^2+V_1^2|1)\}^2]=O(\frac{(\log n)^6}{n^3h^{2d+4}})$$

respectively due to (i)", an elementary inequality, (A.5) and Lemmas 1-(d), 4. Thus, Borel-Cantelli lemma with assumption (ix)' (which implies $h^{-1}=O(n^{1/(d+2)})$) gives (B.13) is $o(1)$. (B.14) has mean zero and variance smaller than

$$\frac{C(\log n)^6}{n^6}\sum_{k=3}^{n}\sum_{j=2}^{k-1}E\{\sum_{i=1}^{j-1}(W_{ij}+V_j)(W_{ik}+V_k)\}^2$$
$$\leq\frac{C(\log n)^6}{n^4}[n\,E(W_{12}^2W_{13}^2)+n^2E(W_{13}W_{14}W_{23}W_{24})]$$
$$=O(\frac{(\log n)^6}{n^3h^{2d+4}}+\frac{(\log n)^6}{n^2h^{d+2}}) \ ,$$

similarly to the above argument so that (B.14) is $o(1)$ by Borel-Cantelli lemma. Typical term of the fourth term on the right of (B.12) multiplied by $(\log n)^3$ is

$$\frac{C(\log n)^3}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(W_{ij}+V_j)V_i \qquad \text{(B.15)}$$

which has mean zero by Lemma 1-(b), (c) and variance bounded by

$$\frac{C(\log n)^6}{n^4}[n^2E\{(W_{12}+V_2)V_1\}^2+n^3E\{(W_{13}+V_3)V_1(W_{23}+V_3)V_2\}] \ .$$

due to $E\{(W_{ij}+V_j)V_i(W_{kl}+V_l)V_k\}=0$ for $j\neq l$. An elementary inequality and Lemmas 1-(d) and 4 give that the first expectation above is bounded by

$$CE(W_{12}^2V_1^2+V_1^2V_2^2)=CE\{V_1^2E(W_{12}^2|1)+V_1^2V_2^2\}=O(h^{-d-2})$$

and Lemma 1-(b), (c) and (A.3) give the second expectation above is equal to

$$E(W_{13}W_{23}V_1V_2)=E\{E(W_{13}V_1|3)E(W_{23}V_2|3)\}\leq CE(|Y_3|^2+1)\leq C.$$

Thus the variance of (B.15) is $O(\frac{(\log n)^6}{n^2h^{d+2}}+\frac{(\log n)^6}{n})$ so that (B.15) is

$o(1)$. Since $\nu^\tau(EU-E^*U^*)$ is $o(1)$ by Lemma 34-(c) and $\frac{(\log n)^3}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(W_{ij}+V_j)$ has mean zero and variance smaller than

$$\frac{C(\log n)^6}{n^4}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\{E(W_{ij}^2)+E(V_j^2)\} = O(\frac{(\log n)^6}{n^2h^{d+2}}) ,$$

the fifth term of (B.12) is $o((\log n)^{-3})$, which completes the proof of the first equation of (d). The second one is straightforward from the first.

To prove (e), using Lemma 34-(d),

$$E^*(V_1^{*4}) = \frac{1}{n^5}\sum_{i=1}^{n}|\sum_{j=1}^{n}\sigma^{*-1}\nu^\tau(U_{ij}-E^*U^*)|^4$$
$$\leq \frac{C}{n^5}\sum_{i=1}^{n}\{|(n-1)\nu^\tau(E^*U^*-EU)+\nu^\tau E^*U^*|^4\}$$
$$+\frac{C}{n^5}\sum_{i=1}^{n}\left\{|\sum_{j\neq i}^{n}(W_{ij}+V_i+V_j)|^4\right\} ,$$

the first term is $o(n^{-3})$ by Proposition 1, $E^*U^*=(n-1)U/n$ and Lemma 34-(c). The second term is smaller than a constant times

$$\frac{1}{n^5}\sum_{i=1}^{n}|\sum_{j\neq i}^{n}W_{ij}|^4+\frac{1}{n}\sum_{i=1}^{n}V_i^4 , \tag{B.16}$$

whose second term is bounded since

$$\frac{1}{n}\sum_{i=1}^{n}|V_i|^r \leq \frac{C}{n}\sum_{i=1}^{n}(|Y_i|^r+1) < C \text{ for } r \leq 8+\delta \tag{B.17}$$

due to Lemma 1-(d) and SLLN under (i)". The first term of (B.16) is equal to

$$\frac{1}{n^5}\sum\sum_{i\neq j}W_{ij}^4+\frac{1}{n^5}\sum\sum\sum_{i\neq j\neq k}W_{ij}^3W_{ik}+\frac{1}{n^5}\sum\sum\sum_{i\neq j\neq k}W_{ij}^2W_{ik}^2$$
$$\frac{1}{n^5}\sum\sum\sum\sum_{i\neq j\neq k\neq l}W_{ij}^2W_{ik}W_{il}+\frac{1}{n^5}\sum\sum\sum\sum\sum_{i\neq j\neq k\neq l\neq m}W_{ij}W_{ik}W_{il}W_{im}$$

and similar method of the proof to Theorem 2 (see the proof of convergence of (2.34)-(2.36) to zero) gives the second, fourth and fifth terms converge to zero. For the remaining terms, we show more general results than we need here, namely,

$$\frac{1}{n^2}\sum\sum_{i\neq j}|W_{ij}|^r = E|W_{12}|^r+o(h^{-(r-1)d-r}) \text{ for } r=1,2,3,4 \tag{B.18}$$

and

$$\frac{1}{n^3}\sum\sum\sum_{i\neq j\neq k}|W_{ij}W_{ik}|^r = E|W_{12}W_{13}|^r+o(h^{-2(r-1)d-2r}) \text{ for } r=1,2 \tag{B.19}$$

(B.18) and (B.19) are used to prove Lemma 37 later. Write due to the symmetry of W_{ij} ,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|W_{ij}|^r = \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}|W_{ij}|^r ,$$

then the right side is asymptotically equivalent to

$$\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(Q_{ij}+Q_i+Q_j)+E|W_{12}|^r$$
$$=\frac{2}{n}\sum_{i=1}^{n}Q_i+\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}Q_{ij}+E|W_{12}|^r \tag{B.20}$$

where $Q_i=E(|W_{ij}|^r|i)-E|W_{12}|^r$ and $Q_{ij}=|W_{ij}|^r-Q_i-Q_j-E|W_{12}|^r$. Since Q_i is iid with $E(Q_i)=0$ and $|Q_i|\le C(|Y_i|^r+1)h^{-(r-1)d-r}$ by Lemma 4, setting $\xi=\frac{1}{4}\delta$, and using DFJ,

$$E\left|\frac{h^{(r-1)d+r}}{n}\sum_{i=1}^{n}Q_i\right|^{2+\xi}\le\frac{1}{n^{2+\xi}}n^{\frac{2+\xi}{2}-1}\sum_{i=1}^{n}|h^{(r-1)d+r}Q_i|^{2+\xi}$$
$$\le\frac{C}{n^{(4+\xi)/2}}\sum_{i=1}^{n}(|Y_i|^{(2+\xi)r}+1)=O(n^{-(2+\xi)/2}) ,$$

due to (i)" and SLLN so that the leading term on the right of (B.20) is $o(h^{-(r-1)d-r})$. Similarly, $E(Q_{12})=0$ and

$$E\left\{h^{(r-1)d+r}\binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}Q_{ij}\right\}^2=\frac{2h^{2(r-1)d+2r}}{n(n-1)}E(Q_{12}^2)$$
$$\le\frac{Ch^{2(r-1)d+2r}}{n^2}E|W_{12}|^{2r}=O(n^{-2}h^{-d})$$

by (A.5) so that the second term on the right of (B.20) is $o(h^{-(r-1)d-r})$. Thus (B.18) is true. To prove (B.19), we only show a bound of its typical term $\frac{1}{n^3}\sum\sum_{i<j<k}\sum|W_{ij}W_{ik}|^r$. Putting

$$R_{ijk}=|W_{ij}W_{ik}|^r,\ R_{ij}=E(R_{ijk}|i,j)\ ,\ P_{ijk}=R_{ijk}-R_{ij}-R_{ik}+E(R_{123})\ ,$$

$$\binom{n}{3}^{-1}\sum\sum_{i<j<k}\sum|W_{ij}W_{ik}|^r$$
$$=\binom{n}{3}^{-1}\sum\sum_{i<j<k}\sum P_{ijk}+\binom{n}{3}^{-1}\sum_{i=1}^{n}(n-i-1)\sum_{j>i}\{R_{ij}-E(R_{123})\}+E(R_{123}) . \tag{B.21}$$

The first term times $h^{2(r-1)d+2r}$ has mean zero and variance bounded by

$$\frac{Ch^{4(r-1)d+4r}}{n^6}\left[\sum_{i=1}^{n-2}E(\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}P_{ijk})^2\right.$$
$$\left.+\sum_{i=1}^{n}\sum_{l\ne i}^{n}E(\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}P_{ijk})(\sum_{j=l+1}^{n-1}\sum_{k=j+1}^{n}P_{ljk})\right]$$

$$\leq \frac{Ch^{4(r-1)d+4r}}{n^6}\{nE(R_{123}^2)+n^2E(R_{134}R_{234})\}$$
$$\leq \frac{Ch^{4(r-1)d+4r}}{n^6}\{nE|W_{12}W_{13}|^{2r}+n^2E|W_{13}W_{14}|^r|W_{23}W_{24}|^r\}$$
$$= O\left(\frac{h^{4(r-1)d+4r}}{n^5h^{2(2r-1)+4r}}+\frac{h^{4(r-1)d+4r}}{n^4h^{2(2r-1)+4r}}\right)=O(n^{-4}h^{-2d})$$

due to Hölder's inequality and (A.5), so that it is $o(h^{-2(r-1)d-2r})$ by (ix). Rewriting the second term of (B.21) as

$$\binom{n}{3}^{-1}\sum_{i=1}^{n}(n-i-1)\sum_{j>i}\{R_{ij}-E(R_{ij}|i+1,\cdots,n)\}$$
$$+\binom{n}{3}^{-1}\sum_{i=1}^{n}(n-i-1)\sum_{j>i}\{E(R_{ij}|i+1,\cdots,n)-E(R_{123})\}\ ,$$

the first term times $h^{2(r-1)d+2r}$ has mean zero and variance bounded by

$$\frac{Ch^{4(r-1)d+4r}}{n^2}E(R_{12}^2)\ \leq \frac{Ch^{2(r-1)d+2r}}{n^2}E\{|W_{12}|^{2r}(|Y_1|^{2r}+1)\}$$
$$=\frac{Ch^{2(r-1)d+2r}}{n^2}E(|Y_1|^{4r}+1)h^{-(2r-1)d-2r}=O(n^{-2}h^{-d})$$

using Lemma 4 repeatedly so that the first term is $o(h^{-2(r-1)d-2r})$ by (ix)'. The second term is handled similarly to the first term of (B.20) using

$$E\{h^{2(r-1)d+2r}E(R_{ij}|j)\}^{2+\xi}\leq CE[h^{(r-1)d+r}E\{|W_{ij}|^r(|Y_i|^r+1)|j\}]^{2+\xi}$$
$$\leq CE\{|Y_j|^{(2+\xi)r}+1\}=O(1)\ ,$$

under (i)" and so it is $o(h^{-2(r-1)d-2r})$. The first and second inequalities above uses Lemma 4 and $E(|W_{12}Y_1|^r|2)=O(h^{-(r-1)d-r})$ shown similarly to Lemma 4.

□

LEMMA 35:

$\bar{V}^*-\bar{V}^*(m)$ *and* $\bar{W}^*-\bar{W}^*(m)$ *are independent of* $(X_1^{*\tau},Y_1^*),\cdots,$ $(X_m^{*\tau},Y_m^*)$ *conditional on* $(X_1^\tau,Y_1),\cdots,(X_n^\tau,Y_n)$.

PROOF.

The proof is straightforward. □

LEMMA 36: *Under* (i)", (iii), (iv)', (v)', (vi)', (vii), *and* (viii),

$$|E^*(W_{12}^*V_1^*V_2^*)|=|E(W_{12}V_1V_2)|+o(1)=O(1)\ .$$

PROOF. The second equality is immediate by (A.4). Writing

$$E^*(W_{12}^*V_1^*V_2^*) = \sigma^{*-1}E^*(\nu^\tau U_{12}^*V_1^*V_2^*) - 2E^*(V_1^{*2}V_2^*) - \sigma^{*-1}E^*\nu^\tau U^* E^*(V_1^*V_2^*) \ ,$$

the last two terms are zero by Lemma 34-(a), (e) and the first term is

$$\frac{\sigma^{-3}}{n^4}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n \{\nu^\tau(U_{ij}-EU)\}\{\nu^\tau(U_{ik}-EU)\}\{\nu^\tau(U_{jl}-EU)\}$$
$$= \frac{1}{n^4}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n (W_{ij}+V_i+V_j)(W_{ik}+V_i+V_k)(W_{jl}+V_j+V_l)$$

plus smaller terms involving $\sigma^{*\,-3} - \sigma^{-3}$ and $\nu^\tau(E^*U^* - EU)$, which are $o(1)$. Expanding the summand of the right hand side, the dominating terms are

$$\frac{1}{n^4}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n W_{ij}\,W_{ik}\,W_{jl} \qquad \text{(B.22)}$$

and

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n W_{ij}\,V_iV_j = E(W_{12}V_1V_2) + \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \{W_{ij}\,V_iV_j - E(W_{12}V_1V_2)\} \ . \text{(B.23)}$$

The absolute value of the second term on the right of (B.23) was shown to converge to zero in the proof of Theorem 2 when the normalizer is $n(n-1)$ instead of n^2, which however does not change the asymptotic result. Rewriting (B.22) as

$$\frac{1}{n^4}\sum_{i=1}^n\sum_{k=1}^n\sum_{l=1}^n W_{ii}\,W_{ik}\,W_{il} + \frac{1}{n^4}\sum_{i=1}^n\sum_{j\neq i}^n\sum_{k=1}^n\sum_{l=1}^n W_{ij}\,W_{ik}\,W_{il} \ ,$$

we can easily show that the first term is $o(1)$ by Borel-Cantelli Lemma since it has mean zero and variance bounded by $Cn^{-4}E(W_{11}^2W_{12}^2W_{13}^2) = O(n^{-4}h^{-2d-4})$ by Lemmas 1-(d), 4 and we can handle the second term similarly to (2.34)-(2.36) to show it converges to zero. □

LEMMA 37: *Under* (i)", (iii), (iv)', (v)', (vi)', (vii), *and* (viii),

(a) $E^*|W_{12}^*|^r = E|W_{12}|^r + o(h^{-(r-1)d-r})$ *for* $r = 1, 2, 3, 4$.

