This is to certify that the

dissertation entitled

Minimum Distance Regression
and
Autoregressive Model Fitting

presented by

Pingping Ni

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

H. Koul

Major professor
Hira L. Koul

Date May 20, 2002

MINIMUM DISTANCE REGRESSION AND AUTOREGRESSIVE MODEL FITTING

By

Pingping Ni

A DISSERTATION

Submitted to
Michigan State University
in partial fufillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2002

ABSTRACT

MINIMUM DISTANCE REGRESSION AND AUTOREGRESSIVE MODEL FITTING

By

Pingping Ni

This work proposes a class of tests for fitting a parametric regression model to a regression function when the underlying design variables are random and the model is possibly hetroscedastic. These tests are based on certain minimized L_2 distances between a nonparametric regression function estimator and the parametric model being fitted. The work obtains the asymptotic distribution of the proposed statistic under the null hypthesis. It also derives the asymptotic distribution of the corresponding minimum distance estimator. A class of tests based on a slightly different L_2 distance for fitting a parametric autoregressive model to a autoregressive function is also proposed in this thesis. The asymptotic properties of underlying parameter estimator and corresponding minimized distanced is derived.

ACKNOWLEDGMENTS

I would like to thank my advisor Professor Hira L. Koul for his guidance and many helpful discussions on the subject of this thesis. He was always available when I had doubts or questions. His general thinking of statistical problem and ways to solve the problem will help my future research and working. I would also like to thank all the other committee members, Professors Connie Page, Habib Salehi, and Lijian Yang, for serving on my guidance committee.

Many thanks to Professor Connie Page for her advice when I was at the consulting service, to Professors Vincent Melfi, Habib Salehi, and James Stapleton, and to Cathy Sparks for their help on my simulation study. Finally I would like to thank the department of Statistics and Probabilities for offering me graduate assistantships so that I could come to the states to complete my graduate studies at the Michigan State University.

This research was partly supported by the NSF Grant DMS 0071619, under the P.I.: Professor Hira Koul.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

This thesis is concerned with the classical problem of using a set of variables, say d-dimensional variable X, to explain the response Y, a 1- dimensional real variable. As in the practice this is often done in terms of the conditional mean function of Y, given X, known as the the regression function, and defined as

$$\mu(x) = E(Y|X = x), \quad x \in \mathbb{R}^d,$$

assuming, of course, $E|Y| < \infty$. In the context of time series where X may be the vector of the previous d lagged variables, μ is called the autoregressive function.

To be specific, let $\{(X_i, Y_i) : i = 1, ..., n,\}$ be observable random variables, where (X_i, Y_i) has the same distributions as (X, Y), for all $1 \leq i \leq n$. They are said to obey a regression model with regression function μ if in addition $\{(X_i, Y_i) : i = 1, ..., n.\}$ are independent and identically distributed (i.i.d). The data is said to have come from an autoregressive model of order $d = 1$ with autoregressive function μ, if in addition, X_{n+1} is also observable and $Y_i = X_{i+1}$, $1 \leq i \leq n$.

Let $\Theta \subset \mathbb{R}^m$, and $\{m_\theta(\cdot) : \theta \in \Theta\}$, be a given set of parametric models. The statistical problem addressed in this thesis is that of model checking, i.e., to test the goodness-of-fit hypothesis

$$(1.0.1) \qquad H_0 : \mu(x) = m_{\theta_0}(x), \quad \text{for some } \theta_0 \in \Theta, \text{ and for all } x \in \mathcal{I} \ vs.$$

$$H_1 : H_0 \text{ is not true,}$$

based on the given data, where $\mathcal{I}$ is a compact subset of $\mathbb{R}^d$.

Several researchers have used nonparametric techniques on model checking in regression and autoregressive setting since the late 1980's. For instance, Eubank and Spiegelman (1990), Eubank and Hart (1992, 1993), Härdle and Mammen (1993), Stute (1996), and Stute, Thies, and Zhu (1998) address this problem in regression setting, while An and Cheng (1991), Vidar, Yao, and Tjøstheim (1997), and Koul and Stute (1999) in the autoregressive setting. In the regression context, some of these works focus on fixed design rather than random and under some restrictive assumptions on the error distribution. The proposed tests in these papers, except Stute (1996), Stute et al. (1998), and Koul and Stute (1995), are based on some nonparametric estimator of the regression function while the tests in the latter papers are based on a certain partial sum empirical process of the residuals.

Here we shall briefly summarize the contents of some of these papers. Eubank and Spiegelman (1990) consider the sequence of models where $d = 1$, at stage n,

$X_i = x_{in}$ with $0 \le x_{1n} < \cdots < x_{nn} \le 1$, known nonrandom, and where

$$\mu(x_{in}) = \beta_0 + \beta_1 x_{in} + f(x_{in}), \qquad 1 \le i \le n,$$

and f is a smooth unknown function. Moreover, here the errors $Y_i - \mu(x_{in})$ are assumed to be i.i.d. $N(0, \tau^2)$ with τ^2 unknown. It is also assumed that x_{in} are generated by a continuous positive density w on $[0,1]$ through the relation $\int_0^{x_{jn}} w(x)dx = (2j-1)/2n$. The problem addressed in this paper is to test the hypothesis $f = 0$ versus the alternative that $f \in L_2(w)/\{1, x\}$, f is absolutely continuous and its a.e. derivative f' is absolutely continuous and square integrable. Here the space $L_2(w)/\{1, x\}$ consists of all functions in $L_2(w)$ orthogonal to 1 and the identity function.

The paper proposes two tests. For one, they assume that $f = \boldsymbol{T}_{np}\boldsymbol{\alpha}$, $\boldsymbol{\alpha} \in \mathbb{R}^p$, where $\boldsymbol{T}_{np}$ is a vector of known functions orthogonal to 1 and the identity function. Then the test is based on the the least square estimators of β_0, β_1 and $\boldsymbol{\alpha}$. The other test is based on the spline estimation of f and the least squares estimators of β_0, β_1. They prove the asymptotic normality of their proposed statistics under their null hypothesis. We note that the problem addressed in this paper may be thought to be equivalent to fitting a simple linear regression model, i.e., to test H_0 of (1.0.1), with $m = 2$, $m_\theta(x) = (1, x)\theta$, against a nonparametric class of alternatives.

Härdle and Mammen (1993) consider the problem of testing H_0 based on the model

$$Y_i = \mu(X_i) + \varepsilon_i, \tag{1.0.2}$$

where ε_i's are allowed to be heteroscedastic with $E(\varepsilon_i|X_i) = 0$ and $E(\varepsilon_i^2|X_i = x) = \sigma^2(x)$. They propose a class of tests based on

$$
\begin{aligned}
(1.0.3)\ M_{hh}(\theta) \ &:= \ \int_{\mathcal{I}} \left[n^{-1} \sum_{i=1}^{n} K_h(x - X_i) \left(Y_i - m_\theta(X_i) \right) \right]^2 \{\hat{f}_h(x)\}^{-2} dG(x), \\
\hat{f}_h(x) \ &:= \ n^{-1} \sum_{i=1}^{n} K_h(x - X_i), \ x \in \mathbb{R}^d, \quad K_h(u) = \frac{1}{h^d} K(\frac{u}{h}),
\end{aligned}
$$

where K is a kernel density function on $[-1, 1]^d$ and G is a σ-finite measure on $\mathbb{R}^d$. Their test is based on the statistic $\tilde{T}_n := nh^{d/2} M_{hh}(\tilde{\theta})$, where the estimator $\tilde{\theta}$ and the null model are assumed to satisfy the condition

$$
(1.0.4) \qquad m_{\tilde{\theta}}(x) - m_{\theta_0}(x) = (1/n) \sum_{i=1}^{n} < \eta(x), \gamma(X_i) > \varepsilon_i + o_p((n \log n)^{-1/2}),
$$

uniformly in x. Here η and γ are bounded functions taking values in $\mathbb{R}^k$ for some k. It is pointed out in the paper that this assumption holds for linear models and the weighted least squares estimators in nonlinear models if $\mu(\cdot)$ is "smooth" with $\eta(x) = (\partial/\partial\theta) m_\theta(\cdot)$ at $\theta = \theta_0$.

Apart from the usual assumptions such as the kernel K is a symmetric, twice continuously differentiable with compact support, X lies in a compact set with probability 1 and the density f of X is bounded away from zero and infinity, they also assumed that $h_n = cn^{-1/(d+4)}$ for some known constant $c > 0$, the regression function μ and the density f are twice continuously differentiable, and $Eexp(t\varepsilon_i)$ is uniformly bounded in i for $|t|$ small enough.

Under some additional assumptions, they concluded that the asymptotic null distribution of $nh^{d/2}(M_{hh}(\tilde{\theta}) - \tilde{C}_n)$ is $N(0, \tilde{V})$, where $\tilde{C}_n$ depends on the $\mu = m_{\theta_0}$, the second derivative of K and $h^{-d/2}$, and where

$$\tilde{V} = 2\int \frac{\sigma^4(x)g^2(x)}{f^2(x)}dx \int \left(K^{(2)}(t)\right)^2 dt,$$

and g is the Lebesgue density of G.

The choice of bandwidth $h_n = cn^{-1/(d+4)}$ is asymptotically optimal for the class of twice continuously differentiable regression functions. It is also crucial in getting the rates of uniform consistency of nonparametric estimators of μ and f, which in turn play a crucial part in the proofs of this paper.

The paper gives details of the proof for one dimensional case only, i.e, for the case $d = 1$. But it is not clear how their proof can be extended to the case $d > 1$, without a concern for bandwidth selection.

These authors also did Monte Carlo simulations on both distribution of the test statistic and its asymptotic distribution. Their studies show that the simulation of the null distribution of the test statistic has a non-negligible large departure from the limiting distribution in its mean, variance, and shape. It is also proved in the paper that the naive bootstrap does not work for degenerate U statistics. So they suggested to use wild bootstrap to calculate critical values.