(b) $E^*(|W_{12}^*W_{13}^*|^r) \le O(h^{-2(r-1)d-2r})$ *for* $r = 1, 2$.

PROOF. To prove (a), write using an elementary inequality and Lemma 34-(d),

$$E^*|W_{12}^*|^r - E^*|\sigma^{*-1}\nu^\tau(U_{12}^* - EU)|^r \le C[\,|\nu^\tau(E^*U^* - EU)|^r + E^*|V_1^*|^r + E^*|V_2^*|^r$$
$$+ E^*|\nu^\tau(U_{12}^* - EU)|^{r-1}\{|\nu^\tau(E^*U^* - EU)| + E^*|V_1^*| + E^*|V_2^*|\}] \ .$$

$|\nu^\tau(E^*U^* - EU)|^r + E^*|V_1^*|^r + E^*|V_2^*|^r = O(1)$ by Lemma 34-(c), (e).

We have $\sigma^{*-1} = \sigma^{-1} + o(1)$ and, using an elementary inequality,

$$E^*|\sigma^{-1}\nu^\tau(U_{12}^*-EU)|^r = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |\sigma^{-1}\nu^\tau(U_{ij}-EU)|^r$$
$$= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |W_{ij}+V_i+V_j|^r$$
$$= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |W_{ij}|^r + O\Big(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |W_{ij}|^{r-1}|V_i+V_j| + \frac{1}{n}\sum_{i=1}^n |V_i|^r\Big) ,$$

where we can show

$$\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n |W_{ij}|^{r-1}|V_i+V_j| = E|W_{12}|^{r-1}|V_1+V_2| + o(h^{-(r-2)d-(r-1)})$$

similarly to (B.18) and the expectation is $O(h^{-(r-2)d-(r-1)})$ by Lemmas 1-(d) and 4 so that, using (B.17), the last term in the above is $o(h^{-(r-1)d-r})$. The first term is $E|W_{12}|^r + o(h^{-(r-1)d-r})$ by (B.18).

(b) Similar manipulation to the above gives

$$E^*|W_{12}^*W_{13}^*|^r \le \frac{C}{n^3}\sum\sum\sum_{i\ne j\ne k} |W_{ij}\,W_{ik}|^r \ + \text{smaller order terms.}$$

Apply (B.19). □

LEMMA 38: *Under* (i)", (iv)' *and* (viii),

$$E^*[W_{12}^*w_1^*(t)w_2^*(t)]$$
$$= \frac{4(it)^2}{n}E(W_{12}V_1V_2) + o\Big(\frac{t^2}{n}\Big) + O\Big(\Big(\frac{|t|^4}{n^2}+\frac{|t|^3}{n^{3/2}}\Big)h^{-\frac{2}{3}d-1}\Big) .$$

PROOF.

Write $\zeta_i^* = \frac{2}{n^{1/2}}V_i^*$.

$E^*[W_{12}^*w_1^*(t)w_2^*(t)] = E^*[W_{12}^*\{w_1^*(t)-1\}\{w_2^*(t)-1\}]$ by Lemma 34-(b)

$$= E^*\Big[W_{12}^*\Big\{(e^{it\zeta_1^*}-1)(e^{it\zeta_2^*}-1)-(it\zeta_1^*)(it\zeta_2^*)\Big\}\Big] + E^*[W_{12}^*(it\zeta_1^*)(it\zeta_2^*)]$$
$$= E^*\Big[W_{12}^*\Big\{(e^{it\zeta_1^*}-1-it\zeta_1^*)(e^{it\zeta_2^*}-1-it\zeta_2^*) + it\zeta_1^*(e^{it\zeta_2^*}-1-it\zeta_2^*)$$
$$+ it\zeta_2^*(e^{it\zeta_1^*}-1-it\zeta_1^*)\Big\}\Big] + E^*[W_{12}^*(it\zeta_1^*)(it\zeta_2^*)]$$
$$= E^*[W_{12}^*(it\zeta_1^*)(it\zeta_2^*)] + O\big(E^*|W_{12}^*||t\zeta_1^*|^2|t\zeta_2^*|^2 + E^*|W_{12}^*||t\zeta_1^*|^2|t\zeta_2^*|\big)$$
$$= \frac{4(it)^2}{n}E^*(W_{12}^*V_1^*V_2^*) + O\Big(\frac{t^4}{n^2}E^*|W_{12}^*||V_1^*|^2|V_2^*|^2 + \frac{|t|^3}{n^{3/2}}E^*|W_{12}^*||V_1^*|^2|V_2^*|\Big)$$
$$= \frac{4(it)^2}{n}E(W_{12}V_1V_2) + o\Big(\frac{t^2}{n}\Big)$$
$$+ O\Big(\frac{t^4}{n^2}E^*|W_{12}^*||V_1^*|^2|V_2^*|^2 + \frac{|t|^3}{n^{3/2}}E^*|W_{12}^*||V_1^*|^2|V_2^*|\Big)$$

by Lemma 3. The fourth equality uses (2.6). By Hölder's inequality,

$$E^*|W_{12}^*||V_1^*|^2|V_2^*|^2 \le \{E^*|W_{12}^*|^3\}^{1/3}\{E^*|V_1^*|^3E^*|V_2^*|^3\}^{2/3} \le Ch^{-\frac{2}{3}d-1} .$$

The last inequality is because of Lemma 34-(e) with (iii) and (i)" and $E^*|W_{12}^*|^3 \le Ch^{-2d-3}$ due to Lemma 37-(b). Similarly,

$$E^*|W_{12}^*||V_1^*|^2|V_2^*| \le \left\{E^*|W_{12}^*|^3\right\}^{1/3}\left\{E^*|V_1^*|^3E^*|V_2^*|^{3/2}\right\}^{2/3} \le Ch^{-\frac{2}{3}d-1} .\Box$$

LEMMA 39: *Under* (i)", (iii), (iv)', (v)', (vi)', (vii), *and* (viii),

$$E^*\left[W_{12}^{*\,2}w_1^*(t)\,w_2^*(t)\right] = h^{-d-2}\kappa_2 + o(h^{-d-2}) + O\left(\frac{|t|}{n^{1/2}}h^{-\frac{4}{3}d-2}\right) .$$

PROOF.

$$\begin{aligned} E\left[W_{12}^{*\,2}w_1^*(t)\,w_2^*(t)\right] &= E^*(W_{12}^{*\,2}) + E^*\left[W_{12}^{*\,2}\{e^{it(\zeta_1^*+\zeta_2^*)}-1\}\right] \\ &= E^*(W_{12}^{*\,2}) + O\left(E^*|W_{12}^*|^2|t\,\zeta_1^*|\right) \text{ by (2.6)} \\ &= E^*(W_{12}^{*\,2}) + O\left(\frac{|t|}{n^{1/2}}E^*|W_{12}^*|^2|V_1^*|\right) . \end{aligned}$$

By Hölder's inequality, Lemma 1-(d) and 4-(a),

$$E^*|W_{12}^*|^2|V_1^*| \le (E^*|W_{12}^*|^3)^{2/3}(E^*|V_1^*|^3)^{1/3} \le Ch^{-\frac{4}{3}d-2} .$$

Apply Lemma 37-(a) and Lemma 12. □

LEMMA 40: *Under* (i)", (iii), (iv)', (v)', (iv)', (vii), *and* (viii),

$$E^*\left[W_{12}^*W_{13}^*w_1^*(t)\,w_2^*(t)\,w_3^*(t)\right] = O\left(\frac{|t|^3}{n^{3/2}h^{\frac{4}{3}d+2}}\right) .$$

PROOF. Using Lemma 34-(b) and (2.6),

$$\begin{aligned} E^*\left[W_{12}^*W_{13}^*w_1^*(t)\,w_2^*(t)\,w_3^*(t)\right] &= E^*\left[W_{12}^*W_{13}^*(e^{it\zeta_1^*}-1)(e^{it\zeta_2^*}-1)(e^{it\zeta_3^*}-1)\right] \\ &= O\left(E^*(|W_{12}^*||W_{13}^*||t\,\zeta_1^*||t\,\zeta_2^*||t\,\zeta_3^*|)\right) \\ &= O\left(\frac{|t|^3}{n^{3/2}}E^*(|W_{12}^*||W_{13}^*||V_1^*||V_2^*||V_3^*|)\right) . \end{aligned}$$

By Hölder's inequality, Lemmas 34-(e), 37-(a) and (A.5),

$$\begin{aligned} &E^*(|W_{12}^*||W_{13}^*||V_1^*||V_2^*||V_3^*|) \\ &\le (E^*|W_{12}^*|^3)^{1/3}(E^*|W_{13}^*|^3)^{1/3}(E^*|V_1^*|^3E^*|V_2^*|^3E^*|V_3^*|^3)^{1/3} \\ &\le Ch^{-\frac{4}{3}d-2} . \end{aligned}$$

□

LEMMA 41: *Under* (i)", (iii), (iv)', (v)', (vi)', (vii), *and* (viii),

$$E^*\left[W_{12}^*W_{34}^*w_1^*(t)\,w_2^*(t)\,w_3^*(t)\,w_4^*(t)\right]$$

$$= O\left(\frac{t^4}{n^2} + \frac{t^6}{n^3}h^{-\frac{4}{3}d-2} + \frac{t^8}{n^4}h^{-\frac{4}{3}d-2}\right) .$$

PROOF. By (iii) and Lemma 38,

$$E^*\left[W_{12}^*W_{34}^*w_1^*(t)\,w_2^*(t)\,w_3^*(t)\,w_4^*(t)\right] = \left\{E^*[\,W_{12}^*w_1^*(t)\,w_2^*(t)\,]\right\}^2$$
$$= \left[\frac{4(it)^2}{n}E(\,W_{12}V_1V_2) + o\,(\frac{t^2}{n}) + O\left((\frac{t^4}{n^2} + \frac{|t|^3}{n^{3/2}})\,h^{-\frac{2}{3}d-1}\right)\right]^2 .$$

Apply (A.4). □

LEMMA 42: *Under* (i)", (iii), (iv)', (v)', (vi)', (vii), (viii) *and* (ix)',

$$E^*|\bar{W}^*(m)|^3 = O\left((\frac{m}{n^2h^{d+2}})^{\frac{3}{2}}\right) ,$$

in particular $E^*|W^*|^3 = E^*|\bar{W}^*(n-1)|^3 = O\left((n^{-1}h^{-d-2})^{\frac{3}{2}}\right)$.

PROOF. Writing $S_{i,n}^* = \sum_{j=i+1}^{n} W_{ij}^*$, $S_{i,n}^*$ is a reverse martingale difference sequence, i.e.

$$E^*(S_{i,n}^*|i+1^*, \cdots, n^*) = 0 \tag{B.24}$$

due to Lemma 34-(b). Put $\bar{\bar{W}}^*(m) = n^{1/2}\binom{n}{2}^{-1}\sum_{k=1}^{m} S_{m-k+1,n}^*$, then $\bar{W}^*(m) = \bar{\bar{W}}^*(m)$. Taylor's expansion and $\bar{\bar{W}}^*(m) - \bar{\bar{W}}^*(m-1) = n^{1/2}\binom{n}{2}^{-1}S_{1,n}^*$ give

$$|\bar{\bar{W}}^*(m)|^3 - |\bar{\bar{W}}^*(m-1)|^3 = 3sgn\,(\bar{\bar{W}}^*(m-1))\,\bar{\bar{W}}^*(m-1)^2n^{1/2}\binom{n}{2}^{-1}S_{1,n}^*$$
$$+ 3\,|\bar{\bar{W}}^*(m-1) + \theta n^{1/2}\binom{n}{2}^{-1}S_{1,n}^*|\{n^{1/2}\binom{n}{2}^{-1}S_{1,n}^*\}^2 , \tag{B.25}$$

for some $\theta\in[0, 1]$. Since $\bar{\bar{W}}^*(m-1)$ does not involve $(X_1^{*\tau}, Y_1^*)$, taking conditional expectation of (B.25) given $(X_1^\tau, Y_1), \cdots, (X_n^\tau, Y_n)$ using (B.24), we have

$$E^*|\bar{\bar{W}}^*(m)|^3 - E^*|\bar{\bar{W}}^*(m-1)|^3$$
$$= 3E^*\left[\,|\bar{\bar{W}}^*(m-1) + \theta n^{1/2}\binom{n}{2}^{-1}S_{1,n}^*|\{n^{1/2}\binom{n}{2}^{-1}S_{1,n}^*\}^2\right]$$
$$\le \frac{C}{n^3}\left\{E^*|\bar{\bar{W}}^*(m-1)|S_{1,n}^{*2} + \frac{1}{n^{3/2}}E^*|S_{1,n}^*|^3\right\}$$
$$\le \frac{C}{n^3}\left[\left\{E^*\{\bar{\bar{W}}^*(m-1)^2\}\right\}^{1/2}\left\{E^*(S_{1,n}^{*4})\right\}^{1/2} + \frac{1}{n^{3/2}}E^*|S_{1,n}^*|^3\right] . \tag{B.26}$$

The last inequality uses Hölder's inequality. For $i=1, \ldots , n$, Lemmas 34-(b), 37, (A.5) and (ix)' yield

$$E^*(S_{i,n}^{*4}) = \sum_{j=i+1}^{n} E^*(W_{ij}^{*4}) + 6\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} E^*(W_{ij}^{*2}W_{ik}^{*2})$$
$$\le C\{nE^*(W_{12}^{*4}) + n^2E^*(W_{12}^{*2}W_{13}^{*2})\} = O(n^2h^{-2(d+2)}) \quad \text{(B.27)}$$

so that Jensen's inequality and (B.27) give

$$E^*|S_{i,n}^*|^3 = O((nh^{-d-2})^{3/2}) \ , \quad \text{(B.28)}$$

$$E^*(S_{i,n}^{*2}) = O(nh^{-d-2}) \ . \quad \text{(B.29)}$$

The first wave bracketed term in (B.26) is equal to

$$E^*\left\{n^{1/2}\binom{n}{2}^{-1}\sum_{k=1}^{m-1} S_{m-k+1,n}^*\right\}^2 = n\binom{n}{2}^{-2}\sum_{i=2}^{m} E^*(S_{i,n}^{*2}) = O(\frac{m}{n^2h^{d+2}}) \ . \quad \text{(B.30)}$$