Stute, Thies and Zhu (1998) also considered the problem of testing H_0 of (1.0.1) for model (1.0.2) with $d = 1$. They constructed a class of test statistics by first splitting the whole sample into two parts, 1 to n_1 and n_1 to n with $n_1 \longrightarrow \infty$ and $n - n_1 \longrightarrow \infty$. The test statistic is based on the cusum process of the residuals of the second half. Let F_{n_1} be the empirical distribution function of $X_{n_1+1}, ..., X_n$, θ_{n_1}

be a $\sqrt{n-n_1}$-consistent estimator of the true parameter under the null hypothesis based on (X_i, Y_i), $n_1+1 \le i \le n$, and let

$$\tilde{R}_n^1(x) = (n-n_1)^{-1/2} \sum_{i=n_1+1}^{n} 1_{\{X_i<x\}} \sigma_{n_1}^{-1}(X_i)[Y_i - m_{\theta_{n_1}}(X_i)].$$

Define the transformation T_n by

$$\begin{aligned}(T_n f)(x) \\ = \; & f(x) - \int_{-\infty}^{x} \sigma_{n_1}^{-1}(y)\dot{m}_{\theta_{n_1}}^T(y)A_{n_1}^{-1}(y) \times \left[\int_y^\infty \sigma_{n_1}^{-1}(z)\dot{m}_{\theta_{n_1}}(z)f(dz)\right] F_{n_1}(dy).\end{aligned}$$

Here

$$A_{n_1}(y) = \int_y^\infty \dot{m}_{\theta_{n_1}}(u)\dot{m}_{\theta_{n_1}}^T(u)\sigma_{n_1}^{-2}(u)F_{n_1}(du),$$

$\sigma_{n_1}^2$ is a consistent estimator of σ^2 based on the first half sample, and $\dot{m}_\theta = \frac{\partial}{\partial\theta}m_\theta$.

Under the assumption that $\int_{x_0}^\infty \dot{m}_{\theta_0}(u)\dot{m}_{\theta_0}^T(u)\sigma^{-2}(u)F(du)$ is positive definite for some $x_0 < \infty$, and under some additional smoothness assumptions on the null model, they proved that under H_0, $T_n\tilde{R}_n^1 \longrightarrow B \circ F$ in distribution on $D[-\infty, x_0]$, where B is a standard Brownian motion and F is the distribution function of X. They then propose the test statistic $\sigma_{n1}^{-2}F_n^{-2}(x_0)\int_{-\infty}^{x_0}[T_n\tilde{R}_n^1(x)]^2F_n(dx)$. It is also proved in this paper that their test statistic converges in distribution to $\int_0^1 B^2(u)du$ under their null hypothesis.

An and Cheng (1991) considered a problem of testing linearity of an autoregressive function. They proposed a Kolmogorov-Simirnov type of test statistic based on a process similar to $\tilde{R}_n$ with

$$\hat{e}_i = (X_i - \bar{X}) - \hat{\rho}(X_{i-1} - \bar{X}),$$

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad \text{and} \qquad \hat{\rho} = \frac{\sum_{k=1}^{n}(X_k - \bar{X})(X_{k-1} - \bar{X})}{\sum_{k=1}^{n}(X_k - \bar{X})^2}.$$

The test statistic is defined to be

$$\sup_{-\infty<t<\infty}\left|\frac{m^{-1/2}}{\hat{\sigma}}\sum_{k=2}^{m} \hat{e}_k I_{(X_{k-1}<t)}\right|,$$

where $m = m(n)$ is a subsequence of n satisfying $m \longrightarrow \infty$ and $m(lnlnn)/n \longrightarrow 0$. It is proved in the paper that this test statistic converges in distribution to $\sup_{0<t<1}|B(t)|$ under the null hypothesis, where B stands for standard Brownian motion.

Koul and Stute (1999) proposed a class of tests for testing the goodness-of-fit of an autoregressive model also based on an analogue of $\tilde{R}_n$. The test statistic is defined as follows,

$$\sup_{x \le x_0} \frac{|T_n V_n(x)|}{\sigma_n \sqrt{G_n(x_0)}},$$

where, for an $x \in \mathbb{R}$,

$$\begin{aligned} T_n V_n(x) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[I(X_{i-1} \le x) - \frac{1}{n}\sum_{j=1}^{n} \dot{m}_{\theta_n}(X_{j-1}) A_n^{-1}(X_{j-1}) \dot{m}_{\theta_n}(X_{i-1}) \\ &\qquad \times I(X_{j-1} \le X_{i-1} \wedge x)\Big](X_i - m_{\theta_n}(X_{i-1})), \\ G_n(x) &= \frac{1}{n}\sum_{i=1}^{n} I(X_{i-1} \le x), \qquad A_n(x) = \int \dot{m}_{\theta_n}(y)\dot{m}_{\theta_n}^T(y) I(y \ge x) G_n(dy). \end{aligned}$$

Here θ_n is the least square estimator of θ_0, σ_n^2 is a consistent estimator of the variance of the error. It is also proved that under H_0 and when $\sigma^2(x) = \sigma^2$, the test statistic converges to $\sup_{0<t<1}|B(t)|$, in distribution.

Vidar, Yao, and Tjϕstheim (1997) considered a problem of fitting a linear autoregressive function by using local polynomial approximation. They pointed out that one can construct new tests of linearity by exploiting that the first order derivative is a constant, and the second order derivative is zero for a linear model. From the estimation point of view, the local polynomial method does overcome some draw backs of kernel type nonparametric estimate provided that the regression/autoregressive function is "smooth", for example the existence of higher order derivatives. If continuity is the only smoothness condition that is put on the regression/autoregressive function, then either a kernel type estimates or a local polynomial estimate of the regression/autoregressive function yield exactly the same estimates. That means that some tests proposed in this paper can not be extended easily to nonlinear case without relatively strong smoothness conditions on the regression/autoregressive function.

Our work uses the minimum distance ideas as developed by Wolfowitz (1952, 1954, 1957), to propose tests of fit for the problem. The inference procedures based on various L_2-distances have proved to be successful in producing tests for fitting a distribution and/or a density function, and in producing asymptotically efficient and robust estimators of the underlying parameters in the fitted model, as is evidenced in the works of Beran (1977, 1978), Donoho and Liu (1988a, 1988b), and Koul (1985), among others.

Beran (1977) focuses on fitting a parametric family $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ of densities to the common density in the one sample setup. The question raised in the paper is how to estimate θ in order to investigate the fit of the model to the data. This paper introduces a new efficient parametric estimator based on the minimum Hellinger distance. The Hellinger distance is defined to be the L_2 norm of the difference of the square roots of two nonnegative densities. The parametric estimator θ_n is the $\theta \in \Theta$ that minimizes the Hellinger distance between f_θ and $\hat{f}_h$. It is proved that under some conditions the estimator θ_n is stable under small perturbations, and $\sqrt{n}(\theta_n - \theta_0)$ converges in distribution to a normal random variable with mean zero and variance $4^{-1}[\int \dot{s}_{\theta_0}(x)\dot{s}_{\theta_0}^T(x)dx]^{-1}$, where θ_0 is the true parameter, $\dot{s}_{\theta_0}$ is $\frac{\partial}{\partial\theta}s_\theta$ at $\theta = \theta_0$, and $s_\theta = f_\theta^{1/2}$. The test statistic for testing the null hypothesis that f is a member of $\mathcal{F}$, against the alternative hypothesis that f is not a member of $\mathcal{F}$, is the corresponding minimum Hellinger distance. It is also proved in the paper that under some conditions the suitably standardized minimum Hellinger distance converges in distribution to $N(0,1)$ under the null hypothesis.

In the context of density fitting problem in the one sample set up, Beran (1977, 1978) showed that the inference procedures based on the Hellinger distance have desirable properties. In the regression model fitting context, this motivates one to consider the square distance

$$M_{hh}^*(\theta) = \int_I (\hat{\mu}_{hh}(x) - m_\theta(x))^2 dG(x), \qquad \theta \in \mathbb{R}^q, \tag{1.0.5}$$

and the corresponding minimum distance estimator $\alpha_n^* = \text{argmin}_{\theta\in\Theta} M_{hh}^*(\theta)$, where $\hat{\mu}_{hh}(x)$ is a nonparametric estimator of the regression function $\mu(x)$ based on the window width $h = h_n$:

$$\hat{\mu}_{hh}(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i}{\hat{f}_h(x)}.$$

But because the integrand inside the square of M^*_{hh} is not centered, and because of the non-negligible asymptotic bias in the nonparametric estimator $\hat{\mu}_{hh}$, the goodness-of-fit statistic $M^*_{hh}(\alpha^*_n)$ does not have a desirable asymptotic null distribution. Moreover, the estimator α^*_n, though consistent, is not asymptotically normal. In fact it can be shown that generally the sequence $(nh^d)^{1/2}\|\alpha^*_n - \theta_0\|$ may not even be tight. For example, see Remark 2.4.3 at the end of Chapter 2. To overcome this difficulty, one may think of using M_{hh} defined in (1.0.3) and let $\hat{\alpha}_n = \text{argmin}_{\theta\in\Theta} M_{hh}(\theta)$.

Now, under the null hypothesis H_0, the i^{th} summand inside the square integrand of $M_{hh}(\theta_0)$ is now conditionally centered, given the i^{th} design variable, $1 \le i \le n$. But the asymptotic bias in $n^{1/2}(\hat{\alpha}_n - \theta_0)$ and $M_{hh}(\hat{\alpha}_n)$ caused by the nonparametric estimator $\hat{f}_h$ of f in the denominator of $\hat{\mu}_{hh}$ still exists. It turns out that this difficulty can be overcome if we use optimal window width, different from h, and possibly a different kernel, to estimate f. This leads us to consider the following modification of the above distance and estimator. Define

$$\begin{aligned}
(1.0.6)\; \hat{f}_{w_n}(x) &= n^{-1}\sum_{i=1}^{n} K^*_w(x - X_i),\ x \in \mathbb{R}^d, \quad w_n \sim (\log n/n)^{\frac{1}{d+4}}, \\
M_{hw}(\theta) &= \int_{\mathcal{I}} \left[n^{-1}\sum_{i=1}^{n} K_h(x - X_i)\left(Y_i - m_\theta(X_i)\right)\right]^2 \{\hat{f}_{w_n}(x)\}^{-2} dG(x),
\end{aligned}$$

where K^* is a density kernel function, possibly different from K, satisfying a Lipschitz condition. The proposed minimum distance estimators of θ are

$$\hat{\theta}_n := \operatorname{argmin}_{\theta\in\Theta} M_{hw}(\theta),$$

and the proposed tests of H_0, one for each G, will be based on

$$\inf_{\theta\in\Theta} M_{hw}(\theta) = M_{hw}(\hat{\theta}_n).$$

We also consider the following square distance and estimator:

$$\begin{aligned} M^*_{hw}(\theta) &= \int_{\mathcal{I}} (\hat{\mu}_{hw}(x) - m_\theta(x))^2 dG(x), \qquad \theta \in \mathbb{R}^q, \\ \theta^*_n &= \operatorname{argmin}_{\theta\in\Theta} M^*_{hw}(\theta), \end{aligned}$$

where

$$\hat{\mu}_{hw}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i) Y_i \Big/ \hat{f}_{w_n}(x), \quad x \in \mathbb{R}^p.$$

For the sake of simplicity, we write h for h_n and w for w_n.