The first equality is due to $E^*(S_{i,n}^*S_{j,n}^*) = 0$ for $i \neq j$ by (B.24) and the second equality uses (B.29). Substituting (B.27), (B.28), (B.30) into (B.26),

$$E^*|\bar{\bar{W}}^*(m)|^3 - E^*|\bar{\bar{W}}^*(m-1)|^3$$

$$\le \frac{C}{n^3}\left\{(\frac{m}{n^2h^{d+2}})^{1/2}(\frac{n^2}{h^{2(d+2)}})^{1/2} + \frac{1}{n^{3/2}}(\frac{n}{h^{d+2}})^{3/2}\right\}$$

$$= O\left(\frac{m^{1/2}}{(n^2h^{d+2})^{3/2}}\right) \ .$$

Solving the difference equation, we get $\bar{\bar{W}}^*(m) = O\left((\frac{m}{n^2h^{d+2}})^{3/2}\right)$. The second equation follows immediately from the first. □

Appendix C

Here, we show how some of the terms appearing in the proof of Theorem 3 can be expressed. For $1 \le m \le n-1$,

(a) $E(\bar{b}_2' e^{itb_2})$

$$= -\{\gamma(t)\}^{n-1}\left\{it\frac{2}{n}E(W_{12}^2) + O(\frac{|t|}{n^2h^{d+2}} + \frac{t^2}{n^{3/2}h^{d+2}} + \frac{|t|h^L}{nh^{d+2}})\right\}$$
$$- \{\gamma(t)\}^{n-1}\left\{\frac{4E(V_1^3) + 8E(W_{12}V_1V_2)}{n^{1/2}} + O(\frac{|t|}{n})\right\}$$

$$- \{\gamma(t)\}^{n-2}\left[\frac{(it)^2}{n^{1/2}}\{4E(V_1^3) + 8E(W_{12}V_1V_2)\} + O(\frac{t^2 + |t|^3}{n} + \frac{t^4}{n^{3/2}})\right]$$

$$+ \{\gamma(t)\}^{n-3}\left[O\left(\frac{|t|^3 + |t|}{n} + \frac{t^2}{n^{3/2}h^{d+2}} + \frac{t^6}{n^3h^{d+2}} + \frac{|t|^5}{n^{5/2}h^{d+2}} + \frac{t^4 + |t|^3}{n^2h^{d+2}}\right)\right] ,$$

(b) $|E(\delta_{2m}e^{it\bar{B}_m})| \le \frac{Cm}{n^{1/2}h^2}|\gamma(t)|^{m-4}$,

(c) $|E(\delta'_{3m}e^{it\bar{B}_m})| \le \frac{Cm}{n^{1/2}h^3}|\gamma(t)|^{m-4}$,

(d) $E|\delta_{2m}|^2 \le Cm(\frac{1}{n^3h^{2d+4}} + \frac{1}{n^2})$,

(e) $E|\delta'_{3m}|^2 \le \frac{Cm}{n^4h^{3d+6}}$,

(f) $|E\delta''_3 e^{it\bar{B}_m}| \le \frac{Cn^{1/2}}{h^2}|\gamma(t)|^{m-5}$.

PROOF.

(a) Write

$$E(\delta'_2 e^{itb_2}) = E(T\bar{V}e^{itb_2})$$
$$= E(T_1\frac{2}{\sqrt{n}}\sum_{j=1}^{n}V_je^{itb_2}) + E(T_2\frac{2}{\sqrt{n}}\sum_{j=1}^{n}V_je^{itb_2}) + E(T_3\frac{2}{\sqrt{n}}\sum_{i=1}^{n}V_je^{itb_2})$$
$$= (A) + (B) + (C) \quad . \tag{C.1}$$

Thus

$$(A) = -\frac{4n^{1/2}}{(n-2)^2s^3}E(W_{12}^2)\sum_{j=1}^{n}E(V_je^{itb_2}) \quad . \tag{C.2}$$

Due to (iii), (2.6), $\gamma(t) = E(e^{it\frac{2}{\sqrt{ns}}V_1})$, $4E(V_1^2) = s^2$, (3.8) and Lemma 1-(d),

$$E(V_je^{itb_2}) = E(V_je^{it\frac{2}{\sqrt{ns}}V_j})\,E(e^{it\frac{2}{\sqrt{ns}}\sum_{k\neq j}^{n}V_k})$$
$$= \left[E\left\{V_j(e^{it\frac{2}{\sqrt{ns}}V_j} - 1 - it\frac{2V_j}{n^{1/2}s})\right\} + (it)E(\frac{2V_j^2}{n^{1/2}s})\right]\{\gamma(t)\}^{n-1}$$
$$= \{\gamma(t)\}^{n-1}\left\{\frac{its}{2n^{1/2}} + O(\frac{t^2}{n})\right\}$$
$$= \{\gamma(t)\}^{n-1}\left\{\frac{it}{2n^{1/2}} + O(\frac{t^2}{n} + \frac{|t|h^L}{n^{1/2}})\right\} \tag{C.3}$$

Substituting (C.3) into (C.2),

$$(A) = -\{\gamma(t)\}^{n-1}\frac{4n^{3/2}}{(n-2)^2 s^3}E(W_{12}^2)\left\{\frac{it}{2n^{1/2}} + O(\frac{t^2}{n} + \frac{|t|h^L}{n^{1/2}})\right\}$$

$$= -\frac{\{\gamma(t)\}^{n-1}}{s^3}\left\{\frac{2it}{n}E(W_{12}^2) + O(\frac{|t|}{n^2h^{d+2}} + \frac{t^2}{n^{3/2}h^{d+2}} + \frac{|t|h^L}{nh^{d+2}})\right\} . \tag{C.4}$$

Now, we write $(B) = (B)' + (B)''$ where

$$(B)' = -\frac{1}{n^{3/2}s^3}\sum_{j=1}^{n}E(4V_j^2 - s^2 + 8\tilde{V}_j)V_j e^{itb_2} . \tag{C.5}$$

$$(B)'' = -\frac{1}{n^{3/2}s^3}\sum_{j=1}^{n}\sum_{k\neq j}^{n}E(4V_j - s^2 + 8\tilde{V}_j)V_k e^{it\frac{2}{\sqrt{n}s}\sum_{l=1}^{n}V_l} \tag{C.6}$$

The summand of $(B)'$ is, using (2.6) and Lemma 1-(d),

$$E\{(4V_1^2 - s^2 + 8\tilde{V}_1)V_1 e^{it\frac{2V_1}{\sqrt{n}s}}\}E(e^{it\frac{2}{\sqrt{n}s}\sum_{l\neq 1}V_l})$$

$$= \{\gamma(t)\}^{n-1}E\{(4V_1^2 - s^2 + 8\tilde{V}_1)V_1 e^{it\frac{2V_1}{\sqrt{n}s}}\}$$

$$= \{\gamma(t)\}^{n-1}\left\{4E(V_1^3) + 8E(W_{12}V_1V_2) + O(\frac{|t|}{n^{1/2}})\right\} . \tag{C.7}$$

Substituting (C.7) into (C.5),

$$(B)' = -\frac{1}{n^{1/2}s^3}\{\gamma(t)\}^{n-1}\left\{4E(V_1^3) + 8E(W_{12}V_1V_2) + O(\frac{|t|}{n^{1/2}})\right\} . \tag{C.8}$$

For $j \neq k$, the summand of $(B)''$ is, due to (iii),

$$E\{(4V_1^2 - s^2 + 8\tilde{V}_1)V_2 e^{it\frac{2(V_1+V_2)}{\sqrt{n}s}}\}E(e^{it\frac{2}{\sqrt{n}s}\sum_{l\neq 1,2}V_l})$$

$$= \{\gamma(t)\}^{n-2}E\{(4V_1^2 - s^2) + 8\tilde{V}_1)V_2 e^{it\frac{2(V_1+V_2)}{\sqrt{n}s}}\}$$

$$= \{\gamma(t)\}^{n-2}E\{(4V_1^2 - s^2) + 8\tilde{V}_1)e^{it\frac{2V_1}{\sqrt{n}s}}\}E(V_2 e^{it\frac{2V_2}{\sqrt{n}s}})$$

$$= \{\gamma(t)\}^{n-2}\left[E\left\{(4V_1^2 - s^2 + 8\tilde{V}_1)(e^{it\frac{2V_1}{\sqrt{n}s}} - 1 - it\frac{V_1}{\sqrt{n}s})\right\} + itE\frac{V_1}{\sqrt{n}s}(4V_1^2 - s^2 + 8\tilde{V}_1)\right]$$

$$\times\left[E\left\{V_2(e^{it\frac{2V_2}{\sqrt{n}s}} - 1 - it\frac{V_2}{\sqrt{n}s})\right\} + itE\frac{V_2^2}{\sqrt{n}s}\right]$$

$$= \{\gamma(t)\}^{n-2}\left[itE\frac{V_1}{\sqrt{n}s}(4V_1^2 - s^2 + 8\tilde{V}_1) + O(\frac{|t|^2}{n})\right]\times\left[it\frac{E(V_2^2)}{\sqrt{n}s} + O(\frac{|t|^2}{n})\right]$$

$$= \{\gamma(t)\}^{n-2}\left[\frac{it}{\sqrt{n}s}\{E(4V_1^3) + 8E(W_{12}V_1V_2)\} + O(\frac{|t|^2}{n})\right]\times\left[\frac{its}{\sqrt{n}} + O(\frac{|t|^2}{n})\right]$$

$$= \{\gamma(t)\}^{n-2}\left[\frac{(it)^2}{n}\{E(4V_1^3)+8E(\tilde{W}_{12}V_1V_2)\}+O\left(\frac{|t|^3}{n^{3/2}}+\frac{|t|^4}{n^2}\right)\right] \quad \text{(C.9)}$$

by (2.6). Therefore, substituting (C.9) into (C.6) yields

$$(B)'' = -\frac{n(n-1)}{n^{3/2}s^3}$$
$$\times\left[\frac{(it)^2}{n}\{4E(V_1^3)+8E(\tilde{W}_{12}V_1V_2)\}+O\left(\frac{|t|^3}{n^{3/2}}+\frac{|t|^4}{n^2}\right)\right]\{\gamma(t)\}^{n-2}\ . \quad \text{(C.10)}$$

By (C.8) and (C.10),

$$(B) = -\frac{\{\gamma(t)\}^{n-1}}{s^3}\left\{\frac{4E(V_1^3)+8E(\tilde{W}_{12}V_1V_2)}{n^{1/2}}+O\left(\frac{|t|}{n}\right)\right\}$$
$$-\frac{\{\gamma(t)\}^{n-2}}{s^3}\left[\frac{(it)^2}{n^{1/2}}\{4E(V_1^3)+8E(\tilde{W}_{12}V_1V_2)\}\right.$$
$$\left.+O\left(\frac{t^2+|t|^3}{n}+\frac{t^4}{n^{3/2}}\right)\right] \quad \text{(C.11)}$$

Now, write

$$(C) = -\frac{4}{n^{1/2}s^3}E\left\{\binom{n-1}{2}^{-1}\sum_{1\le j<k\le n}\tilde{W}_{jk}\right\}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}V_l e^{itb_2}$$
$$= -\frac{4}{n^{1/2}s^3}\binom{n-1}{2}^{-1}\sum_{l=1}^{n}\sum_{j<k}{}^{(l)}E(\tilde{W}_{jk}V_l e^{itb_2})$$
$$-\frac{4}{n^{1/2}s^3}\binom{n-1}{2}^{-1}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}E(\tilde{W}_{jk}V_j e^{itb_2})$$
$$-\frac{4}{n^{1/2}s^3}\binom{n-1}{2}^{-1}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}E(\tilde{W}_{jk}V_k e^{itb_2})$$
$$= (C)'+(C)''+(C)'''.$$

Using (iii), Lemma 1-(d), $E(V_j)=0$ and $E(\tilde{W}_{km})=E(\tilde{W}_{km}|k)=E(\tilde{W}_{km}|m)=0$, the summand of $(C)'$ is

$$E\left\{\tilde{W}_{12}V_3 e^{it\frac{2}{\sqrt{ns}}(V_1+V_2+V_3)}\right\}E\left(e^{it\frac{2}{\sqrt{ns}}\sum_{l\ne 1,2,3}^{n}V_l}\right)$$
$$=E\left\{\tilde{W}_{12}V_3 e^{it\frac{2}{\sqrt{ns}}(V_1+V_2+V_3)}\right\}\{\gamma(t)\}^{n-3}$$
$$=\left\{E\tilde{W}_{12}\left(e^{it\frac{2V_1}{\sqrt{ns}}}-1-it\frac{2V_1}{\sqrt{ns}}\right)\left(e^{it\frac{2V_2}{\sqrt{ns}}}-1-it\frac{2V_2}{\sqrt{ns}}\right)\right.$$
$$\left.+(it)^2\frac{4}{ns^2}E(\tilde{W}_{12}V_1V_2)+(it)\frac{2}{\sqrt{ns}}E\{\tilde{W}_{12}V_1\left(e^{it\frac{2V_2}{\sqrt{ns}}}-1-it\frac{2V_2}{\sqrt{ns}}\right)\}\right.$$

$$+(it)\frac{2}{\sqrt{ns}}E\{\tilde{W}_{12}V_2(e^{it\frac{2V_1}{\sqrt{ns}}}-1-it\frac{2V_1}{\sqrt{ns}})\}\Bigg\}$$

$$\times\left\{E\{V_3(e^{it\frac{2V_3}{\sqrt{ns}}}-1-it\frac{2V_3}{\sqrt{ns}})\}+(it)\frac{2}{\sqrt{ns}}E(V_3^2)\right\}\{\gamma(t)\}^{n-3}$$

$$=\left\{(it)^2\frac{4}{ns^2}E(\tilde{W}_{12}V_1V_2)+O(\frac{t^4}{n^2h^{d+2}}+\frac{|t|^3}{n^{3/2}h^{d+2}})\right\}$$

$$\times\left\{(it)\frac{2}{\sqrt{ns}}E(V_3^2)+O(\frac{t^2}{n})\right\}\{\gamma(t)\}^{n-3}\ .$$

The last equality uses (2.6) and

$$E|\tilde{W}_{12}V_1^2V_2^2|\leq\{E|\tilde{W}_{12}|^3\}^{1/2}(E|V_1V_2|^3)^{2/3}\leq Ch^{-\frac{2}{3}d-2}=O(h^{-d-2})$$

due to Hölder's inequality, Lemma 17-(b), (i), (iii) and Lemma 1-(d).