This thesis proves the consistency of θ^*_n, $\hat{\theta}_n$, and the asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta_0)$. The asymptotic null distribution of the statistic

$$T_n := nh^{d/2} \left(M_{hw}(\hat{\theta}_n) - \hat{C}_n \right) / \hat{\Gamma}_n^{1/2}$$

is shown to be standard normal. This result is similar in nature as the corollary to Theorem 8 of Beran (1977, p459). A test of H_0 can be thus based on T_n. Here, $\hat{C}_n$ is an $nh^{d/2}$-consistent approximation of an asymptotic centering sequence $\tilde{C}_n$ and

$\hat{\Gamma}_n$ is a consistent estimator of the asymptotic variance Γ,

$$\begin{aligned}
\tilde{C}_n &:= n^{-2}\sum_{i=1}^{n}\int_{\mathcal{I}} K_h^2(x-X_i)\,\varepsilon_i^2\,\{f(x)\}^{-2}dG(x),\\
\hat{C}_n &= n^{-2}\sum_{i=1}^{n}\int_{\mathcal{I}} K_h^2(x-X_i)\,\hat{\varepsilon}_i^2\,\{\hat{f}_w(x)\}^{-2}dG(x),\ \ \hat{\varepsilon}_i = Y_i - m_{\hat{\theta}_n}(X_i),\ 1\le i\le n,\\
\Gamma &:= 2\int_{\mathcal{I}}\sigma^4(x)\,\frac{g^2(x)}{f^2(x)}dx\int\Big(\int K(u)K(v+u)du\Big)^2 dv,\\
\hat{\Gamma}_n &= h^d n^{-2}\sum_{i\neq j}\Big(\int_{\mathcal{I}} K_h(x-X_i)K_h(x-X_j)\,\hat{\varepsilon}_i\hat{\varepsilon}_j\,\{\hat{f}_w(x)\}^{-2}dG(x)\Big)^2,
\end{aligned}$$

where $\sigma^2(x) := E\Big\{(Y-\mu(x))^2\Big|X=x\Big\}$, $x\in\mathbb{R}^d$.

In autoregressive setup, where autoregressive function is defined to be

$$\mu(x) = E(X_n|X_{n-1}=x),\ n\in\mathbb{Z}, \tag{1.0.7}$$

and $\mathbb{Z}$ stands for the set of integers, we propose a class of test statistics for testing H_0 of (1.0.1) based on a slightly different L_2-distance $M_h(\theta)$ defined as

$$M_h(\theta) := \int_{\mathcal{I}}\Big(\frac{1}{n}\sum_{i=1}^{n}K_h(x-X_{i-1})(X_i - m_\theta(X_{i-1}))\Big)^2 dG(x). \tag{1.0.8}$$

The underlying parameter estimator is defined as

$$\theta_n := \operatorname{argmin}_{\theta\in\Theta} M_h(\theta). \tag{1.0.9}$$

It is proved in this thesis that when $d=1$ and ε_i's are i.i.d with $\sigma^2(x)\equiv\sigma^2$, under some conditions, $\sqrt{n}(\theta_n-\theta_0)$ converges in distribution to a normal random variable with mean zero and covariance matrix $\Sigma_0^{-1}\boldsymbol{\eta}^2\Sigma_0^{-1}$ under H_0, where

$$\begin{aligned}
\boldsymbol{\eta}^2 :&= \sigma^2\int_{\mathcal{I}}\dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)g^2(x)f^3(x)dx,\quad \xi(x) := \dot{m}_{\theta_0}(x)f(x),\\
\Sigma_0 :&= \int_{\mathcal{I}}\xi(x)\xi^T(x)dG(x).
\end{aligned} \tag{1.0.10}$$

It is also proved that under some conditions, a suitably standardized minimized distance $M_h(\theta_n)$ converges in distribution to a standard normal random variable under the null hypothesis.

This thesis is organized as follows. Chapter 2 discusses the model fitting for regression function. Theorem 2.2.1 and Theorem 2.3.1 give the asymptotic properties of the underlying parameter estimate. Theorem 2.4.1 gives the asymptotic distribution of the minimized distance under the null hypothesis. A test statistic therefore can be constructed based on this theorem. Chapter 3 discusses a parallel results for autoregressive model fitting.

Chapter 4 shows a simulation results in the autoregressive setting. This simulation compares the level and power performance of a minimum distance test with that of Koul and Stute and An and Cheng (1991) tests for the sample sizes $50, 100, 200, 500$. The minimum distance test is seen to perform better at some of the chosen altenatives and for all chosen sample sizes. For additional details see Chapter 4.

In the sequel, all limits are taken as $n \longrightarrow \infty$, unless specified otherwise.

Chapter 2

Minimum Distance Regression Model Fitting

2.1 Introduction

This chapter discusses a minimum distance method for fitting a parametric model to the regression function, i.e. to test H_0 of (1.0.1) based on the random sample $\{(X_i, Y_i) : i = 1, ..., n\}$ from the distribution of (X, Y), for which (X_i, Y_i) satisfy (1.0.2), where $\mathcal{I}$ is a compact subset of $\mathbb{R}^d$. Moreover, assuming that the given parametric family of models holds, one is interested in finding the model in the given family that best fits the data.

In this chapter, we will construct a class of tests based on $M_{hw}(\hat{\theta}_n)$ in (1.0.6). In contrast to Härdle and Mammen (1993), our results do not require the null regression function to be twice continuously differentiable nor do the proofs in this chapter need the rate for uniform consistency of $\hat{\mu}_{hw}$ for μ. Moreover, we derive the asymptotic distributions of $n^{1/2}(\hat{\theta}_n - \theta_0)$ and T_n under H_0. This was made feasible by recognizing to use different window widths for the estimation of the numerator and denominator in the nonparameteric regression function estimator.

The rest of the chapter is organized as follows. Section 2 states various assumptions, and section 3 contains the consistency proofs. The claimed asymptotic normality of $\hat{\theta}_n$ and $M_{hw}(\hat{\theta}_n)$ are proved in sections 4 and 5, respectively. A simulation study is presented in section 6 to illustrate the asymptotics for the sample sizes $50, 100$, and 200. The results are presented in terms of densities of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \hat{C}_n)$. These graphs show that the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ resembles the asymptotic normal distribution quite well even for the sample of size 50. The distribution of $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \hat{C}_n)$ has a small negative bias compared with the asymptotic normal distribution for all three sample sizes. But the bias decreases as n increases.

2.2 Assumptions

Here we shall state the needed assumptions. About the errors, the underlying design and G we assume the following:

(e1) The random variables $\{(X_i, Y_i);\ X_i \in \mathbb{R}^d,\ Y_i \in \mathbb{R}, i = 1, \cdots, n\}$, are *i.i.d.* with the regression function $\mu(x) = E(Y|X = x)$ satisfying $\int \mu^2(x)dG(x) < \infty$, where G is a σ-finite measure on $\mathbb{R}^d$.

(e2) $E\big(Y-\mu(X)\big)^2 < \infty$ and the function $\sigma^2(x) := E\{\big(Y-\mu(x)\big)^2 | X = x\}$ is a.s. (G) continuous on $\mathcal{I}$.

(f) The design variable X has a uniformly continuous Lebesgue density f that is bounded from below on $\mathcal{I}$.

(g) G has a continuous Lebesgue density g.

About the kernel functions K, K^* we shall assume the following:

(k) The kernels K, K^* are positive symmetric density functions on $[-1,1]^d$ with finite variances and $\int |u|^r K^2(u)du + \int K^{*2}(u)du < \infty$, for $r = 0, 1, 2$. In addition, K^* satisfies a Lipschitz condition.

About the parametric family of functions to be fitted we need to assume the following:

(m1) For each θ, $m_\theta(x)$ is a.s continuous in x w.r.t integrating measure G.

(m2) The parametric family of models $m_\theta(x)$ is identifiable w.r.t θ. i.e., if $m_{\theta_1}(x) = m_{\theta_2}(x)$, for almost all x (G), then $\theta_1 = \theta_2$.

(m3) For some positive continuous function ℓ on $\mathcal{I}$ and for some $\beta > 0$,

$$|m_{\theta_2}(x) - m_{\theta_1}(x)| \leq ||\theta_2 - \theta_1||^\beta \, \ell(x), \quad \forall\, \theta_2,\, \theta_1 \in \Theta,\, x \in \mathcal{I}.$$

(m4) The model m_θ is differentiable in θ in a neighborhood of θ_0 with the vector of derivatives $\dot{m}_\theta$, such that for every $\epsilon > 0$, $k < \infty$,

$$\limsup_n P\Big(\sup_{1 \leq i \leq n,\, (nh^d)^{1/2}||\theta-\theta_0|| \leq k} \frac{|m_\theta(X_i) - m_{\theta_0}(X_i) - (\theta-\theta_0)^T \dot{m}_{\theta_0}(X_i)|}{||\theta - \theta_0||} > \epsilon \Big)$$

is 0.

(m5) For every $\epsilon > 0$, there is an $N_\epsilon < \infty$ such that for every $0 < k < \infty$,

$$P\left(\max_{1\le i\le n,\,(nh^d)^{1/2}\|\theta-\theta_0\|\le k} h^{-d/2}\|\dot{m}_\theta(X_i) - \dot{m}_{\theta_0}(X_i)\| \ge \epsilon\right) \le \epsilon, \qquad \forall\, n > N_\epsilon.$$

About the bandwidth h_n we shall make the following assumptions:

(h1) $h_n \to 0$ as $n \to \infty$.

(h2) $nh_n^{2d} \to \infty$ as $n \to \infty$.

(h3) $h \sim n^{-a}$, where $a < min(1/2d, 4/(d(d+4)))$.

Conditions (h1) and (h2) suffice for the consistency of $\hat{\theta}_n$, while (h3) is needed for the asymptotic normality of $\hat{\theta}_n$ and $M_{hw}(\hat{\theta}_n)$. Of course, (h3) implies (h1) and (h2).

It is well known that under (f), (k), (h1) and (h2), cf., Mack and Silverman (1982),

$$\text{(2.2.1)} \qquad \sup_{x\in\mathcal{I}}\left|\hat{f}_h(x) - f(x)\right| = o_p(1), \quad \sup_{x\in\mathcal{I}}\left|\hat{f}_w(x) - f(x)\right| = o_p(1),$$

$$\text{(2.2.2)} \qquad \sup_{x\in\mathcal{I}}\left|\frac{f(x)}{\hat{f}_w(x)} - 1\right| = o_p(1).$$

These conclusions are often used in the proofs below.

In the sequel, we write h for h_n, w for w_n; the true parameter θ_0 is assumed to be an inner point of Θ; and the integrals with respect to the G-measure are understood to be over the set $\mathcal{I}$. The inequality $(a+b)^2 \le 2(a^2+b^2)$, for any real numbers a, b, is often used without mention in the proofs below.

2.3 Consistency of θ_n^* and $\hat{\theta}_n$

This section proves the consistency of θ_n^* and $\hat{\theta}_n$. To state and prove these results we need some more notation. Let $L_2(G)$ denote a class of square integrable real valued functions on $\mathbb{R}^d$ with respect to G. Define

$$\rho(\nu_1, \nu_2) \;:=\; \int_{\mathcal{I}} (\nu_1(x) - \nu_2(x))^2 dG(x), \quad \nu_1,\, \nu_2 \in L_2(G),$$

and the map

$$T(\nu) = \operatorname{argmin}_{\theta \in \Theta} \rho(\nu, m_\theta), \quad \nu \in L_2(G).L_2(G).$$

In the sequel we shall often use the following notation

$$d\hat{\varphi}_h := \hat{f}_h^{-2} dG, \quad d\varphi = f^{-2} dG.$$

Moreover, for any integral $L := \int \gamma d\hat{\varphi}_h$, $\tilde{L} := \int \gamma d\varphi$. Thus, e.g., $\tilde{T}(\theta)$ stands for $T(\theta)$ with $\hat{\varphi}_w$ replaced by φ, i.e., with $\hat{f}_w$ replaced by f. We also need to define

$$\begin{aligned}
\mu_n(x,\theta) &:= n^{-1} \sum_{i=1}^n K_h(x - X_i)\, m_\theta(X_i), \\
\dot{\mu}_n(x,\theta) &:= n^{-1} \sum_{i=1}^n K_h(x - X_i) \dot{m}_\theta(X_i), \\
U_n(x,\theta) &:= n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i - \mu_n(x,\theta), \\
&= n^{-1} \sum_{i=1}^n K_h(x - X_i)(Y_i - m_\theta(X_i)), \qquad U_n(x) = U_n(x,\theta_0), \\
Z_n(x,\theta) &:= \mu_n(x,\theta) - \mu_n(x,\theta_0) \\
&= n^{-1} \sum_{i=1}^n K_h(x - X_i)[m_\theta(X_i) - m_{\theta_0}(X_i)], \quad \theta \in \mathbb{R}^q,
\end{aligned}$$

$$\begin{aligned}\bar{K}_n(x) &:= n^{-1}\sum_{i=1}^{n} K_h(x - X_i), \quad \bar{K}_n^*(x) := n^{-1}\sum_{i=1}^{n} K_w^*(x - X_i), \ x \in \mathbb{R}^d, \\ \Sigma_0 &:= \int \dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)dG(x).\end{aligned}$$

To begin with we state

Lemma 2.3.1 *Let m satisfy the conditions (m1), (m2), and (m3). Then the following hold.*

(a) $T(\nu)$ always exists, $\forall\, \nu \in L_2(G)$.

(b) If $T(\nu)$ is unique, then T is continuous at ν in the sense that for any sequence of $\{\nu_n\} \in L_2(G)$ converging to ν in $L_2(G)$, $T(\nu_n) \to T(\nu)$, i.e.,

$$\rho(\nu_n, \nu) \longrightarrow 0 \quad \textit{implies} \quad T(\nu_n) \longrightarrow T(\nu), \quad \textit{as } n \to \infty.$$

(c) $T(m_\theta(\cdot)) = \theta$, uniquely for $\forall\ \theta \in \Theta$.

Proof. The main ideas of the following proof are essentially as in Beran (1977).

Proof of Part (a). Because Θ is compact, it suffices to show that for every $\nu \in L_2(G)$, the map $\theta \mapsto \rho(\nu, m_\theta)$ is continuous. Accordingly, let θ_n be a sequence in Θ, converging to a $\theta \in \Theta$. Then, by the Cauchy-Schwarz inequalities, we obtain

$$|\rho(\nu, m_{\theta_n}) - \rho(\nu, m_\theta)| \le \rho(m_{\theta_n}, m_\theta) + 2\rho^{1/2}(\nu, m_\theta)\rho^{1/2}(m_{\theta_n}, m_\theta) \longrightarrow 0,$$

by (m3).

Proof of part (b). Let $\{\nu_n\}$, ν in $L_2(G)$ be such that

(2.3.1) $$\rho(\nu_n, \nu) \to 0.$$

Set $\vartheta = T(\nu)$, $\vartheta_n = T(\nu_n)$. Then, by the definition of T,

$$\rho(\nu_n, m_{\vartheta_n}) \leq \rho(\nu_n, m_\vartheta).$$

By subtracting and adding ν and expanding the quadratic and using the the Cauchy-Schwarz inequality on the cross product term, the above bound is bounded above by

$$\rho(\nu_n, \nu) + \rho(\nu, m_\vartheta) + 2\rho^{1/2}(\nu_n, \nu)\rho^{1/2}(\nu, m_\vartheta).$$

In view of (2.3.1), we thus obtain

$$\limsup_n \rho(\nu_n, m_{\vartheta_n}) \leq \rho(\nu, m_\vartheta). \tag{2.3.2}$$

On the other hand, again by the definition of T, ϑ, and ϑ_n here, $\rho(\nu, m_\vartheta) \leq \rho(\nu, m_{\vartheta_n})$ which, together with an argument like the above, implies

$$\begin{aligned} \rho(\nu_n, m_{\vartheta_n}) - \rho(\nu, m_\vartheta) &\geq \rho(\nu_n, m_{\vartheta_n}) - \rho(\nu, m_{\vartheta_n}) \\ &\geq \rho(\nu_n, \nu) - 2\rho^{1/2}(\nu_n, \nu)\rho^{1/2}(\nu, m_{\vartheta_n}). \end{aligned}$$

But

$$\rho(\nu, m_{\vartheta_n}) \leq 6\rho(\nu_n, \nu) + 4\rho(\nu, m_\vartheta) = O(1).$$

Thus, again in view of (2.3.1), $\liminf_n \rho(\nu_n, m_{\vartheta_n}) \geq \rho(\nu, m_\vartheta)$, which together with (2.3.2), yields

$$\rho(\nu_n, m_{\vartheta_n}) \longrightarrow \rho(\nu, m_\vartheta). \tag{2.3.3}$$

From this it follows that $\vartheta_n \to \vartheta$. For, suppose $\vartheta_n \nrightarrow \vartheta$. Then, by the compactness of Θ, there is a subsequence $\{\vartheta_{n_k}\} \subset \{\vartheta_n\}$ such that $\vartheta_{n_k} \to \vartheta_1 \neq \vartheta_0$, and by the continuity of the map $\theta \mapsto \rho(\nu, \theta)$, and by (2.3.1), we obtain $\rho(\nu_{n_k}, m_{\vartheta_{n_k}}) \longrightarrow \rho(\nu, m_{\vartheta_1})$. Hence, by (2.3.3), $\rho(\nu, m_{\vartheta_1}) = \rho(\nu, m_\vartheta)$, implying, in view of the uniqueness of $T(\nu)$, a contradiction, unless $\vartheta_1 = \vartheta$.

Proof of part (c) follows from the identifiability condition (m2), which implies that $T(m_\theta) = \theta$. □

A consequence of this lemma is the following

Corollary 2.3.1 *Suppose H_0, (e1), (e2), (f), (m1), (m2), and (m3) hold. Then, $\theta_n^* \longrightarrow \theta_0$, in probability under H_0.*

Proof. We shall use part (b) of the Lemma 2.3.1 with $\nu_n = \hat{\mu}_{hw}$, $\nu = m_{\theta_0}$. Note that $M_{hw}^*(\theta_0) = \rho(\hat{\mu}_{hw}, m_{\theta_0})$, $\theta_n^* = T(\nu_n)$, and by the identifiability condition (m2), $T(\nu) = \theta_0$ is unique. It thus suffices to prove

$$\rho(\hat{\mu}_{hw}, m_{\theta_0}) = o_p(1). \tag{2.3.4}$$

To show this, we note that by plugging in $Y_i = \mu(X_i) + \varepsilon_i$ and note that $\mu = m_{\theta_0}$ under H_0, and expanding the quadratic integrand, $\rho(\hat{\mu}_{hw}, \mu)$ is bounded above by the sum $2[C_{n1} + C_{n2}(\theta_0)]$, where,

$$\begin{aligned} C_{n1} &:= \int U_n^2(x) d\hat{\varphi}_w(x), \\ C_{n2}(\theta) &:= \int \Big[\mu_n(x,\theta) - \bar{K}_n^*(x) m_\theta(x)\Big]^2 d\hat{\varphi}_w(x), \ \theta \in \mathbb{R}^q. \end{aligned}$$

It thus suffices to show that both of these two terms are $o_p(1)$.

By Fubini, the continuity of f and σ^2, assured by (e2) and (f), and by (k) and (h2),

$$(2.3.5)\quad E\int U_n^2(x)d\varphi(x) = n^{-1}\int EK_h^2(x-X)\sigma^2(X)d\varphi(x) = O(1/nh^d) = o(1),$$

we obtain that

$$(2.3.6)\qquad \int U_n^2(x)d\varphi(x) = O_p((nh^d)^{-1}).$$

Hence, by (2.2.2),

$$C_{n1} \;\leq\; \sup_{x\in\mathcal{I}} |f(x)/\hat{f}_w(x)|^2 \int U_n^2(x)d\varphi(x) = O_p((nh^d)^{-1}).$$

Next, we shall show

$$(2.3.7)\qquad C_{n2}(\theta_0) = o_p(1).$$

Let

$$\begin{aligned} e_h(x,\theta) &= EK_h(x-X)m_\theta(X) = \int K(u)m_\theta(x-uh)f(x-uh)du, \\ e_w^*(x,\theta) &= EK_w^*(x-X)m_\theta(x) = \int K^*(u)m_\theta(x)f(x-uw)du. \end{aligned}$$

By adding and subtracting $e_h(x,\theta)$ and $e_w^*(x,\theta)$ in the quadratic term of the integrand, one obtains that

$$(2.3.8)\qquad C_{n2}(\theta) \leq 3C_{n21}(\theta) + 3C_{n22}(\theta) + 3C_{n23}(\theta), \quad \theta\in\Theta,$$

where

$$(2.3.9)\qquad \begin{aligned} C_{n21}(\theta) &= \int [\mu_n(x,\theta) - e_h(x,\theta)]^2 d\hat{\varphi}_w(x), \\ C_{n22}(\theta) &= \int \left[\bar{K}_n^*(x)m_\theta(x) - e_w^*(x,\theta)\right]^2 d\hat{\varphi}_w(x), \\ C_{n23}(\theta) &= \int [e_h(x,\theta) - e_w^*(x,\theta)]^2 d\hat{\varphi}_w(x). \end{aligned}$$