$$(C)'=-\frac{\{\gamma(t)\}^{n-3}}{s^3}\frac{8n(n-1)(n-2)}{n^{5/2}}$$

$$\times\left\{\frac{8(it)^3}{n^{3/2}s^3}E(\tilde{W}_{12}V_1V_2)s^2+O(\frac{t^6}{n^3h^{d+2}}+\frac{|t|^5}{n^{5/2}h^{d+2}}+\frac{t^4}{n^2h^{d+2}})\right\}$$

$$=\frac{\{\gamma(t)\}^{n-3}}{s^3}\left[O\left(\frac{|t|^3}{n}+\frac{t^6}{n^3h^{d+2}}+\frac{|t|^5}{n^{5/2}h^{d+2}}+\frac{t^4+|t|^3}{n^2h^{d+2}}\right)\right]\ . \qquad (C.12)$$

Here we use, due to (iii) and Lemma 16,

$$E(\tilde{W}_{12}V_1V_2)=E(W_{13}W_{23}V_1V_2)$$

$$=E[\,E(W_{13}V_1|3)E(W_{23}V_2|3)\,]=E(\tilde{V}_3^2)<C. \qquad (C.13)$$

The summand of (C)" can be expressed as follows using (iii), $E(\tilde{W}_{12}V_1e^{it\frac{2}{\sqrt{ns}}V_1})=0$, Lemma 17-(b) and (2.6).

$$E(\tilde{W}_{jk}V_je^{itb_2})=\{\gamma(t)\}^{n-2}E(\tilde{W}_{12}V_1e^{it\frac{2}{\sqrt{ns}}(V_1+V_2)})$$

$$=\{\gamma(t)\}^{n-2}E\left\{\tilde{W}_{12}V_1e^{it\frac{2V_1}{\sqrt{ns}}}(e^{it\frac{2V_2}{\sqrt{ns}}}-1-it\frac{2V_2}{\sqrt{ns}})+\frac{2it}{\sqrt{ns}}\tilde{W}_{12}V_1V_2\right\}$$

$$=\{\gamma(t)\}^{n-2}\left\{\frac{2it}{\sqrt{ns}}E(\tilde{W}_{12}V_1V_2)+O(\frac{|t|^2}{nh^{d+2}})\right\}\ .$$

Thus, using (3.8) and (C.13),

$$(C)''=\frac{\{\gamma(t)\}^{n-2}}{s^3}\left[-\frac{8n(n-1)}{n^{5/2}}\left\{\frac{2it}{\sqrt{ns}}E(\tilde{W}_{12}V_1V_2)+O(\frac{t^2}{nh^{d+2}})\right\}\right]$$

$$=\frac{\{\gamma(t)\}^{n-2}}{s^3}O(\frac{|t|}{n}+\frac{t^2}{n^{3/2}h^{d+2}})\ . \qquad (C.14)$$

Similarly,

$$(C)''' = \frac{\{\gamma(t)\}^{n-2}}{s^3} O\left(\frac{|t|}{n} + \frac{t^2}{n^{3/2}h^{d+2}}\right) . \tag{C.15}$$

By (C.12), (C.14) and (C.15),

$$(C) = (C)' + (C)'' + (C)'''$$
$$= \frac{\{\gamma(t)\}^{n-3}}{s^3}\left[O\left(\frac{|t|^3+|t|}{n} + \frac{t^2}{n^{3/2}h^{d+2}} + \frac{t^6}{n^3h^{d+2}} + \frac{|t|^5}{n^{5/2}h^{d+2}} + \frac{t^4+|t|^3}{n^2h^{d+2}}\right)\right] . \tag{C.16}$$

Therefore, by (3.8), (C.1), (C.4), (C.11) and (C.16),

$$E(\tilde{b}_2' e^{itb_2}) = (A) + (B) + (C)$$
$$= -\{\gamma(t)\}^{n-1}\left\{it\,\frac{2}{n}E(W_{12}^2) + O\left(\frac{|t|}{n^2h^{d+2}} + \frac{t^2}{n^{3/2}h^{d+2}} + \frac{|t|h^L}{nh^{d+2}}\right)\right\}$$
$$- \{\gamma(t)\}^{n-1}\left\{\frac{4E(V_1^3) + 8E(W_{12}V_1V_2)}{n^{1/2}} + O\left(\frac{|t|}{n}\right)\right\}$$
$$- \{\gamma(t)\}^{n-2}\left[\frac{(it)^2}{n^{1/2}}\{4E(V_1^3) + 8E(W_{12}V_1V_2)\} + O\left(\frac{t^2+|t|^3}{n} + \frac{t^4}{n^{3/2}}\right)\right]$$
$$+ \{\gamma(t)\}^{n-3}\left[O\left(\frac{|t|^3+|t|}{n} + \frac{t^2}{n^{3/2}h^{d+2}} + \frac{t^6}{n^3h^{d+2}} + \frac{|t|^5}{n^{5/2}h^{d+2}} + \frac{t^4+|t|^3}{n^2h^{d+2}}\right)\right] .$$

(b) Writing, using (3.8) and (A.5),

$$|E(\tilde{b}_{2m} e^{it\bar{B}_m})| \le \frac{C}{n^{3/2}h^{d+2}}\sum_{j=1}^{m} |E(V_j e^{it\bar{B}_m})|$$
$$+ \frac{C}{n^{3/2}}\left\{\sum_{j=1}^{n}\sum_{k=1}^{m} |E(d_jV_k e^{it\bar{B}_m})| + \sum_{j=1}^{m}\sum_{k=m+1}^{n} |E(d_jV_k e^{it\bar{B}_m})|\right\}$$
$$+ \frac{C}{n^{5/2}}\left\{\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sum_{s=1}^{m} |E(e_{jk}V_s e^{it\bar{B}_m})| + \sum_{j=1}^{m}\sum_{k=j+1}^{n}\sum_{s=m+1}^{n} |E(e_{jk}V_s e^{it\bar{B}_m})|\right\}$$
$$+ \frac{C}{n^{7/2}}\left[\sum_{j=1}^{n}\sum_{k<l}^{n\,(j)}\sum_{s=1}^{m} |E(V_jW_{kl}V_s e^{it\bar{B}_m})|\right.$$
$$+ \sum_{j=1}^{n}\sum_{k=1}^{m\,(j)}\sum_{l=k+1}^{n\,(j)}\sum_{s=m+1}^{n} |E(V_jW_{kl}V_s e^{it\bar{B}_m})|$$
$$\left. + \sum_{j=1}^{m}\sum_{k=m+1}^{n-1\,(j)}\sum_{l=k+1}^{n\,(j)}\sum_{s=m+1}^{n} |E(V_jW_{kl}V_s e^{it\bar{B}_m})|\right] , \tag{C.17}$$

$$|E(V_j e^{it\bar{B}_m})| = |E(V_j e^{it\frac{2}{\sqrt{ns}}V_j}) E\{e^{it(\frac{2}{\sqrt{ns}}\sum_{k\ne j}^{n}V_k + b_3 - b_{3m} + \tilde{b}_2 - \tilde{b}_{2m} + \tilde{b}_3' - \tilde{b}_{3m}')}\}|$$
$$\le E|V_j|\,|\gamma(t)|^{m-1} \tag{C.18}$$

for $j = 1, \cdots, m$, since $b_3 - b_{3m} + \tilde{b}_2' - b_{2m}' + \tilde{b}_3' - \tilde{b}_{3m}'$ is independent of $V_1, \cdots, V_m$.

For $j \le m$, $k \le m$ and $j \ne k$,

$$|E(d_j V_k e^{it\bar{B}_m})|$$
$$= |E[d_j V_k e^{it\{\frac{2}{\sqrt{ns}}(V_j + V_k + \sum_{l=m+1}^{n} V_l) + b_3 - b_{3m} + \bar{b}_2 - \bar{b}_{2m} + \bar{b}_3' - \bar{b}_{3m}'\}}]\,||E\{e^{it\frac{2}{\sqrt{ns}}\sum_{l\neq j,k}^{m} V_l}\}|$$
$$\leq E|d_j V_k||\gamma(t)|^{m-2} \ . \tag{C.19}$$

For $j = k \leq m$,

$$|E(d_j V_j e^{it\bar{B}_m})|$$
$$= |E[d_j V_j e^{it\{\frac{2}{\sqrt{ns}}(V_j + \sum_{k=m+1}^{n} V_k) + b_3 - b_{3m} + \bar{b}_2 - \bar{b}_{2m} + \bar{b}_3' - \bar{b}_{3m}'\}}]\,||E\{e^{it\frac{2}{\sqrt{ns}}\sum_{k\neq j}^{m} V_k}\}|$$
$$\leq E|d_j V_j||\gamma(t)|^{m-1} \quad \leq E|d_j V_j||\gamma(t)|^{m-2} \ . \tag{C.20}$$

For $j \leq m$ and $k \geq m+1$,

$$|E(d_j V_k e^{it\bar{B}_m})|$$
$$= |E[d_j V_k e^{it\{\frac{2}{\sqrt{ns}}(V_j + \sum_{l=m+1}^{n} V_l) + b_3 - b_{3m} + \bar{b}_2 - \bar{b}_{2m} + \bar{b}_3' - \bar{b}_{3m}'\}}]\,||E\{e^{it\frac{2}{\sqrt{ns}}\sum_{l\neq j}^{m} V_l}\}|$$
$$\leq E|d_j V_k||\gamma(t)|^{m-1} \quad \leq E|d_j V_k||\gamma(t)|^{m-2} \ . \tag{C.21}$$

For $j \geq m+1$ and $k \leq m$, similarly to (C.21),

$$|E(d_j V_k e^{it\bar{B}_m})| \quad \leq E|d_j V_k||\gamma(t)|^{m-2} \tag{C.22}$$

Therefore, by (C.19)-(C.22) and Lemma 18, for all j, k,

$$|E(d_j V_k e^{it\bar{B}_m})| \leq E|d_j V_k|\,|\gamma(t)|^{m-2} \tag{C.23}$$

Similarly to the derivation of (C.23), for any j, k, l, s,

$$|E(e_{jk} V_s e^{it\bar{B}_m})| \leq E|e_{jk} V_s|\,|\gamma(t)|^{m-3} \ , \tag{C.24}$$

$$|E(V_j W_{kl} V_s e^{it\bar{B}_m})| \leq E|V_j W_{kl} V_s|\,|\gamma(t)|^{m-4} \ . \tag{C.25}$$

Substituting (C.18), (C.23)-(C.25) into (C.17), using $|\gamma(t)| \leq 1$,

$$|E(\bar{b}_{2m} e^{it\bar{B}_m})| \leq C\,|\gamma(t)|^{m-4} \times$$
$$\Bigg[\frac{1}{n^{3/2}h^{d+2}}\sum_{j=1}^{m} E|V_j| + \frac{1}{n^{3/2}}\left(\sum_{j=1}^{n}\sum_{k=1}^{m} E|d_j V_k| + \sum_{j=1}^{m}\sum_{k=m+1}^{n} E|d_j V_k|\right)$$
$$+ \frac{1}{n^{5/2}}\left(\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sum_{s=1}^{m} E|e_{jk} V_s| + \sum_{j=1}^{m}\sum_{k=j+1}^{n}\sum_{s=m+1}^{n} E|e_{jk} V_s|\right)$$
$$+ \frac{1}{n^{7/2}}\Bigg(\sum_{j=1}^{n}\sum_{k<l}^{n\,(j)}\sum_{s=1}^{m} E|V_j W_{kl} V_s| + \sum_{j=1}^{n}\sum_{k=1}^{m\,(j)}\sum_{l=k+1}^{n\,(j)}\sum_{s=m+1}^{n} E|V_j W_{kl} V_s|$$
$$+ \sum_{j=1}^{m}\sum_{k=m+1}^{n-1\,(j)}\sum_{l=k+1}^{n\,(j)}\sum_{s=m+1}^{n} E|V_j W_{kl} V_s|\Bigg)\Bigg] \ . \tag{C.26}$$

The summations in the square brackets have the following bounds.

$$\sum_{j=1}^{m} E|V_j| \le Cm \quad \text{by Lemma 1-(d).} \tag{C.27}$$

$$\sum_{j=1}^{n}\sum_{k=1}^{m} E|d_j V_k| = \sum_{j=1}^{m} E|d_j V_j| + \sum_{j=1}^{n} \sum_{k=1}^{m(j)} E|d_j V_k| \tag{C.28}$$

$$\le C(m+mn) \quad \text{by Lemma 18.}$$

$$\sum_{j=1}^{m} \sum_{s=m+1}^{n} E|d_j V_s| \le Cmn \quad \text{by Lemma 18-(a).} \tag{C.29}$$

$$\sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \sum_{s=1}^{m} E|e_{jk} V_s| = \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} \sum_{s=1}^{m(j,k)} E|e_{jk} V_s| + \sum_{j=1}^{m} \sum_{k=j+1}^{n} E|e_{jk} V_j| + \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} E|e_{jk} V_k|$$

$$\le C(mn^2 + mn + m^2) h^{-2} \tag{C.30}$$

by Lemma 20, $\sum_{s}^{(i_1, i_2, \cdots, i_r)}$ denoting summations excluding $s = i_1, i_2, \cdots, i_r$.

$$\sum_{j=1}^{m} \sum_{k=j+1}^{n} \sum_{s=m+1}^{n} E|e_{jk} V_s| = \sum_{j=1}^{m} \sum_{k=j+1}^{n} \sum_{s=m+1}^{n(j)} E|e_{jk} V_s| + \sum_{j=1}^{m} \sum_{k=j+1}^{n} E|e_{jk} V_k|$$

$$\le C(mn^2 + mn) h^{-2} \quad \text{by Lemma 20.} \tag{C.31}$$

$$\sum_{j=1}^{n} \sum_{k<l}^{n(j)} \sum_{s=1}^{m} E|V_j W_{kl} V_s|$$

$$= \sum_{j=1}^{n} \sum_{k<l}^{n(j)} \sum_{s=1}^{m(j,k,l)} E|V_j| E|W_{kl}| E|V_s| + \sum_{j=1}^{m} \sum_{k<l}^{n(j)} E|V_j^2 W_{kl}| + \sum_{j=1}^{n} \sum_{k<l}^{m(j)} E|V_j W_{kl} V_k| + \sum_{j=1}^{n} \sum_{k<l}^{m(j)} E|V_j W_{kl} V_l|$$

$$\le C(mn^3 + mn^2 + m^2 n) h^{-1} \tag{C.32}$$

by (iii), Lemmas 1-(d), 19 and (A.5).

$$\sum_{j=1}^{n} \sum_{k=1}^{m(j)} \sum_{l=k+1}^{n(j)} \sum_{s=m+1}^{n} E|V_j W_{kl} V_s|$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{m(j)} \sum_{l=k+1}^{n(j)} \sum_{s=m+1}^{n(j,l)} E|V_j| E|W_{kl}| E|V_s| + \sum_{j=m+1}^{n} \sum_{k=1}^{m(j)} \sum_{l=k+1}^{n(j)} E|V_j^2 W_{kl}| + \sum_{j=1}^{n} \sum_{k=1}^{m(j)} \sum_{l=k+1}^{n(j)} E|V_j W_{kl} V_l|$$

$$\le C(mn^3 + mn^2) h^{-1} \tag{C.33}$$

by (iii), Lemmas 1-(d), 19 and (A.5).

$$\sum_{j=1}^{m} \sum_{m<k<l}^{n\,(j)} \sum_{s=m+1}^{n} E|V_j W_{kl} V_s|$$

$$= \sum_{j=1}^{m} \sum_{m<k<l}^{n\,(j)} \sum_{s=m+1}^{n\ (k,l)} E|V_j| E|W_{kl}| E|V_s|$$

$$+ \sum_{j=1}^{m} \sum_{m<k<l}^{n\,(j)} (E|V_j W_{kl} V_k| + E|V_j W_{kl} V_l|)$$

$$\leq C (mn^3 + mn^2) h^{-1} \tag{C.34}$$

by (iii), Lemmas 1-(d), 19 and (A.5). Therefore, substituting (C.27)-(C.34) into (C.26), using $1 \leq m \leq n-1$,

$$|E(\tilde{b}^{\,l}_{2m} e^{\,i t \bar{B}_m}) | \leq C |\gamma(t)|^{m-4} \left(\frac{m}{n^{3/2} h^{d+2}} + \frac{mn}{n^{3/2}} + \frac{mn^2}{n^{5/2} h^2} + \frac{mn^3}{n^{7/2} h} \right)$$

$$\leq \frac{C_1 m}{n^{1/2} h^2} |\gamma(t)|^{m-4} , \tag{C.35}$$

the third term in parentheses dominating the others for sufficiently large n by assumption (ix).

(c) Using (3.8) and (A.5), we start with writing

$$|E(\tilde{b}^{\,l}_{3m} e^{\,i t \bar{B}_m}) |$$

$$\leq C \Bigg[\frac{1}{n^{5/2} h^{d+2}} \sum_{l=1}^{m} \sum_{s=l+1}^{n} |E(W_{ls} e^{\,i t \bar{B}_m}) |$$

$$+ \frac{1}{n^{5/2}} \left(\sum_{j=1}^{m} \sum_{l<s}^{n} |E(d_j W_{ls} e^{\,i t \bar{B}_m}) | + \sum_{j=m+1}^{n} \sum_{l=1}^{m} \sum_{s=l+1}^{n} |E(d_j W_{ls} e^{\,i t \bar{B}_m}) | \right)$$

$$+ \frac{1}{n^{7/2}} \left(\sum_{j=1}^{m} \sum_{k=j+1}^{n} \sum_{l<s}^{n} |E(\tilde{W}_{jk} W_{ls} e^{\,i t \bar{B}_m}) | + \sum_{m<j<k}^{n} \sum_{l=1}^{m} \sum_{s=l+1}^{n} |E(\tilde{W}_{jk} W_{ls} e^{\,i t \bar{B}_m}) | \right) \Bigg] . \tag{C.36}$$

Similarly to (C.23)-(C.25), for all j, k, l, s,

$$|E(W_{ls} e^{\,i t \bar{B}_m}) | \leq E|W_{ls}| \, |\gamma(t)|^{m-2} , \tag{C.37}$$

$$|E(d_j W_{ls} e^{\,i t \bar{B}_m}) | \leq E|d_j W_{ls}| \, |\gamma(t)|^{m-3} , \tag{C.38}$$

$$|E(\tilde{W}_{jk} W_{ls} e^{\,i t \bar{B}_m}) | \leq E|\tilde{W}_{jk} W_{ls}| \, |\gamma(t)|^{m-4} . \tag{C.39}$$

Substituting (C.37)-(C.39) into (C.36), we have, due to $|\gamma(t)| \le 1$,

$$|E(\tilde{b}_{3m}' e^{it\bar{B}_m})|$$

$$\le C|\gamma(t)|^{m-4}\left[\frac{1}{n^{5/2}h^{d+2}}\sum_{l=1}^{m}\sum_{s=l+1}^{n}E|W_{ls}| + \frac{1}{n^{5/2}}\left(\sum_{j=1}^{m}\sum_{l<s}^{n}E|d_j W_{ls}| + \sum_{j=m+1}^{n}\sum_{l=1}^{m}\sum_{s=l+1}^{n}E|d_j W_{ls}|\right) + \frac{1}{n^{7/2}}\left(\sum_{j=1}^{m}\sum_{k=j+1}^{n}\sum_{l<s}^{n}E|\tilde{W}_{jk}W_{ls}| + \sum_{j=m+1}^{n-1}\sum_{k=j+1}^{n}\sum_{l=1}^{m}\sum_{s=l+1}^{n}E|\tilde{W}_{jk}W_{ls}|\right)\right].$$

Applying (A.5), Lemmas 21 and 22, and (ix),

$$|E(\tilde{b}_{3m}' e^{it\bar{B}_m})| \le C|\gamma(t)|^{m-4}\left\{\frac{mn}{n^{5/2}h^{d+3}} + \frac{mn^2}{n^{5/2}h} + \frac{1}{n^{7/2}}\left(\frac{mn}{h^{d+3}} + \frac{mn^3}{h^3}\right)\right\}$$

$$= Cm|\gamma(t)|^{m-4}\left(\frac{1}{n^{3/2}h^{d+3}} + \frac{1}{n^{1/2}h} + \frac{1}{n^{5/2}h^{d+3}} + \frac{1}{n^{1/2}h^3}\right)$$

$$\le \frac{Cm}{n^{1/2}h^3}|\gamma(t)|^{m-4} .$$

(d) Write, using (3.8) and (A.5),

$$E|\tilde{b}_{2m}|^2 \le C\left[\frac{1}{n^3h^{2d+4}}E|\sum_{i=1}^{m}V_i|^2 + \frac{1}{n^3}\left(E|\sum_{i=1}^{n}\sum_{s=1}^{m}d_iV_s|^2 + E|\sum_{i=1}^{m}\sum_{s=m+1}^{n}d_iV_s|^2\right) + \frac{1}{n^5}\left(E|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{s=1}^{m}e_{ij}V_s|^2 + E|\sum_{i=1}^{m}\sum_{j=i+1}^{n}\sum_{s=m+1}^{n}e_{ij}V_s|^2\right) + \frac{1}{n^7}\left(E|\sum_{i=1}^{n}\sum_{k<l}^{(i)}\sum_{s=1}^{m}V_iW_{kl}V_s|^2 + E|\sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{l=k+1}^{n\,(i)}\sum_{s=m+1}^{n\,(i)}V_iW_{kl}V_s|^2 + E|\sum_{i=1}^{m}\sum_{k=m+1}^{n-1\,(i)}\sum_{l=k+1}^{n\,(i)}\sum_{s=m+1}^{n}V_iW_{kl}V_s|^2\right)\right]. \qquad (C.40)$$

We show bounds only of some typical terms. Since V_i is an iid sequence with zero mean, due to Lemma 1-(d), $E|\sum_{i=1}^{m}V_i|^2 = mE|V_1|^2 \le Cm$. Writing

$$E|\sum_{i=1}^{n}\sum_{s=1}^{m}d_iV_s|^2 \le C\left(E|\sum_{i=1}^{m}d_iV_i|^2 + E|\sum_{i=1}^{m-1}\sum_{s=i+1}^{m}d_iV_s|^2 + E|\sum_{s=1}^{m}\sum_{i=s+1}^{n}d_iV_s|^2\right), \qquad (C.41)$$

the first term in parentheses is bounded by

$$mE|d_1V_1|^2 + m(m-1)E|d_1V_1|E|d_2V_2| \le Cm^2 \qquad (C.42)$$

due to (iii) and Lemma 18-(b). Since d_i and V_s are iid with zero mean,

$$E|\sum_{i=1}^{m-1}\sum_{s=i+1}^{m} d_i V_s|^2 = \sum_{i=1}^{m-1} E(d_i^2)\sum_{s=i+1}^{m} E(V_s^2) \le C m^2 \tag{C.43}$$

by Lemma 1-(d) and (B.6) under (i)'. Similarly, using Lemma 18-(a),

$$E|\sum_{s=1}^{m}\sum_{i=s+1}^{n} d_i V_s|^2 \le \sum_{s=1}^{m}\sum_{i=s+1}^{n} E(d_i^2)E(V_s^2) \le Cmn \ . \tag{C.44}$$

From (C.41)-(C.44),

$$E|\sum_{i=1}^{n}\sum_{s=1}^{m} d_i V_s|^2 \le C(m^2+mn) \ .$$

Similarly,

$$E|\sum_{i=1}^{m}\sum_{j=m+1}^{n} d_i V_s|^2 = \sum_{i=1}^{m} E(d_i^2)\sum_{s=m+1}^{n} E(V_s^2) \le C mn.$$

We next consider

$$E|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{s=1}^{m} e_{ij} V_s|^2 \le C\Big\{E|\sum_{s=1}^{m-1}\sum_{i=s+1}^{n-1}\sum_{j=i+1}^{n} e_{ij} V_s|^2 + E|\sum_{i=1}^{m}\sum_{j=i+1}^{n} e_{ij} V_i|^2$$

$$+E|\sum_{i=1}^{m-1}\sum_{s=i+1}^{m}\sum_{j=s+1}^{n} e_{ij} V_s|^2 + E|\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} e_{ij} V_j|^2 + E|\sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1}\sum_{s=j+1}^{m} e_{ij} V_s|^2\Big\} \ . \tag{C.45}$$

Due to (iii), $E(e_{ij}\,|i) = E(e_{ij}\,|j) = 0$, $E(V_s) = 0$ and Lemma 20, the triple summation terms on the right of (C.45) is $O((m^3+m^2n+mn^2)h^{-d-4})$. Using Lemma 20 and Hölder's inequality, the second term in (C.45) equals

$$\sum_{i=1}^{m}\sum_{j=i+1}^{n} E(e_{ij} V_i)^2 + 2\sum_{i=1}^{m-1}\sum_{k=i+1}^{m}\sum_{j=k+1}^{n} E(e_{ij} V_i e_{kj} V_k)$$

$$\le C[mnE(e_{12}V_1)^2 + m^2n\{E(e_{13}V_1)^2 E(e_{23}V_2)^2\}^{1/2}] \le Cm^2nh^{-d-4} \ . \tag{C.46}$$

Similarly, the fourth term of (C.45) is $O(m^3h^{-d-4})$. Using Lemma 19, as above, the terms involving $V_i W_{kl} V_s$ in (C.40) are $O((m^4+m^3n+m^2n^2+mn^3)h^{-d-2})$, so by (ix)

$$E|\tilde{b}_{2m}|^2 \le \frac{C}{nh^{d+2}}\left(\frac{m}{n^2h^{d+2}}\right) + \frac{C}{n^3}(m^2+mn)$$
$$+ \frac{C}{n^5}(m^3+m^2n+mn^2)h^{-d-4}$$
$$+ \frac{C}{n^7}(m^4+m^3n+m^2n^2+mn^3)h^{-d-2}$$
$$\le C\, m\Big(\frac{1}{n^3h^{2d+4}} + \frac{1}{n^2}\Big) \ .$$

(e) The derivation is similar using (A.5), Lemmas 21 and 22. As in (d), we can show

$$E|\tilde{b}'_{3m}|^2 \le \frac{C}{n^5h^{2d+4}} mnh^{-d-2} + \frac{C}{n^5}(m^3+m^2n+mn^2)h^{-d-2}$$
$$+ \frac{C}{n^7}(m^4+m^3n+m^2n^2+mn^3)h^{-3d-6}$$

$$\leq \frac{Cm}{n^4 h^{3d+6}} \; .$$

(f) Write

$$|E\delta_3'' e^{it\bar{B}_m}| = |EQ\bar{W}e^{it\bar{B}_m}| \leq |EQ_1\bar{W}e^{it\bar{B}_m}| + |EQ_2\bar{W}e^{it\bar{B}_m}| \; . \qquad \text{(C.47)}$$

By (3.8),

$$|E(Q_1\bar{W}e^{it\bar{B}_m})|$$

$$\leq \frac{C}{n^{7/2}}|\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}\sum_{l=1}^{n-1}\sum_{s=l+1}^{n} E\{(V_j+V_k)W_{jk} - \tilde{V}_j - \tilde{V}_k\}W_{ls}e^{it\bar{B}_m}|$$

$$\leq \frac{6C}{n^{7/2}}\sum_{j=1}^{n-3}\sum_{k=j+1}^{n-2}\sum_{l=k+1}^{n-1}\sum_{s=l+1}^{n} E|\{(V_j+V_k)W_{jk} - \tilde{V}_j - \tilde{V}_k\}W_{ls}||\gamma(t)|^{m-4}$$

$$+ \frac{6C}{n^{7/2}}\sum_{j=1}^{n-2}\sum_{k=j+1}^{n-1}\sum_{s=k+1}^{n} E|\{(V_j+V_k)W_{jk} - \tilde{V}_j - \tilde{V}_k\}W_{ks}||\gamma(t)|^{m-3}$$

$$+ \frac{C}{n^{7/2}}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} E|\{(V_j+V_k)W_{jk} - \tilde{V}_j - \tilde{V}_k\}W_{jk}||\gamma(t)|^{m-2}$$

$$\leq Cn^{1/2}E|\{(V_1+V_2)W_{12} - \tilde{V}_1 - \tilde{V}_2\}W_{34}||\gamma(t)|^{m-4}$$

$$+ \frac{C}{n^{1/2}}E|\{(V_1+V_2)W_{12} - \tilde{V}_1 - \tilde{V}_2\}W_{13}||\gamma(t)|^{m-3}$$

$$+ \frac{C}{n^{3/2}}E|\{(V_1+V_2)W_{12} - \tilde{V}_1 - \tilde{V}_2\}W_{12}||\gamma(t)|^{m-3} \; . \qquad \text{(C.48)}$$