By Fubini, the fact that the variance is bounded above by the second moment, and by (f), (k) and (m1), one obtains hat

$$(2.3.10)\quad E\tilde{C}_{n21}(\theta_0) \;\leq\; n^{-1}\int EK_h^2(x-X)m_{\theta_0}^2(X)d\varphi(x) = O((nh^d)^{-1}).$$

Hence $C_{n21}(\theta_0) = O_p((nh^d)^{-1})$ follows from (2.2.2). Similarly, one can obtain that $C_{n22}(\theta_0) = O_p((nh^d)^{-1})$. The claim $C_{n23}(\theta_0) = o(1)$ follows from the continuity of m_{θ_0} and f. This completes the proof of (2.3.7), and hence that of (2.3.4) and the corollary. □

Before stating the next result we give a fact that is often used in the proofs below. Under (f), (k), and (h2),

$$\begin{aligned}(2.3.11)\int E\left[n^{-1}\sum_{i=1}^{n}K_h(x-X_i)\alpha(X_i)\right]^2 d\varphi(x) \\ &= n^{-1}\int EK_h^2(x-X)\alpha^2(X)d\varphi(x) + \int [EK_h(x-X)\alpha(X)]^2 d\varphi(x) \\ &= o(1)+O(1)=O(1), \qquad \text{for any continuous function } \alpha \text{ on } \mathcal{I}.\end{aligned}$$

We now proceed to state and prove

Theorem 2.3.1 *Under H_0, (e1), (e2), (f), (k), (m1), (m2), (m3), (h1), and (h2),*

$$(2.3.12)\qquad \hat{\theta}_n \longrightarrow \theta_0, \quad \textit{in probability under } H_0.$$

Proof. We shall again use part (b) of Lemma 2.3.1 with $\nu(x) \equiv m_{\theta_0}(x)$, $\nu_n(x) \equiv m_{\hat{\theta}_n}(x)$. Then by (m2), $\hat{\theta}_n = T(\nu_n)$, $\theta_0 = T(\nu)$, uniquely. It thus suffices to show that

$$(2.3.13)\qquad \rho(m_{\hat{\theta}_n}, m_{\theta_0}) = o_p(1).$$

But observe that

$$\rho(m_{\hat{\theta}_n}, m_{\theta_0}) \le 2[\rho(\hat{\mu}_{hw}, m_{\hat{\theta}_n}) + \rho(\hat{\mu}_{hw}, m_{\theta_0})].$$

Thus, in view of (2.3.4), it suffices to show that

$$M^*_{hw}(\hat{\theta}_n) \equiv \rho(\hat{\mu}_{hw}, m_{\hat{\theta}_n}) = o_p(1). \tag{2.3.14}$$

But this will be implied by the following result.

$$\sup_{\theta} |M_{hw}(\theta) - M^*_{hw}(\theta)| = o_p(1). \tag{2.3.15}$$

For, (2.3.15) implies that

$$M^*_{hw}(\hat{\theta}_n) = M_{hw}(\hat{\theta}_n) + o_p(1), \qquad M^*_{hw}(\theta^*_n) = M_{hw}(\theta^*_n) + o_p(1),$$

$$M^*_{hw}(\hat{\theta}_n) - M^*_{hw}(\theta^*_n) = M_{hw}(\hat{\theta}_n) - M_{hw}(\theta^*_n) + o_p(1). \tag{2.3.16}$$

By the definitions of $\hat{\theta}_n$ and θ^*_n, for every n, the left hand size of (2.3.16) is nonnegative, while the first term on the right hand side is nonpositive. Hence,

$$M^*_{hw}(\hat{\theta}_n) - M^*_{hw}(\theta^*_n) = o_p(1).$$

This together with the fact that $M^*_{hw}(\theta^*_n) \le M^*_{hw}(\theta_0)$ and (2.3.4) then proves (2.3.14).

We now focus on proving (2.3.15). Add and subtract $\mu_n(x,\theta)/\hat{f}_w(x)$ inside the parenthesis of $M^*_{hw}(\theta)$, expand the quadratic, and use the Cauchy-Schwarz inequality on the cross product, to obtain that the left hand side of (2.3.15) is bounded above by

$$\sup_{\theta} C_{n2}(\theta) + 2 \sup_{\theta} \left(C_{n2}(\theta) M_{hw}(\theta)\right)^{1/2}.$$

It thus suffices to show that

$$\text{(2.3.17)} \qquad \sup_\theta C_{n2}(\theta) = o_p(1), \qquad \sup_\theta M_{hw}(\theta) = O_p(1).$$

Recall the notation at (2.3.9). Using the same argument as for (2.3.10), and by the boundedness of m on $\mathcal{I} \times \Theta$, one obtains that

$$\sup_\theta E\tilde{C}_{n21}(\theta) = o(1) = \sup_\theta E\tilde{C}_{n22}(\theta).$$

By the continuity of m_θ and f, one also readily sees that $\tilde{C}_{n23}(\theta) = o(1)$, for each $\theta \in \Theta$. In view of an inequality like (2.3.8) for $\tilde{C}_{n2}$, we thus obtain that $\tilde{C}_{n2}(\theta) = o_p(1)$, for each $\theta \in \Theta$. This and (2.2.2) in turn imply that

$$\text{(2.3.18)} \qquad C_{n2}(\theta) \;\leq\; \sup_{x\in\mathcal{I}} \frac{f^2(x)}{\hat{f}_w^2(x)} \tilde{C}_{n2}(\theta) = o_p(1), \quad \forall\, \theta \in \Theta.$$

Finally, by (m3),

$$\begin{aligned}
& |C_{n2}(\theta_2) - C_{n2}(\theta_1)| \\
\leq\; & 2\|\theta_2 - \theta_1\| \sup_{x\in\mathcal{I}} \frac{f^2(x)}{\hat{f}_w^2(x)} \Big[\int \Big[n^{-1} \sum_{i=1}^n K_h(x - X_i)\ell(X_i) \Big]^2 d\varphi(x) \\
& + \int [\bar{K}_n^*(x)\ell(x)]^2 d\varphi(x) \Big].
\end{aligned}$$

But (2.3.11) applied once with $\alpha \equiv \ell$ and once with $\alpha \equiv 1$ implies that the third factor of this bound is $O_p(1)$. This bound and (2.2.2) together with the compactness o Θ and (2.3.18) completes the proof of the first part of (2.3.17).

To prove the second part of (2.3.17), note that by adding and subtracting $m_{\theta_0}(X_i)$ to the i^{th} summand in $M_{hw}(\theta)$, we obtain

$$M_{hw}(\theta) \leq 2 \sup_{x\in\mathcal{I}} (f(x)/\hat{f}_w(x))^2 \left(\int U_n^2(x) d\varphi(x) + \int Z_n^2(x,\theta) d\varphi(x) \right).$$

But, by the boundedness of m over $\mathcal{I} \times \Theta$ and by (2.3.11) applied with $\alpha \equiv 1$,

$$\text{(2.3.19)} \qquad \sup_{\theta} \int Z_n^2(x,\theta)d\varphi(x) \leq C \int \left(\bar{K}_n(x)\right)^2 d\varphi(x) = O_p(1).$$

This together with (2.3.6) then completes the proof of the second part of (2.3.17), and hence that of the Theorem 2.3.1. □

2.4 Asymptotic distribution of $\hat{\theta}_n$

In this section we shall prove the asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta_0)$. The first step towards this goal is to show that

$$\text{(2.4.1)} \qquad nh^d \|\hat{\theta}_n - \theta_0\|^2 = O_p(1).$$

Recall the definition of Z_n from (2.3.1) and let $D_n(\theta) := \int Z_n^2(x,\theta)d\varphi(x)$. We claim

$$\text{(2.4.2)} \qquad nh^d D_n(\hat{\theta}_n) = O_p(1).$$

To see this, observe that

$$\begin{aligned} nh^d M_{hw}(\theta_0) &= nh^d \int \left(n^{-1} \sum_{i=1}^{n} K_h(x - X_i)\varepsilon_i\right)^2 d\hat{\varphi}_w(x) \\ &\leq nh^d \int U_n^2(x)d\varphi(x) + nh^d \int U_n^2(x)d\varphi(x) \sup_{x \in \mathcal{I}} |f^2(x)/\hat{f}_w^2(x) - 1| \\ &= O_p(1), \end{aligned}$$

by (2.3.5) and (2.2.2). But, by definition,

$$M_{hw}(\hat{\theta}_n) \leq M_{hw}(\theta_0),$$

implying that

$$nh^d M_{hw}(\hat{\theta}_n) = O_p(1).$$

These facts together with the inequality

$$D_n(\theta) \le 2[M_{hw}(\theta_0) + M_{hw}(\hat{\theta}_n)]$$

proves (2.4.2).

Next, we shall show that for any $a > 0$, there exists an N_a such that

$$\text{(2.4.3)} \quad P\left(D_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2 \ge a + \inf_{\|b\|=1} b^T\Sigma_0 b\right) > 1 - a, \qquad \forall\, n > N_a,$$

where Σ_0 is as in (2.3.1). The claim (2.4.1) then will follow from (2.4.3), (2.4.2), the positive definiteness of Σ_0, and the fact

$$nh^d D_n(\hat{\theta}_n) = nh^d \|\hat{\theta}_n - \theta_0\|^2 \left[D_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2\right].$$

To that effect, let

$$\text{(2.4.4)} \; u_n \;:=\; (\hat{\theta}_n - \theta_0),\; d_{ni} := m_{\hat{\theta}_n}(X_i) - m_{\theta_0}(X_i) - u_n^T \dot{m}_{\theta_0}(X_i),\; 1 \le i \le n.$$

We have

$$\begin{aligned}
\frac{D_n(\hat{\theta}_n)}{\|\hat{\theta}_n - \theta_0\|^2} &\le D_{n1} + D_{n2}, \quad \text{where} \\
D_{n1} &= \int \left[n^{-1} \sum_{i=1}^{n} K_h(x - X_i) \left(\frac{d_{ni}}{\|u_n\|}\right)\right]^2 d\varphi(x) \\
D_{n2} &= \int \left[\frac{u_n' \dot{\mu}_n(x, \theta_0)}{\|u_n\|}\right]^2 d\varphi(x).
\end{aligned}$$

By the assumption (m4) and the consistency of $\hat{\theta}_n$, one verifies by a routine argument that $D_{n1} = o_p(1)$. For the second term we notice that

$$D_{n2} \geq \inf_{||b||=1} \Sigma_n(b), \tag{2.4.5}$$

where

$$\Sigma_n(b) := \int \left[b^T \dot{\mu}_n(x, \theta_0)\right]^2 d\varphi(x), \qquad b \in \mathbb{R}^d.$$

By the usual calculations one sees that for each $b \in \mathbb{R}^d$, $\Sigma_n(b) \to b^T \Sigma_0 b$, in probability. Also, note that for any $\delta > 0$, and any two unit vectors b, $b_1 \in \mathbb{R}^d$, $||b - b_1|| \leq \delta$, we have

$$|\Sigma_n(b) - \Sigma_n(b_1)| \leq \delta(\delta + 2) \left[\int n^{-1} \sum_{i=1}^{n} K_h(x - X_i) \, ||\dot{m}_{\theta_0}(X_i)|| d\varphi(x)\right]^2.$$

But the expected value of the r.v.'s inside the square of the second factor tends to $\int ||\dot{m}(x)|| \; f(x) d\varphi(x)$, and hence this factor is $O_p(1)$. From these observations and the compactness of the set $\{b \in \mathbb{R}^d; ||b|| = 1\}$, we obtain that

$$\sup_{||b||=1} |\Sigma_n(b) - b^T \Sigma_0 b| = o_p(1).$$

This fact together with (2.4.5) implies (2.4.3) in a routine fashion, and also concludes the proof of (2.4.1).