Using (i)', (iii), Lemmas 1-(d), 4, (B.5) and (A.5), the first expectation of (C.48) is bounded by

$$CE\{(|Y_1|+|Y_2|+1)|W_{12}|\}E|W_{34}| \leq Ch^{-2} \; .$$

Using (i)', Lemmas 1-(d), 4, (B.5) and (A.5), the second expectation of (C.48) is bounded by

$$CE\{(|Y_1|+|Y_2|+1)|W_{12}||W_{13}|\}$$

$$\leq CE\{(|Y_1|+|Y_2|+1)|W_{12}|E(|W_{13}|\,|1)\}$$

$$\leq Ch^{-1}E\{(|Y_1|+|Y_2|+1)(|Y_1|+1)|W_{12}|\} \; .$$

Similarly to Lemma 15 and (B.9), $E|Y_1W_{12}|+E|Y_1^2W_{12}|+E|Y_1Y_2W_{12}| = O(h^{-1})$ so that the above quantity is $O(h^{-2})$. The third expectation of (C.48) is bounded by

$$CE|V_1W_{12}^2| + E|\tilde{V}_1W_{12}| \leq C(h^{-d-2}+h^{-1}) = O(h^{-d-2})$$

due to Lemmas 1-(d), 4, (A.5) and Lemma 16. Therefore,

$$|E(Q_1\bar{W}e^{it\bar{B}_m})| \quad \leq C(\frac{n^{1/2}}{h^2} + \frac{1}{n^{3/2}h^{d+2}})\,|\gamma(t)|^{m-4} \leq \frac{Cn^{1/2}}{h^2}|\gamma(t)|^{m-4}\,.$$

The second term of (C.47) is bounded by, using (3.8),

$$\frac{C}{n^{9/2}}|\sum_{r=1}^{n}\sum_{j=1}^{n-1\,(r)}\sum_{k=j+1}^{n\,(r)}\sum_{l=1}^{n-1}\sum_{s=l+1}^{n}E(V_r W_{jk} W_{ls}e^{it\bar{B}_m})|$$

$$\leq \frac{C}{n^{9/2}}\sum_{r=1}^{n-4}\sum_{j=r+1}^{n-3}\sum_{k=j+1}^{n-2}\sum_{l=k+1}^{n-1}\sum_{s=l+1}^{n}E|V_r W_{jk} W_{ls}|\,|\gamma(t)|^{m-5}$$

$$+ \frac{C}{n^{9/2}}\sum_{r=1}^{n-3}\sum_{j=r+1}^{n-2}\sum_{k=j+1}^{n-1}\sum_{s=k+1}^{n}E|V_r W_{jk} W_{ks}|\,|\gamma(t)|^{m-4}$$

$$+ \frac{C}{n^{9/2}}\sum_{r=1}^{n-3}\sum_{j=r+1}^{n-2}\sum_{k=j+1}^{n-1}\sum_{s=k+1}^{n}E|V_r W_{jk} W_{rs}|\,|\gamma(t)|^{m-4}$$

$$+ \frac{C}{n^{9/2}}\sum_{r=1}^{n-2}\sum_{j=r+1}^{n-1}\sum_{k=j+1}^{n}E|V_r W_{jk} W_{jk}|\,|\gamma(t)|^{m-3}$$

$$+ \frac{C}{n^{9/2}}\sum_{r=1}^{n-2}\sum_{j=r+1}^{n-1}\sum_{k=j+1}^{n}E|V_r W_{jk} W_{rk}|\,|\gamma(t)|^{m-3}$$

$$\leq Cn^{1/2}E|V_1|E|W_{23}|E|W_{45}|\,|\gamma(t)|^{m-5}$$

$$+ \frac{C}{n^{1/2}}(E|V_1|E|W_{23}W_{24}| + E|V_1W_{14}|E|W_{23}|)\,|\gamma(t)|^{m-4}$$

$$+ \frac{C}{n^{3/2}}(E|V_1|E|W_{23}|^2 + E|V_1W_{23}W_{13}|)\,|\gamma(t)|^{m-3}$$

$$\leq C(\frac{n^{1/2}}{h^2} + \frac{1}{n^{1/2}h^2} + \frac{1}{n^{3/2}h^{d+2}})\,|\gamma(t)|^{m-5}$$

by (i)', (iii), Lemmas 1-(d), 4, (A.5) and Lemma 15. Then apply (ix). □

Bibliography

AHN, H. (1997) "Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors," Econometric Theory, 13, 3-31.

AHN, H. AND C.F.MANSKI (1993) "Distribution Theory for the Analysis of Binary Choice under Uncertainty with Nonparametric Estimation of Expectations," Journal of Econometrics, 56, 291-321.

AHN, H. AND J.L.POWELL (1993) "Semiparametric Estimation of Censored Models with a Nonparametric Selection Mechanism," Journal of Econometrics, 58, 3-29.

AI, C. (1997) "A Semiparametric Maximum Likelihood Estimator," Econometrica, 65, 933-963.

AI, C. AND D.MCFADDEN (1997) "Estimation of Some Partially Specified Nonlinear Models," Journal of Econometrics, 76, 1-37.

AMEMIYA, T. (1973) "Regression Analysis when the Dependent Variable is Truncated Normal," Econometrica, 41, 6, 997-1016.

AMEMIYA, T. (1974) "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal," Econometrica, 42, 6, 999-1012.

AMEMIYA, T. (1977) "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," Econometrica, 45, 4, 955-968.

AMEMIYA, T. (1978) "The Estimation of a Simultaneous Equation Generalized Probit Model," Econometrica, 46, 1193-1205.

AMEMIYA, T. (1979) "The Estimation of a Simultaneous-Equation Tobit Model," International Economic Review, 20, 169-181.

AMEMIYA, T. (1981) "Qualitative Response Models: A Survey," Journal of Economic Literature, 19, 1483-1536.

AMEMIYA, T. (1984) "Tobit Models: A Survey," Journal of Econometrics, 24, 3-61.

AMEMIYA, T. (1985) Advanced Econometrics : Blackwell.

ANDERSON, T.W. (1974) "An Asymptotic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a

Simultaneous Equation System," Journal of the American Statistical Association, 69, 565-573.

ANDERSON, T.W. AND T. SAWA (1973) "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions," Econometrica, 41, 683-714.

ANDREWS, D.W.K. (1991) "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," Econometrica, 59, 307-345.

ANDREWS, D.W.K. (1995) "Nonparametric Kernel Estimation for Semiparametric Models," Econometric Theory, 11, 560-596.

AYER, M., H.BRUNK, G.M. EWING, W.T. REID, AND E. SILVERMAN (1955) "An Empirical Distribution Function for Sampling with Incomplete Information," Annals of Mathematical Statistics, 26, 641-647.

BABU, G.J. AND K. SINGH (1984) "On One Term Edgeworth Correction by Efron's Bootstrap, Sankhya, 46, Series A, 219-232.

BARNETT, W.A., J.L. POWELL, AND G. TAUCHEN (1991) Nonparametric And Semiparametric Methods in Econometrics And Statistics: Cambridge University Press.

BEGUN, J.M., W.J.HALL, W.M. HUANG, AND J.A. WELLNER (1983) "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," Annals of Statistics, 11, 432-452.

BELLMAN, R.E. (1961) Adaptive Control Processes. Princeton University Press.

BENTKUS, V., F.GÖTZE, AND W.R. VAN ZWET (1997) "An Edgeworth Expansion for Symmetric Statistics," Annals of Statistics, 25, 851-896.

BERAN, R. (1982) "Estimated Sampling Distributions: the Bootstrap and Competitors," Annals of Statistics, 10, 212-225.

BERAN, R. (1984) "Jackknife Approximations to Bootstrap Estimates," Annals of Statistics, 12, 101-118.

BERK, R.H. (1966) "Limiting Behaviour of Posterior Distributions When the Models is Incorrect," Annals of Mathematical Statistics, 37, 51-58.

BERRY, A.C. (1941) "The Accuracy of the Gaussian Approximation to the Sum of Independent Variables," Transactions of the American Mathematical Society, 49, 122-136.

BHATTACHARYA, P.K. (1967) "Estimation of a Probability Density Function and

Its Derivatives," Sankhya, Ser. A, 29, 373-382.

BHATTACHARYA, R.N. (1968) "Berry-Esseen Bounds for the Multidimensional Central Limit Theorem," Bulletin of the American Mathematical Society, 75, 285-287.

BHATTACHARYA, R.N. (1971) "Rates of Weak Convergence and Asymptotic Expansions with Applications," in P.R. Krishnaiah (ed.), Multivariate Analysis VI, North Holland, 57-75.

BHATTACHARYA, R.N. AND J.K. GHOSH (1978) "On the Validity of the Formal Edgeworth Expansion," Annals of Statistics, 6, 434-451.

BHATTACHARYA, R.N. AND M.L. PURI (1995) "Asymptotic Expansions in Statistics: a Review of Methods and Applications," Advances in Econometrics and Quantitative Economics, ed. by G.S. Maddala et.al., Blackwell.

BHATTACHARYA, R.N. AND M. QUMSIYEH (1989) "Second Order and L_p - Comparisons Between the Bootstrap and Empirical Edgeworth Expansion Methodologies,"Annals of Statistics, 17, 160-169.

BHATTACHARYA, R.N. AND R.RANGA RAO (1976) Normal Approximation and Asymptotic Expansions. Wiley.

BICKEL, P.J. (1974) "Edgeworth Expansions in Nonparametric Statistics," Annals of Statistics, 2, 1-20.

BICKEL, P.J. (1982) "On Adaptive Estimation," Annals of Statistics, 10, 647-671.

BICKEL, P.J. AND D.A. FREEDMAN (1981) "Some Asymptotic Theory for the Bootstrap," Annals of Statistics, 9, 1196-1217.

BICKEL, P.J., F.GÖTZE AND W.R. VAN ZWET (1986) "The Edgeworth Expansion for U-Statistics of Degree Two," Annals of Statistics, 14, 1463-1484.

BICKEL, P.J., C.A.J. KLAASSEN, Y. RITOV AND J.A. WELLNER (1993) Efficient And Adaptive Estimation for Semiparametric Models: Johns Hopkins University Press.

BOX, G.E.P. AND W.J. HILL (1974) "Correcting Inhomogeneity of Variance with Power Transformation Weighting," Technometrics, 16, 385-389.

BREIMAN, L., Y. TSUR AND A. ZEMEL (1987) "Distribution-Free Estimators of Censored Regression Models," Manuscript, Department of Economics, Ben Gurion University of the Negev, Israel.

CACOULLOS, T. (1966) "Estimation of a Multivariate Density," Annals of Institute of Statistical Mathematics, 18, 178-189.

CALLAERT, H. AND P.JANSSEN (1978) "The Berry-Esseen Theorem for *U*-Statistics," Annals of Statistics, 6, 417-421.

CALLAERT, H., P.JANSSEN, AND N.VERAVERBEKE (1980) "An Edgeworth Expansion for *U*-Statistics," Annals of Statistics, 8, 299-312.

CALLAERT, H. AND N.VERAVERBEKE (1981) "The Order of the Normal Approximation for a Studentized *U*-Statistics," Annals of Statistics, 9, 194-200.

CARROLL, R.J. (1982) "Adapting for Heteroscedasticity in Linear Models," Annals of Statistics, 10, 4, 1224-1233.

CARROLL, R.J. AND D. RUPPERT (1982) "Robust Estimation in Heteroscedastic linear Models," Annals of Statistics, 10, 429-441.

CAVANAGH, C.L. (1987) "Limiting Behaviour of Estimators Defined by Optimization," unpublished manuscript.

CHAMBERLAIN, G. (1992) "Efficiency Bounds for Semiparametric Regression," Econometrica, 60, 3, 567-596.

CHAMBERS, J.M. (1967) "On Methods of Asymptotic Approximation for Multivariate Distributions," Biometrika, 54, 367-383.

CHAN, Y.-K. AND J. WIERMAN (1977) "On the Berry-Esseen Theorem for *U*-Statistics," Annals of Statistics, 5, 136-139.

CHAUDHURI, P., K.DOKSUM, AND A.SAMAROV (1991) "On Average Derivative Quantile Regression," Unpublished Manuscript.

CHEBYSHEV, P.L. (1890) "Sur Deux Theoremes Relatifs auf Probabilites," Acta Mathematica, 14, 305-315.

CHEN, S. (1997) "Semiparametric Estimation of the Type-3 Tobit Model," Journal of Econometrics, 80, 1-34.

CHEN, S. AND L.-F. LEE (1998) "Efficient Semiparametric Scoring Estimation of Sample Selection Models," Econometric Theory, 14, 423-462.

CHEN, H.Z. AND A.RANDALL (1997) "Semi-nonparametric Estimation of Binary Response Models with an Application to Natural Resource Valuation," Journal of Econometrics, 76, 323-340.

CHEN, H. AND J.J.H. SHIAU (1991) "A Two-Stage Spline Smoothing Method for

Partially Linear Models," Journal of Statistical Planning and Inference, 27, 187-201.

CHEN, H. AND J.J.H. SHIAU (1994) "Data-Driven Efficient Estimation for a Partially Linear Model," Annals of Statistics, 22, 211-237.

CHENG, B. AND P.M.ROBINSON (1994) "Semiparametric Estimation from Time Series with Long-range Dependence," Journal of Econometrics, 64, 335-353.

COSSLETT, S.R. (1983) "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model," Econometrica, 51, 765-782.

COSSLETT, S.R. (1987) "Efficiency Bounds for Distribution-free Estimators of the Binary Choice and the Censored Regression Models." Econometrica, 55, 559-585.

COSSLETT, S.R. (1991) "Semiparametric Estimation of a Regression Model with Sampling Selectivity," in Nonparametric and Semiparametric Methods in Econometrics and Statistics: Cambridge University Press.

CRAIG, J.G. (1983) "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," Econometrica, 51, 751-763.

CRAMÈR, H. (1928) "On the Composition of Elementary Errors," Skand. Aktuarietidskr, 11, 13-74, 141-180.

CRAMÈR, H. (1946) Mathematical Methods of Statistics, Princeton University Press.

DHAMADHIKARI, S.W., V.FABIAN, AND K.JOGDEO (1968) "Bounds on the Moments of Martingales," Annals of Mathematical Statistics, 39, 1719-1723.

DELGADO, M.A. (1992) "Semiparametric Generalized Least Squares in the Multivariate Nonlinear Regression Model," Econometric Theory, 8, 203-222.

DELGADO, M.A. AND J.MORA (1995) "Nonparametric and Semiparametric Estimation with Discrete Regressors," Econometrica, 63, 1477-1484.

DELGADO, M.A. AND P.M.ROBINSON (1992) "Nonparametric and Semiparametric Methods for Economic Research," Journal of Economic Surveys, 6, 201-249.

DEVROYE, L. (1987) A Course in Density Estimation. Birkhäuser, Boston.

DEVROYE, L. AND L. GYÖRFI (1985) Nonparametric Density Estimation: The L_1 View: John Wiley and Sons.

DEVROYE, L. AND T.J. WAGNER (1980a) "Distribution Free Consistency Results in Nonparametric Discrimination and Regression Function Estimates," Annals of Statistics, 8, 231-239.

DEVROYE, L. AND T.J. WAGNER (1980b) "On the L_1 - Convergence of kernel Estimators of Regression Function with Application in Discrimination," Zeitschrift für Wahrscheinlichkeitstheorie Verwandte Gebiete, 51, 15-25.

EDGEWORTH, F.Y. (1896) "The Asymmetrical Probability Curve," Philosophical Magazine, 5th Series, 41, 90-99.

EDGEWORTH, F.Y. (1905) "The Law of Error," Transactions of the Cambridge Philosophical Society, 20, 36-65, 113-141.

EFRON, B. (1979) "Bootstrap Methods: Another look at the Jackknife," Annals of Statistics," 7, 1-26.

EFRON, B. (1982) The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics.

EFRON, B. (1990) "More Efficient Bootstrap Computation," Journal of the American Statistical Association, 85, 79-89.

EFRON, B. AND C. STEIN (1981) "The Jackknife Estimate of Variance," Annals of Statistics, 9, 586-596.

EFRON, B. AND R.J. TIBSHIRANI (1993) An Introduction to the Bootstrap: Chapman and Hall.

ENGLE, R.F., C.W.J.GRANGER, J.RICE, AND A.WEISS (1986) "Semiparametric Estimates of the Relation Between Weather and Electricity Demand," Journal of the American Statistical Association, 81, 310-320.

EPANECHNIKOV, V.A. (1969) "Nonparametric Estimation of a Multi-dimensional Probability Density," Theory of Probability and Its Applications, 14, 153-158.

ESSEEN, C.G. (1945) "Fourier Analysis of Distribution Functions," Acta Mathematica, 77, 1-125.

FARAWAY, J.J. AND M. JHUN (1990) "Bootstrap Choice of Bandwidth for Density Estimation," Journal of the American Statistical Association, 85, 1119-1122.

FELLER, W. (1971) An Introduction to Probability Theory and Its Application: John Wiley & Sons.

FEUERVERGER, A. AND R.A. MUREIKA (1977) "The Empirical Characteristic

Function and Its Applications," Annals of Statistics, 5, 1, 88-97.

GABLER, S. (1981), F. LAISNEY, AND M. LECHNER (1993) "Seminonparametric Estimation of Binary-Choice Models with an Application to Labour-Force Participation," Journal of the American Statistical Association, 11, 61-80.

GALLANT, A.R. (1981) "On the Bias in Flexible Functional Forms and an Essentially Unbiased form: the Fourier Flexible Form," Journal of Econometrics, 15, 211-245.

GALLANT, A.R. AND D.W.NYCHKA (1987) "Semi-Nonparametric Maximum Likelihood Estimation," Econometrica, 55, 363-390.

GERFIN, M. (1996) "Parametric and Semi-parametric Estimation of the Binary Response Model of labour Market Participation," Journal of Applied Econometrics, 11, 321-339.

GLESJER, H. (1969) "A New Test for Heteroskedasticity," Journal of the American Statistical Association," 64, 316-323.

GNEDENKO, B.V. AND A.N. KOLMOGOROV (1954) Limit Distributions for Sums of Independent Random Variables, Addison-Wesley.

GOLDFELD, I.J. AND R.A. QUANDT (1965) "Some Tests for Homosckedasticity," Journal of the American Statistical Association, 60, 539-559.

GÖTZE, F. (1987) "Approximations for Multivariate *U*-Statistics," Journal of Multivariate Analysis, 22, 212-229.

GRAMS, W.F. AND R.J. SERFLING (1973) "Convergence Rates for *U*-Statistics and Related Statistics," Annals of Statistics, 1, 153-160.

GREBLICKI, W. (1974) "Asymptotically Optimal Probabilistic Algorithms for Pattern Recognition and Identification," Seria Monografie, 3, Naukowe Instytue Cybernetyki.

GRONAU, R. (1973) "The Effects of Children on the Housewife's Value of Time," Journal of Political Economy, 81, S168-S199.

HALL, P. (1991) "Edgeworth Expansions for Nonparametric Density Estimators, with Applications," Statistics, 22, 215-232.

HALL, P. (1992) The Bootstrap and Edgeworth Expansions. Springer.

HALL, P. AND J.L.HOROWITZ (1990) "Bandwidth Selection in Semiparametric Estimation of Censored Linear Regression Models," Econometric Theory, 6, 123-150.

HALL, P. AND S.J. SHEATHER (1988) "On the Distribution of a Studentized Quantile," Journal of the Royal Statistical Society, B, 50, 3, 381-391.

HÄRDLE, W. (1990) Applied Nonparametric Regression. Cambridge University Press.

HÄRDLE, W., P.HALL, AND H. ICHIMURA (1993) "Optimal Smoothing in Single-Index Models," Annals of Statistics, 21, 157-178.

HÄRDLE, W., P.HALL, AND J.S. MARRON (1988) "How Far Are Automatically Chosen Regression Smoothing Parameters from Their Optimum?," Journal of American Statistical Association, 83, 86-95.

HÄRDLE, W., J. HART, J.S. MARRON, AND A.B. TSYBAKOV (1992) "Bandwidth Choice for Average Derivative Estimation," Journal of the American Statistical Association, 87, 218-226.

HÄRDLE, W., W.HILDENBRAND, AND M. JERISON (1991) "Empirical Evidence on the Law of Demand," Econometrica, 59, 6, 1525-1549.

HÄRDLE, W., AND O. LINTON (1994) "Applied Nonparametric Methods," The Handbook of Econometrics, Vol. IV, ed. by D.F. McFadden and R.F. Engle III. New York: North Holland.

HÄRDLE, W., J.S. MARRON, AND M.P.WAND (1990) "Bandwidth Choice for Density Derivatives," Journal of Royal Statistical Society, B, 52, 223-232.

HÄRDLE, W., AND T.M. STOKER (1989) "Investigating Smooth Multiple Regression by the Method of Average Derivatives," Journal of the American Statistical Association, 84, 986-995.

HÄRDLE, W., AND A.B. TSYBAKOV (1993) "How Sensitive Are Averaged Derivatives?," Journal of Econometrics, 58, 31-48.

HARVEY, A.C. (1976) "Estimating Regression Models with Multiplicative Heteroscedasticity," Econometrica, 44, 461-466.

HECKMAN, J. (1974) "Shadow Prices, Market, and Labour Supply," Econometrica, 42, 4, 679-694.

HECKMAN, J. (1976) "The Common Structure of of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, 5, 475-492.

HECKMAN, J. (1979) "Sample Selection Bias as a Specification Error," Econometrica, 47, 153-161.

HECKMAN, N.E. (1986) "Spline Smoothing in a Partly Linear Model," Journal of the Royal Statistical Society, B, 48, 2, 244-248.

HECKMAN, J., H. ICHIMURA, J.SMITH, AND P.TODD (1998) "Characterizing Selection Bias Using Experimental Data," Econometrica, 66, 5, 1017-1098.

HELMERS, R. (1985) "The Berry-Esseen Bound for Studentized *U*-Statistics," Canadian Journal of Statistics, 13, 79-82.

HELMERS, R. (1991) "On the Edgeworth Expansion and the Bootstrap Approximation for a Studentized *U*-Statistics," Annals of Statistics, 19, 470-484.

HELMERS, R. AND W.R. VAN ZWET (1982) "The Berry-Esseen Bound for *U*-Statistics," Statistical Decision Theory and Related Topics III, Vol.1.

HIDALGO, J. (1992) "Adaptive Estimation in Time Series Regression Models with Heteroskedasticity of Unknown Form," Econometric Theory, 8, 161-187.

HOEFFDING, W. (1948) "A Class of Statistics with Asymptotically Normal Distribution," Annals of Mathematical Statistics, 19, 293-325.

HOEFFDING, W. (1961) "The Strong Law of Large Numbers for U-statistics," Univ. of North Carolina Institute of Statistics Mimeo Series, No. 302.

HONORÈ, B.E. (1992) "Trimmed LAD And Least Squares Estimation of Truncated And Censored Regression Models with Fixed Effects," Econometrica, 60, 3, 533-565.

HONORÈ, B.E. AND J.L.POWELL (1994) "Pairwise Difference Estimators of Censored and Truncated Regression Models," Journal of Econometrics, 64, 241-278.

HOROWITZ, J.L. (1986) "A Distribution-Free Least Squares Estimator for Censored Linear Regression Models," Journal of Econometrics, 32, 59-84.

HOROWITZ, J.L. (1988a) "The Asymptotic Efficiency of Semiparametric Estimators for Censored Linear Regression Models," Empirical Economics, 13, 123-140.

HOROWITZ, J.L. (1988b) "Semiparametric M-Estimation of Censored Linear Regression Models," in Advance in Econometrics, Vol. 7.

HOROWITZ, J.L. (1992) "A Smoothed Maximum Score Estimator for the Binary Response Model," Econometrica, 60, 3, 505-531.

HOROWITZ, J.L. (1993) "Semiparametric Estimation of a Work-Trip Mode Choice

Model," Journal of Econometrics, 58, 49-70.

HOROWITZ, J.L. (1998) Semiparametric Methods in Econometrics : Springer.

HOROWITZ, J.L. AND W. HÄRDLE (1996) "Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates," Journal of the American Statistical Association, 91, 1632-1640.

HOROWITZ, J.L. AND G.R. NEUMANN (1987) "Semiparametric Estimation of Employment Duration Models," Econometric Review, 6, 5-40.

ICHIMURA, H. (1993) "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," Journal of Econometrics, 58, 71-120.

ICHIMURA, H., and L.F. LEE (1991) "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation," in Nonparametric and Semiparametric Methods in Econometrics and Statistics: Cambridge University Press.

IZENMAN, A.J. (1991) "Recent Developments in Nonparametric Density Estimation," Journal of the American Statistical Association," 86, 205-224.

KAPLAN, E.L. AND P. MEIER (1958) "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 53, 457-481.

KIM, J. AND D.POLLARD (1990) "Cube Root Asymptotics," Annals of Statistics, 18, 191-219.

KLEIN, R.W. AND R.H.SPADY (1993) "An Efficient Semiparametric Estimator for Binary Response Models," Econometrica, 61, 387-421.

KOSHEVNIK, Y.A. AND B.Y.LEVIT (1976) "On a Non-parametric Analogue of the Information Matrix," Theory of Probability and its Applications, 21, 738-753.

LEE A.J. (1990) U-Statistics: Dekker.

LEE L.F. (1992) "Semiparametric Nonlinear Least-Square Estimation of Truncated Regression Models," Econometric Theory, 8, 52-94.

LEE L.F. (1998) "Semiparametric Estimation of Simultaneous-equation Microeconometric Models with Index Restriction," Japanese Economic Review, 49, 4, 343-380.

LEE M.-J. (1996) Methods of Moments And Semiparametric Econometrics for Limited Dependent Variable Models: Springer.

LEWBEL A. (1997) "Semiparametric Estimation of Location and Other Discrete

Choice Moments," Econometric Theory, 13, 32-51.

LEWBEL A. (1998) "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," Econometrica, 66, 105-121.

LI, Q. AND T. STENGOS (1996) "Semiparametric Estimation of Partially Linear Panel Data Models," Journal of Econometrics, 71, 389-397.

LINTON, O. (1995a) "Second Order Approximation in the Partially Linear Regression Model," Econometrica, 63, 1079-1112.

LINTON, O. (1995b) "Estimation in Semiparametric Models: a Review," Advances in Econometrics and Quantitative Economics, ed. by G.S. Maddala et.al.

LINTON, O. (1996a)"Edgeworth Approximation for MINPIN Estimators in Semiparametric Regression Models," Econometric Theory, 12, 30-60.

LINTON, O. (1996b) "Second Order Approximation in a Linear Regression with Heteroskedasticity of Unknown Form," Econometric Reviews, 15(1), 1-32.

MACK, Y.P. AND B.W. SILVERMAN (1982) "Weak and Strong Uniform Consistency of Kernel Regression Estimates," Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete, 61, 405-415.

MADDALA, G.S. (1983) Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press.

MADDALA, G.S. AND F. NELSON (1974) "Maximum Likelihood Methods for Markets in Disequilibrium," Econometrica, 42, 1013-1030.

MAGDALINOS, M.A. (1992) "Stochastic Expansions and Asymptotic Approximations," Econometric Theory, 8, 343-367.

MANSKI, C.F. (1975) "The Maximum Score Estimation of the Stochastic Utility Model of Choice," Journal of Econometrics, 3, 205-228.

MANSKI, C.F. (1984) "Adaptive Estimation of Non-Linear Regression Models," Econometric Reviews, 3, 145-194.

MANSKI, C.F. (1985) "Semiparametric Analysis of Discrete Response," Journal of Econometrics, 27, 313-333.

MANSKI, C.F. (1987) "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," Econometrica, 55, 357-362.

MANSKI, C.F. (1991) "Nonparametric Estimation of Expectations in the Analysis of Discrete Choice under Uncertainty," in Nonparametric and Semiparametric

Methods in Econometrics and Statistics: Cambridge University Press.

MARIANO, R.S. (1973a) "Approximations to the Distribution Functions of the Ordinary Least-Squares and Two-Stage Least-Squares Estimators in the Case of Two Included Endogenous Variables," Econometrica, 41, 67-77.

MARIANO, R.S. (1973b) "Approximations to the Distribution Functions of Theil's K-Class Estimators," Econometrica, 41, 715-721.

MARRON, J.S. (1985) "An Asymptotically Efficient Solution to the Bandwidth Problem of Kernel Density Estimation," Annals of Statistics, 13, 1011-1023.

MARRON, J.S. (1987) "A Comparison of Cross-Validation Techniques in Density Estimation," Annals of Statistics, 15, 152-162.

MARRON, J.S. (1988) "Automatic Smoothing Parameter Selection: A Survey," Empirical Economics, 13, 187-208.

MATZKIN, R.L. (1991) "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," Econometrica, 59, 1315-1327.

MORIMUNE, K. (1978) "Improving the Limited Information Maximum Likelihood Estimator When the Disturbances are Small," Journal of the American Statistical Association, 73, 867-871.

MORIMUNE, K. (1981) "Asymptotic Expansions of the Distribution of an Improved Limited Information Maximum Likelihood Estimator," Journal of the American Statistical Association, 76, 476-478.