We shall now prove the asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta_0)$. The proof is classical in nature. Recall the definitions (2.3.1) and (2.4.4), and let

$$\dot{M}_{hw}(\theta) := -2 \int U_n(x, \vartheta) \, \dot{\mu}_n(x, \theta) d\hat{\varphi}_w(x).$$

Since θ_0 is an interior point of Θ, by the consistency, for sufficiently large n, $\hat\theta_n$ will be in the interior of Θ and $\dot M_{hw}(\hat\theta_n) = 0$, with arbitrarily large probability. But the equation $\dot M_{hw}(\hat\theta_n) = 0$ is equivalent to

$$\text{(2.4.6)} \qquad \int U_n(x)\,\dot\mu_n(x,\hat\theta_n)d\hat\varphi_w(x) \;=\; \int Z_n(x,\hat\theta_n)\,\dot\mu_n(x,\hat\theta_n)d\hat\varphi_w(x).$$

We shall show that $n^{1/2}\times$ the left hand side of this equation converges in distribution to a normal r.v., while the right hand side of this equation equals $R_n(\hat\theta_n - \theta_0)$, for all $n \geq 1$, with $R_n = \Sigma_0 + o_p(1)$.

To establish the first of these two claims, rewrite this r.v. as the sum $S_n + S_{n1} + g_{n1} + g_{n2} + g_{n3} + g_{n4}$, where

$$\begin{aligned}
S_n &= \int U_n(x)\dot\mu_h(x)d\varphi(x), \qquad \dot\mu_h(x) = EK_h(x - X)\dot m_{\theta_0}(X),\\
S_{n1} &= \int U_n(x)\dot\mu_h(x)(\hat f_w^{-2}(x) - f^{-2}(x))dG(x),\\
g_{n1} &= \int U_n(x)\left[\dot\mu_n(x,\theta_0) - \dot\mu_h(x)\right]d\varphi(x)\\
g_{n2} &= \int U_n(x)\left[\dot\mu_n(x,\theta_0) - \dot\mu_h(x)\right](\hat f_w^{-2}(x) - f^{-2}(x))dG(x),\\
g_{n3} &= \int U_n(x)\left[\dot\mu_n(x,\hat\theta_n) - \dot\mu_n(x,\theta_0)\right]d\varphi(x)\\
g_{n4} &= \int U_n(x)\left[\dot\mu_n(x,\hat\theta_n) - \dot\mu_n(x,\theta_0)\right](\hat f_w^{-2}(x) - f^{-2}(x))dG(x).
\end{aligned}$$

We need the following lemmas.

Lemma 2.4.1 *Suppose (e1), (e2), (f), (g), (k), (h1), (h2) hold,* $E|\varepsilon|^{2+\delta} < \infty$, *for some* $\delta > 0$, *and* $\dot m_{\theta_0}(x)$ *is continuous in* $x \in \mathcal{I}$. *Then, under* H_0, $n^{1/2}S_n \longrightarrow_d$

$N(0, \Sigma)$, *where*

$$\begin{aligned} \Sigma &= lim_{h \to 0} \int\int EK_h(x - X)K_h(y - X)\ \sigma^2(X)\dot{\mu}_h(x)\dot{\mu}_h^T(y)d\varphi(x)d\varphi(y) \\ &= \int \frac{\sigma^2(x)\dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)g^2(x)}{f(x)}dx. \end{aligned}$$

Moreover, if f is twice continuously differentiable, and h satisfies (h3), then

$$n^{1/2}|S_{n1}| = o_p(1). \tag{2.4.7}$$

Lemma 2.4.2 *Under H_0, (e1), (e2), (f), (k), (m1), (m2), (m4), (m5), (h1), (h2),*

$$(a)\quad n^{1/2}g_{n1} = o_p(1), \qquad (b)\quad n^{1/2}g_{n2} = o_p(1). \tag{2.4.8}$$

$$(c)\quad n^{1/2}g_{n3} = o_p(1), \qquad (d)\quad n^{1/2}g_{n4} = o_p(1). \tag{2.4.9}$$

The proof of (2.4.7) is facilitated by the following lemma, which along with its proof appears as Theorem 2.2 part (2), in Bosq (1998).

Lemma 2.4.3 *Let $\hat{f}_w$ be the kernel estimate associate with a kernel K^* which satisfies a Lipschitz condition. If f is twice continuously differentiable with a compact support, if w_n is chosen to be $a_n (\log n/n)^{\frac{1}{d+4}}$ where $a_n \longrightarrow a_0 > 0$, then*

$$(\log_k n)^{-1} (n/\log n)^{\frac{2}{d+4}} \sup_{x \in \mathcal{I}} |\hat{f}_w(x) - f(x)| \longrightarrow 0, \ \ a.s.$$

for any positive integer k.

Proof of Lemma 2.4.1. For convenience, we shall give the proof here only for the case $d = 1$, i.e., when $\dot{\mu}_h(x)$ is one dimensional. For multidimensional case, the result can be proved by using linear combination of its components instead of $\dot{\mu}_h(x)$, and applying the same argument.

Let $s_{ni} := \int K_h(x - X_i)\varepsilon_i \dot{\mu}_h(x) d\varphi(x)$, and rewrite

$$n^{1/2} S_n = n^{-1/2} \sum_{i=1}^{n} s_{ni}.$$

Note that $\{s_{ni},\ 1 \le i \le n\}$ are i.i.d. centered r.v.'s for each n. By the L-F C.L.T., it suffices to show that as $n \to \infty$,

$$(2.4.10) \qquad Es_{n1}^2 \to \Sigma,$$

$$(2.4.11) \qquad E\left\{s_{n1}^2 I(|s_{n1}| > n^{1/2}\lambda)\right\} \longrightarrow 0, \qquad \forall \lambda > 0.$$

But,

$$\begin{aligned} Es_{n1}^2 &= E\int K_h(x - X)\,\varepsilon\,\dot{\mu}_h(x)\,d\varphi(x) \times \int K_h(y - X)\,\varepsilon\,\dot{\mu}_h(y) d\varphi(y) \\ &= \int\int EK_h(x - X)K_h(y - X)\sigma^2(X)\dot{\mu}_h(x)\dot{\mu}_h(y) d\varphi(x) d\varphi(y). \end{aligned}$$

By the transformation $x - z = uh$, $y - z = vh$, $z = t$, taking the limit, and using the assumed continuity of σ^2, f, and g, we obtain

$$\begin{aligned} \Sigma &= \lim_{h\to 0}\int\int\int K(u)K(v)\sigma^2(t)\dot{\mu}_h(x + uh)\dot{\mu}_h(x + vh)f(x) \\ &\qquad\qquad \times \frac{g(x + uh)g(x + vh)}{f^2(x + uh)f^2(x + vh)}\,du\,dv\,dx \\ &= \int \frac{\sigma^2(x)\dot{m}_{\theta_0}^2(x)g^2(x)}{f(x)}\,dx. \end{aligned}$$

Hence (2.4.10) is proved.

To prove (2.4.11), note that by the Hölder inequality, the L.H.S. of (2.4.11) is bounded above by

$$
\begin{aligned}
& \lambda^{-\delta/2} n^{-\delta/2} E(s_{n1})^{2+\delta} \\
\leq\ & \lambda^{-\delta/2} n^{-\delta/2} E\left[\left(\int (K_h(x-X)\dot{\mu}_h(x))^{\frac{2+\delta}{2}} d\varphi(x)\right)^2 |\varepsilon|^{2+\delta}\right].
\end{aligned}
$$

This upper bound is seen to be of the order $O((nh^d)^{-\delta/2}) = o(1)$, by (h2), thereby proving (2.4.11).

To prove (2.4.7), by the Cauchy-Schwarz inequality, the boundedness of $\dot{\mu}_h(x)$, (2.3.6), and by Lemma 2.4.3, we obtain

$$
\begin{aligned}
nS_{n1}^2 &\leq Cn \int (U_n(x)\dot{\mu}_h(x))^2 d\varphi(x) \sup_{x\in\mathcal{I}} \left|f^2(x)/\hat{f}_w^2(x) - 1\right|^2 \\
&= n\, O_p((nh^d)^{-1})\, O_p((\log_k n)^2 (\log n/n)^{\frac{4}{d+4}}) \\
&= O_p\left((\log_k n)^2 (\log n)^{\frac{4}{d+4}}\, n^{ad-\frac{4}{d+4}}\right) = o_p(1), \text{by (h3)}.
\end{aligned}
$$

This completes the proof of Lemma 2.4.1. □

Proof of Lemma 2.4.2. By the Cauchy-Schwarz inequality,

$$
\left\|n^{1/2} g_{n1}\right\|^2 \leq \left(n^{1/2}\int U_n^2(x) d\varphi(x)\right)\left(n^{1/2}\int \|\dot{\mu}_n(x,\theta_0) - \dot{\mu}_h(x)\|^2 d\varphi(x)\right).
$$

By (2.3.5), and (h2),

$$
En^{1/2}\int U_n^2(x) d\varphi(x) = O(n^{-1/2}h^{-d}) = o(1). \tag{2.4.12}
$$

To handle the second factor, first note that $\dot{\mu}_n(x,\theta_0) - \dot{\mu}_h(x)$ is an average of centered i.i.d. r.v.'s. Using Fubini, and the fact that variance is bounded above by the second moment, we obtain that the expected value of the second factor of the above bound is bounded above by

$$(2.4.13) \quad n^{-1/2} \int E\|K_h(x-X)\dot{m}_{\theta_0}(x)\|^2 \, d\varphi(x) = O(n^{-1/2}h^{-d}) = o(1).$$

This completes the proof of (2.4.8)(a). This together with (2.2.2) implies (2.4.8)(b).

To prove (c), similarly,

$$\left\|n^{1/2} g_{n3}\right\|^2 \leq n \int U_n^2(x) d\varphi(x) \int \left\|\dot{\mu}_n(x,\hat{\theta}_n) - \dot{\mu}_n(x,\theta_0)\right\|^2 d\varphi(x).$$

But, the second integral is bounded above by

$$\max_{1\leq i\leq n} \|\dot{m}_{\hat{\theta}_n}(X_i) - \dot{m}_{\theta_0}(X_i)\|^2 \int \left(\bar{K}_n(x)\right)^2 d\varphi(x) = o_p(h^d) \times O_p(1),$$

by (2.4.1) and the assumption (m5), and by (2.3.11) applied with $\alpha \equiv 1$. This together with (2.3.5) proves (2.4.9)(c). The proof of (2.4.9)(d) uses (2.4.9)(c) and is similar to that of (2.4.8)(b), thereby completing the proof of the Lemma 2.4.2. □

Next, shall show that the right hand side of (2.4.6) equals $R_n(\hat{\theta}_n - \theta_0)$, where

$$(2.4.14) \qquad R_n = \Sigma_0 + o_p(1).$$

Again, recall the definitions (2.3.1) and (2.4.4). The right hand side of (2.4.6) can be written as the sum $W_{n1} + W_{n2}$, where

$$\begin{aligned}
V_n &:= \int \dot{\mu}_n(x,\hat{\theta}_n) \left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\frac{d_{ni}}{\|u_n\|}\right] d\hat{\varphi}_w(x), \\
W_{n1} &:= \int \left[\dot{\mu}_n(x,\hat{\theta}_n)\, n^{-1}\sum_{i=1}^n K_h(x-X_i)d_{ni}\right] d\hat{\varphi}_w(x) = V_n u_n^T u_n, \\
W_{n2} &:= \int \dot{\mu}_n(x,\hat{\theta}_n)\dot{\mu}_n^T(x,\theta_0)\, d\hat{\varphi}_w(x)\, u_n = L_n\, u_n \quad \text{say,}
\end{aligned}$$

so that the right hand side of (2.4.6) equals $[V_n\, u_n^T + L_n]\, u_n$. But,

$$
\begin{aligned}
\|V_n\| &\leq \max_{1\leq i\leq n} \frac{|d_{ni}|}{\|u_n\|}\, V_{n1}, \\
V_{n1} &:= \int \bar{K}_n(x)\, \|\dot{\mu}_n(x, \hat{\theta}_n)\| d\hat{\varphi}_w(x) \\
&\leq \max_{1\leq i\leq n} \|\dot{m}_{\hat{\theta}_n}(X_i) - \dot{m}_{\theta_0}(X_i)\| \int \bar{K}_n(x) d\hat{\varphi}_w(x) \\
&\qquad\qquad + \int \bar{K}_n(x) \|\dot{\mu}_h(x, \theta_0)\| d\hat{\varphi}_w(x) \\
&= o_p(1) + O_p(1),
\end{aligned}
$$

by (2.2.2), the assumption (m5), and by (2.4.1). This together with (m4) then implies that $\|V_n\| = o_p(1)$, and by the consistency of $\hat{\theta}_n$, we also have $\|V_n u_n^T\| = o_p(1)$.

Next, consider L_n. We have

$$
\begin{aligned}
L_n &= \int \dot{\mu}_n(x, \theta_0)[\dot{\mu}_n(x, \hat{\theta}_n) - \dot{\mu}_n(x, \theta_0)]^T d\hat{\varphi}_w(x) + \int \dot{\mu}_n(x, \theta_0) \dot{\mu}_n^T(x, \theta_0) d\hat{\varphi}_w(x) \\
&= L_{n1} + L_{n2}, \qquad \text{say.}
\end{aligned}
$$

But, by (2.2.1) and (m5), $\|L_{n1}\| = o_p(1)$, while

$$
\begin{aligned}
&\left\| L_{n2} - \int \dot{\mu}_h(x, \theta_0) \dot{\mu}_h^T(x, \theta_0) d\hat{\varphi}_w(x) \right\| \\
&\quad \leq \int \|\dot{\mu}_n(x, \theta_0) - \dot{\mu}_h(x, \theta_0)\|^2 d\hat{\varphi}_w(x) \\
&\qquad\qquad + 2 \int \|\dot{\mu}_n(x, \theta_0) - \dot{\mu}_h(x, \theta_0)\|\, \|\dot{\mu}_h(x, \theta_0)\| d\hat{\varphi}_w(x).
\end{aligned}
$$

But, by (2.2.2) and (2.4.13), this upper bound is $o_p(1)$. Moreover, by usual calculations and using (2.2.2), one also obtains

$$
\int \dot{\mu}_h(x, \theta_0) \dot{\mu}_h^T(x, \theta_0) d\hat{\varphi}_w(x) = \Sigma_0 + o_p(1).
$$

This then proves the claim (2.4.14).

Upon combining these results about the left hand side and the right hand side of (2.4.6), we have the following theorem.

Theorem 2.4.1 *Assume (e1), (e2), (f), (g), (k), (m1) - (m5), and (h3) hold. Suppose, in addition, that $E|\varepsilon|^{2+\delta} < \infty$, for some $\delta > 0$, and f is twice continuously differentiable. Then, under H_0,*

$$n^{1/2}(\hat{\theta}_n - \theta_0) = \Sigma_0^{-1} n^{1/2} S_n + o_p(1). \tag{2.4.15}$$

Consequently, $n^{1/2}(\hat{\theta}_n - \theta_0) \Longrightarrow N(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1})$, where Σ is as in Lemma 2.4.1.

Remark 2.4.1 Upon choosing $g \equiv f$, one sees that

$$\Sigma = \int \sigma^2(x)\dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)f(x)dx, \qquad \Sigma_0 = \int \dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)f(x)dx.$$

It thus follows that in this case the asymptotic distribution of $n^{1/2}(\hat{\theta}_n - \theta_0)$ is the same as that of the least square estimator. This analogy is in flavor similar to the one observed by Beran (1977) when pointing out that the minimum Hellinger distance estimator in the context of density fitting problem is asymptotically like the maximum likelihood estimator.

Consider x and θ are one dimensional case. Let $m_\theta(x) = \theta x$, so $\dot{m}_\theta(x) = x$. Let $\hat{\theta}_n$ be the minimum distance (MD) estimator, $\tilde{\theta}_n$ be the lease absolute distance (LAD) estimator. The variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is denoted by V_1, and

$$V_1 = \frac{\sigma_e^2 \int_I x^2 g^2(x) f^{-1}(x)dx}{\left(\int_I x^2 dG(x)\right)^2}.$$

The variance of $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ is denoted by V_2, and

$$V_2 = \frac{1}{4f_e^2(0)EX^2}.$$

Let $g(x) = f^2(x)l(x)$. then

$$V_1 = \frac{\sigma_e^2 \int_{\mathcal{I}} x^2 f^3(x) l^2(x) dx}{\left(\int_{\mathcal{I}} x^2 f^2(x) l(x)\right)^2}.$$

Now consider the example that $X \sim N(0, \tau^2)$, $l(x) = f^{-1}(x)$, the error distribution is $N(0, \sigma_e^2)$, and $\mathcal{I}$ is a finite interval $[-a, a]$, then

$$\begin{aligned} V_1 &= \frac{\sigma_e^2 \int_{-a}^{a} x^2 f(x) dx}{\left(\int_{-a}^{a} x^2 f(x) dx\right)^2} = \frac{\sigma_e^2}{\int_{-a}^{a} x^2 f(x) dx}, \\ V_2 &= \frac{2\pi\sigma_e^2}{4\tau^2} = \frac{\pi\sigma_e^2}{2\tau^2}. \end{aligned}$$

Take $\tau = 1$ and a large enough such that

$$\int_{-a}^{a} x^2 \psi(x) dx > \frac{2}{\pi} \int_{-\infty}^{\infty} x^2 \psi(x) dx,$$

where ψ stands for the standard normal density, then $V_1 < V_2$.

Or take $a = 1$ and τ small enough such that

$$\int_{-\frac{a}{\tau}}^{\frac{a}{\tau}} y^2 \psi(y) dy > \frac{2}{\pi} \int_{-\infty}^{\infty} y^2 \psi(y) dy,$$

then $V_1 < V_2$.

Remark 2.4.2 Linear regression. Consider the linear regression model, where $q = d + 1$, $\Theta = \mathbb{R}^{d+1}$, and $\mu_\theta(x) = \theta_1 + \theta_2' x$, with $\theta_1 \in \mathbb{R}$, $\theta_2 \in \mathbb{R}^d$. Because now the parameter space is not compact the above results are not directly applicable to this model. But, now the estimator has a closed expression and this regression function satisfies the conditions (m1) - (m5) trivially. The same techniques as above yield the following result.