NAGAR, A.L. (1959) "The Bias and Moment Matrix of the General *k*-Class Estimators of the Parameters in Simultaneous Equations," Econometrica, 27, 575-595.

NADARAYA, E. (1964) "On Regression Estimators," Theory of Probability and its Applications, 9, 157-159.

NADARAYA, E. (1965) "On Nonparametric Estimation of Density Function and Regression," Theory of Probability and its Applications, 10, 186-190.

NADARAYA, E. (1970) "Remarks on Nonparametric Estimates of Density Functions and Regression Curves," Theory of Probability and its Applications, 15, 139-142.

NADARAYA, E. (1974) "On the Integral Mean Square Error of Some Nonparametric Estimates of the Probability Density," Theory of Probability and its Applications, 19, 133-141.

NEWEY, W.K. (1990a) "Efficient Instrumental Variables Estimation of Nonlinear Models," Econometrica, 58, 809-837.

NEWEY, W.K. (1990b) "Semiparametric Efficiency Bounds," Journal of Applied Econometrics, 5, 99-135.

NEWEY, W.K. (1991) "Efficient Estimation of Tobit Models under Conditional Symmetry," in Nonparametric and Semiparametric Methods in Econometrics and Statistics: Cambridge University Press.

NEWEY, W.K. (1994) "The Asymptotic Variance of Semiparametric Estimators," Econometrica, 62, 1349-1382.

NEWEY, W.K. AND L.J.POWELL (1993) "Efficiency Bounds for Some Semiparametric Selection Models," Journal of Econometrics, 58, 169-184.

NEWEY, W.K. AND T.M.STOKER (1993) "Efficiency of Weighted Averaged Derivative Estimators and Index Models," Econometrica, 61, 1199-1223.

NISHIYAMA, Y. AND P.M.ROBINSON(1998) "Studentization in Edgeworth Expansions for Estimates of Semiparametric Index Models," forthcoming in Nonlinear Econometric Models (Festschrift for Takeshi Amemiya) (C.Hsiao, K.Morimune and J.Powell eds.).

NISHIYAMA, Y. AND P.M.ROBINSON(2000) "Edgeworth Expansions for Semiparametric Averaged Derivatives," forthcoming in Econometrica.

PARK, B.U. AND J.S. MARRON (1994) "Asymptotically Best Bandwidth Selecters in Kernel Density Estimation," Statistics and Probability Letters, 19, 119-127.

PARZEN, E. (1962) "On Estimation of a Probability Density Function and Mode," Annals of Mathematical Statistics, 33, 1065-1076.

PFANZAGL, J. (1971) "The Berry-Esseen Bound for Minimum Contrast Estimates," Metrika, 17, 82-91.

PFANZAGL, J. (1990) Estimation in Semiparametric Models: Springer-Verlag.

PFANZAGL, J. AND W.WEFELMEYER (1982) Contributions to a General Asymptotic Statistical Theory: Springer-Verlag.

PHILLIPS, P.C.B. (1977) "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators," Econometrica, 45, 1517-1534.

PHILLIPS, P.C.B. (1978) "Edgeworth and Saddlepoint Approximations in the First Order Non-Circular Autoregression," Biometrika, 65, 71-78.

PHILLIPS, P.C.B. (1980) "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume," Review of Economic Studies, 47, 183-224.

PHILLIPS, P.C.B. (1983) "ERA's: A New to Small Sample Theory," Econometrica, 51, 1505-1527.

POWELL, J.L. (1984) "Least Absolute Deviation Estimation for the Censored Regression Model," Journal of Econometrics, 25, 303-325.

POWELL, J.L. (1986a) "Symmetrically Trimmed Least Squares Estimation for Tobit Models," Econometrica, 54, 6, 1435- 1460.

POWELL, J.L. (1986b) "Censored Regression Quantiles," Journal of Econometrics, 32, 143-155.

POWELL, J.L. (1994) "Estimation of Semiparametric Models," Handbook of Econometrics, Vol. IV, edited by R.F.Engle and D.L. McFadden, Elsevier.

POWELL, J.L. (1998) "Semiparametric Estimation of Censored Selection Models," Unpublished Manuscript.

POWELL, J.L., J.H. STOCK, AND T.M. STOKER (1989) "Semiparametric Estimation of Index Coefficients," Econometrica, 57, 1403-1430.

POWELL, J.L., AND T.M. STOKER (1996) "Optimal Bandwidth Choice for Density-Weighted Averages," Journal of Econometrics, 75, 291-316.

PRAKASA RAO, B.L.S. (1977) "Berry-Esseen Bounds for Density Estimators of Stationary Markov Processes," Bulletin of Mathematical Statistics, 17, 15-21.

PRAKASA RAO, B.L.S. (1983) Nonparametric Functional Estimation. Academic Press: Orland.

QUENOUILLE, M.H. (1949) "Approximate Tests of Correlation in Time-Series," Journal of Royal Statistical Society, Ser. B, 11, 68-84.

QUENOUILLE, M.H. (1956) "Notes on Bias in Estimation," Biometrika, 43, 353-360.

RANGA RAO, R. (1961) "On the Central Limit Theorem in R_k ," Bulletin of the American Mathematical Society, 67, 359-361.

RHODES, G.F. JR. AND T.B. FOMBOY (1988) Advances in Econometrics, Vol. 7, Nonparametric and Robust Inference: JAI Press.

RICE, J. (1986) "Convergence rates for Partial Splined Models," Statistics and

Probability Letters, 4, 203-208.

ROBINSON, P.M. (1982) "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables," Econometrica, 50, 1, 27-41.

ROBINSON, P.M. (1987) "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form," Econometrica, 56, 875-891.

ROBINSON, P.M. (1988a) "Semiparametric Econometrics: A Survey," Journal of Applied Econometrics, 3, 35-51.

ROBINSON, P.M. (1988b) "Root-N-Consistent Semiparametric Regression," Econometrica, 56, 931-954.

ROBINSON, P.M. (1989) "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," Review of Economic Studies, 56, 511-534.

ROBINSON, P.M. (1991a) "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," Econometrica, 59, 755-786.

ROBINSON, P.M. (1991b) "Automatic Frequency Domain Inference on Semiparametric and Nonparametric Models," Econometrica, 59, 1329-1363.

ROBINSON, P.M. (1995a) "The Normal Approximation for Semiparametric Averaged Derivatives," Econometrica, 63, 667-680.

ROBINSON, P.M. (1995b) "The Approximate Distribution of Nonparametric Regression Estimates," Statistics and Probability Letters, 23, 193-201.

ROSENBLATT, M. (1956) "Remarks on Some Nonparametric Estimates of a Density Function," Annals of Mathematical Statistics, 27, 832-837.

ROTHENBERG, T.J. (1984a) "Approximate Normality of Generalized Least Squares Estimates," Econometrica, 52, 811-825.

ROTHENBERG, T.J. (1984b) "Approximating the Distributions of Econometric Estimators and Test Statistics," Handbook of Econometrics, Vol. II, Edited by Z. Griliches and M.D. Intriligator, Elsevier.

RUTEMILLER H.C. AND D.A.BOWERS (1968) "Estimation in a Heteroskedastic Regression Model," Journal of the American Statistical Association, 63, 552-557.

RUUD, P.A. (1983) "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinominal Discrete Choice Models," Econometrica, 51, 1, 225-228.

RUUD, P.A. (1986) "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution," Journal of Econometrics, 32, 157-187.

SAMAROV, A.M. (1993) "Exploring Regression Structure Using Nonparametric Functional Estimation," Journal of the American Statistical Association, 88, 836-847.

SARGAN, J.D. (1974) "The Validity of Nagar's Expansion for the Moments of Econometrics Estimators," Econometrica, 42, 169-176.

SARGAN, J.D. (1975a) "Asymptotic Theory and Large Models," International Economic Review, 16, 75-91.

SARGAN, J.D. (1975b) "Gram-Charlier Approximations Applied to *t* Ratios of *k*-Class Estimators," Econometrica, 43, 327-346.

SARGAN, J.D. (1976) "Econometric Estimators and the Edgeworth Approximation," Econometrica, 44, 421-448.

SARGAN, J.D. (1980) "Some Approximations to the Distributions of Econometric Criteria Which are Asymptotically Distributed as Chi-Squared," Econometrica, 48, 1108-1138.

SARGAN, J.D. AND W.M. MIKHAIL (1971) "A General Approximation to the Distribution of Instrumental Variables Estimates," Econometrica, 39, 131-169.

SARGAN, J.D. AND S.E. SATCHELL (1986) "A Theorem of Validity for Edgeworth Expansions," Econometrica, 54, 189-213.

SCHUSTER E.F. (1969) "Estimation of a Probability Density Function and Its Derivatives," Annals of Mathematical Statistics, 27, 832-837.

SCHUSTER E.F.(1972) "Joint Asymptotic Distribution of the Estimated Regression Function at a Finite Number of Distinct Points," Annals of Mathematical Statistics, 43, 84-88.

SCOTT, D.W. (1992) Multivariate Density Estimation. Wiley.

SERFLING, R.J. (1980) Approximation Theorems of Mathematical Statistics. John Wiley and Sons: New York.

SETHURAMAN,J. AND M. SIBUYA (1961) "Nonparametric Methods and Order Statistics," Technical Publ. in Statist., Indian Statist. Inst., Calcutta.

SHEATHER, S.J. (1983) "A Data-Based Algorithm for Choosing the Window Width When Estimating a Density at a Point," Computational Statistics and Data Analysis, 1, 229-238.

SHEATHER,S.J. (1986) "An Improved Data-Based Algorithm for Choosing the Window Width When Estimating a Density at a Point," Computational Statistics and Data Analysis, 4, 61-65.

SHEATHER,S.J. AND M.C. JONES (1991) "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," Journal of the Royal Statistical Society, Ser.B, 53, 683-690.

SHIAU, J., WAHBA, G. AND D.R. JOHNSON (1986) "Partial Spline Models for the Inclusion of Tropopause and Frontal Boundary Information in Otherwise Smooth Two and Three dimensional Objective Analysis," Journal of Atmospheric and Ocean Technology, 3, 713-725.

SHICK, A. (1986) "On Asymptotically Efficient Estimation in Semiparametric Models," Annals of Statistics, 14, 1139-1151.

SILVERMAN, B. (1978) "Weak and Strong Consistency of the Kernel Estimate of a Density and its Derivatives," Annals of Statistics, 6, 177-184.

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

SINGH, K. (1981) "On the Asymptotic Accuracy of Efron's Bootstrap," Annals of Statistics, 9, 1187-1195.

SINGH, R.S. (1976) "Nonparametric Estimation of Mixed Partial Derivatives of a Multivariate Density," Journal of Multivariate Analysis, 6, 111-122.

SINGH, R.S. (1977) "Improvement on Some Known Nonparametric Uniformly Consistent Estimators of Derivatives of a Density," Annals of Statistics, 5, 394-399.

SINGH, R.S. (1979) "On Necessary and Sufficient Conditions for Uniformly Strong Consistency of Estimators of a Density and Its Derivatives," Journal of Multivariate Analysis, 9, 157-164.

SINGH, R.S. AND D.S. TRACY (1977) "Strongly Consistent k-th Order Regression Curves and Rates of Convergence," Zeitschrift für Wahrscheinlichkeitstheorie Verwandte Gebiete, 40,339-348.

SPECKMAN, P. (1988) "Kernel Smoothing in Partial Linear Models," Journal of the Royal Statistical Society, B, 50, 3, 413-436.

STEIN, C. (1956) "Efficient Nonparametric Testing and Estimation," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol.1, University of California at Berkeley.

STERN, S. (1996) "Semiparametric Estimates of the Supply and Demand Effects of

Disability on Labor Force Participation," Journal of Econometrics, 71, 49-70.

STOKER, T.M. (1986) "Consistent Estimation of Scaled Coefficients," Econometrica, 54, 1461-81.

STOKER, T.M. (1991) "Equivalence of Direct, Indirect and Slope Estimators of Averaged Derivatives," in Nonparametric and Semiparametric Methods in Econometrics and Statistics: Cambridge University Press.

STOKER, T.M. (1993) "Smoothing Bias in Density Derivative Estimation," Journal of the American Statistical Association, 88, 855-863.

STOUT, W.F. (1974) Almost Sure Convergence. Academic Press

TANIGUCHI, M. (1983) "On the Second Order Asymptotic Efficiency of Estimators of Gaussian ARMA Processes," Annals of Statistics, 11, 157-169.

TANIGUCHI, M. (1991) Higher Order Asymptotic Theory for Time Series Analysis. Lecture Notes in Statistics 68: Springer-Verlag.

TAPIA, R.A. AND J.R. THOMPSON (1978) Nonparametric Probability Density Estimation. Johns Hopkins University Press.

THOMPSON, T.S. (1991) "Equivalence of Direct, Indirect, and Slope Estimators of Averaged Derivatives: A Comment," in Nonparametric and Semiparametric Methods in Econometrics and Statistics: Cambridge University Press.

TOBIN, J. (1958) "Estimation of Relationships for Limited Dependent Variables," Econometrica, 26, 24-36.

TSUI, K.-L, N.P. JEWELL AND C.F.J. WU (1988) "A Nonparametric Approach to the Truncated Regression Problem," Journal of the American Statistical Association, 83, 785-792.

TUKEY, J.W. (1958) "Bias and Confidence in Not-Quite Large Samples," Annals of Mathematical Statistics, 29, 614.

VON BAHR, B. (1967) "On the Central Limit Theorem in R_k ," Ark. Mat., 7, 61-69.

VON BAHR, B. AND C.G. ESSEEN (1965) "Inequalities for the rth Absolute Moment of a Sum of Random Variables, $1 \leq r \leq 2$," Annals of Mathematical Statistics, 36, 299-303.

WAHBA, G. (1984) "Partial Spline Models for the Semiparametric Estimation of Functions of Several Variables," in Statistical Analysis of Time Series, Proceedings of the Japan U.S. Joint Seminar, Tokyo, 319-329.

WAHBA, G. (1986) "Partial and Interaction Splines for the Semiparametric Estimation of Functions of Several Variables," in Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface, 75-80.

WAND, M.P. AND M.C. JONES (1995) Kernel Smoothing: Chapman Hall.

WATSON, G.S. (1964) "Smooth Regression Analysis," Sankhya, Ser. A, 26, 359-372.

WOODROOFE, M. (1970) "On Choosing a Delta-Sequence," Annals of Mathematical Statistics, 41, 1665-1671.