With the notation in (2.3.1), in this case

$$
\begin{aligned}
\dot{\mu}_n(x,\theta) &\equiv \dot{\mu}_n(x) \equiv \begin{pmatrix} \bar{K}_n(x) \\ \frac{1}{n}\sum_{i=1}^n K_h(x-X_i)X_i \end{pmatrix}, \ \dot{\mu}_h(x) \equiv \begin{pmatrix} EK_h(x-X) \\ EK_h(x-X)X \end{pmatrix} \\
\Sigma_0 &= \int \begin{pmatrix} 1 & x' \\ x & xx' \end{pmatrix} g(x)dx, \qquad \Sigma_n = \int \dot{\mu}_n(x)\dot{\mu}_n(x)'d\hat{\varphi}_w(x), \\
\Sigma &= \int \begin{pmatrix} 1 & x' \\ x & xx' \end{pmatrix} \frac{\sigma^2(x)g^2(x)}{f(x)}dx, \\
M_{hw}(\theta) &= \int [U_n(x) - (\theta-\theta_0)'\dot{\mu}_n(x)]^2 d\hat{\varphi}_w(x).
\end{aligned}
$$

The positive definiteness of Σ_n and direct calculations thus yield

$$(\hat{\theta}_n - \theta_0) = \Sigma_n^{-1}\int \dot{\mu}_n(x)\,U_n(x)\hat{d}\varphi_w(x).$$

From the fact that $\Sigma_n \longrightarrow \Sigma_0$, in probability, parts (a) and (b) of Lemma 2.4.2, and from Lemma 2.4.1 applied to the linear case, we thus obtain that if (e2), (k) and (h3) hold, if the regression function is a linear parametric function, and if $\int \|x\|^2 dG(x) < \infty$, f is twice continuously differentiable, then

$$n^{1/2}(\hat{\theta}_n - \theta_0) \;=\; \Sigma_0^{-1}\int U_n(x)\dot{\mu}_h(x)d\varphi(x) + o_p(1) \Longrightarrow N(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1}).$$

Remark 2.4.3 Tightness. Consider when $d = 1$, from the definition, α_n^* satisfies the equation

$$\int_{\mathcal{I}} \left(m_{\alpha_n^*}(x) - m_{\theta_0}(x)\right) \dot{m}_{\alpha_n^*}(x) dG(x) = A_n + B_n + C_n, \tag{2.4.16}$$

where

$$\begin{aligned}
A_n &= \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)\epsilon_i\right) \dot{m}_{\alpha_n^*}(x) \frac{dG(x)}{\hat{f}_h(x)}, \\
B_n &= \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) m_{\theta_0}(X_i) - \mu_\alpha\right) \dot{m}_{\alpha_n^*}(x) \frac{dG(x)}{\hat{f}_h(x)}, \\
C_n &= \int_{\mathcal{I}} EK_h(x - X_1)\left(m_{\theta_0}(X_1) - m_{\theta_0}(x)\right) \dot{m}_{\alpha_n^*}(x) \frac{dG(x)}{\hat{f}_h(x)},
\end{aligned}$$

and

$$\mu_\alpha = EK_h(x - X_1)\left(m_{\theta_0}(X_1) - m_{\theta_0}(x)\right),$$

and $\dot{m}$ stand for $\partial m / \partial \theta$.

The left hand side of (2.4.16) is approximately

$$(\alpha_n^* - \theta_0) \int_{\mathcal{I}} \dot{m}_{\theta_0}^2(x) dG(x).$$

A_n and B_n are $o_p(1/\sqrt{nh})$ by Cauchy-Schwarz inequality, consistency of α_n^*, and continuity of m on $\theta \in \Theta$. But C_n is approximately

$$\int_{\mathcal{I}} \int K(u)\left(m_{\theta_0}(x - uh) - m_{\theta_0}(x)\right) \dot{m}_{\theta_0}(x) \frac{dG(x)}{f(x)}.$$

So if $m_{\theta_0}(\cdot)$ is not differentiable, and $\sqrt{nh}(m_{\theta_0}(x - h) - m_{\theta_0}(x))$ is divergent, then $\sqrt{nh}(\alpha_n^* - \theta_0)$ is not tight.

2.5 Asymptotic distribution of the minimized distance

This section contains a proof of the asymptotic normality of the minimized distance $M_{hw}(\hat{\theta}_n)$. To state the result precisely, recall the definitions of C_n, $\tilde{C}_n, \hat{C}_n$, Γ, $\hat{\Gamma}_n$ from (1.0.7) and let

$$\Gamma_n \ := \ 2h^d \int\int \left[EK_h(x-X)K_h(y-X)\sigma^2(X)\right]^2 d\varphi(x)d\varphi(y).$$

We shall prove the following

Theorem 2.5.1. *Suppose (e1), (e2), (g), (k), (m1)-(m5) hold, $E\varepsilon^4 < \infty$, h satisfies (h3), and f is twice continuously differentiable.*

Then, under H_0, $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \hat{C}_n)$ asymptotically normally distributed with mean zero and variance Γ. Moreover, $|\hat{\Gamma}_n\Gamma^{-1} - 1| = o_p(1)$.

Consequently, the test that rejects H_0 whenever $\hat{\Gamma}_n^{-1/2} nh^{d/2}|M_{hw}(\hat{\theta}_n) - \hat{C}_n| > z_{\alpha/2}$, is of the asymptotic size α, where z_α is the $100(1-\alpha)\%$ percentile of the standard normal distribution.

Our proof of this theorem is facilitated by the following five lemmas.

Lemma 2.5.1 *If (e1), (e2), (f), (g), (k) hold and if $nh^d \to \infty$, then $nh^{d/2}(\tilde{M}_{hw}(\theta_0) - \tilde{C}_n)$ is asymptotically normally distributed with mean zero and variance Γ.*

Lemma 2.5.2 *Suppose (e1), (e2), (f), (k), (m3), (m4), (m5), (h1), (h2) hold and* $E\varepsilon^4 < \infty$. *Then* $nh^{d/2}\left|M_{hw}(\theta_0) - M_{hw}(\hat{\theta}_n)\right| = o_p(1)$.

Lemma 2.5.3 *Suppose, in addition to (e1), (e2), (k), (m3), (m4), (m5), and* $E\varepsilon^4 < \infty$, f *is twice continuously differentiable and* h *satisfies (h3). Then,*

$$nh^{d/2}\left|M_{hw}(\theta_0) - \tilde{M}_{hw}(\theta_0)\right| = o_p(1).$$

Lemma 2.5.4 *Under the same conditions as in Lemma 2.5.3,*

$$nh^{d/2}(\hat{C}_n - \tilde{C}_n) = o_p(1).$$

Lemma 2.5.5 *Under the conditions of Theorem 2.5.1.,* $\hat{\Gamma}_n - \Gamma = o_p(1)$. *Consequently, the positive definiteness of* Γ *implies,* $|\hat{\Gamma}_n\Gamma^{-1} - 1| = o_p(1)$.

Proof of Lemma 2.5.1. Note that $\tilde{M}_{hw}(\theta_0)$ can be written as the sum of $\tilde{C}_n$ and M_{n2}, where

$$M_{n2} \;=\; n^{-2}\sum_{i\neq j}\int K_h(x - X_i)K_h(x - X_j)\,\varepsilon_i\varepsilon_j\,d\varphi(x).$$

We shall prove that

$$\text{(2.5.1)} \qquad nh^{d/2}M_{n2} \quad \text{is} \quad AN(0, \Gamma_n).$$

To prove (2.5.1), we shall use Theorem 1 of Hall (1984) which is reproduced here for the sake of completeness.

Theorem 2.5.2. *Let* $\tilde{X}_i$, $1 \le i \le n$, *be i.i.d. random vectors, and let*

$$U_n := \sum_{1\le i<j\le n} H_n(\tilde{X}_i, \tilde{X}_j), \qquad G_n(x, y) = EH_n(\tilde{X}_1, x)H_n(\tilde{X}_1, y),$$

where H_n is a sequence of measurable functions symmetric under permutation, with

$$EH_n(\tilde{X}_1, \tilde{X}_2)|\tilde{X}_1) = 0, \text{ a.s., and } EH_n^2(\tilde{X}_1, \tilde{X}_2) < \infty, \text{ for each } n \geq 1.$$

If

$$\left[EG_n^2(\tilde{X}_1, \tilde{X}_2) + n^{-1}EH_n^4(\tilde{X}_1, \tilde{X}_2)\right] \Big/ \left[EH_n^2(\tilde{X}_1, \tilde{X}_2)\right]^2 \longrightarrow 0,$$

then U_n is asymptotically normally distributed with mean zero and variance

$n^2 EH_n^2(\tilde{X}_1, \tilde{X}_2)/2$.

Apply this theorem to $\tilde{X}_i = (X_i^T, \varepsilon_i)^T$ and

$$H_n(\tilde{X}_i, \tilde{X}_j) = n^{-1}h^{d/2} \int K_h(x - X_i)K_h(x - X_j)\, \varepsilon_i \varepsilon_j d\varphi(x),$$

so that

$$nh^{d/2}M_{n2} = 2 \sum_{1 \leq i < j \leq n} H_n(\tilde{X}_i, \tilde{X}_j).$$

Observe that this $H_n(\tilde{X}_1, \tilde{X}_2)$ is symmetric, $E(H_n(\tilde{X}_1, \tilde{X}_2)|\tilde{X}_1) = 0$, and

$$\begin{aligned}
&EH_n^2(\tilde{X}_1, \tilde{X}_2) \\
&= n^{-2}h^d \int\int \left[EK_h(x - X_1)K_h(y - X_1)\sigma^2(X_1)\ \right]^2 d\varphi(x)d\varphi(y) \\
&\leq (n^2h^d)^{-1} \int\int \left[\int K(u)K(\frac{y-x}{h} + u)\sigma^2(x - uh)f(x - uh)du\right]^2 d\varphi(x)d\varphi(y) \\
&< \infty, \qquad \text{for each } n \geq 1.
\end{aligned}$$

Hence, in view of Theorem 2.5.2., we only need to show that

$$EG_n^2(\tilde{X}_1, \tilde{X}_2) \Big/ \left[EH_n^2(\tilde{X}_1, \tilde{X}_2)\right]^2 = o(1), \tag{2.5.2}$$

$$n^{-1}EH_n^4(\tilde{X}_1, \tilde{X}_2) \Big/ \left[EH_n^2(\tilde{X}_1, \tilde{X}_2)\right]^2 = o(1). \tag{2.5.3}$$

To prove (2.5.2) and (2.5.3), it suffices to prove the following three results.

(2.5.4) $$EG_n^2(\tilde{X}_1, \tilde{X}_2) = O(n^{-4}h^d),$$

(2.5.5) $$EH_n^4(\tilde{X}_1, \tilde{X}_2) = O(n^{-4}h^{-d}),$$

(2.5.6) $$EH_n^2(\tilde{X}_1, \tilde{X}_2) = O(n^{-2}).$$

To prove (2.5.4), write a $t \in \mathbb{R}^{d+1}$ as $t^T = (t_1^T, t_2)$, with $t_1 \in \mathbb{R}^d$. Then, for any $t, s \in \mathbb{R}^{d+1}$,

$$\begin{aligned} G_n(t,s) &= n^{-2}h^d \int\int K_h(x-t_1)K_h(z-s_1)t_2 s_2 \\ &\qquad \times E\Big[K_h(x-X_1)K_h(z-X_1)\sigma^2(X_1)\Big]\, d\varphi(x)d\varphi(z). \end{aligned}$$

For the sake of brevity write $d\varphi_{xzwv} = d\varphi(x)d\varphi(z)d\varphi(w)d\varphi(v)$, and

$$\begin{aligned} & EK_h(x-X_1)K_h(z-X_1)\sigma^2(X_1) \\ &= \int K_h(x-t)K_h(z-t)\sigma^2(t)f(t)dt \\ &= h^{-d}\int K(u)K(\frac{z-x}{h}+u)\sigma^2(x-uh)f(x-uh)du \\ &= B_h(z-x), \qquad \text{say.} \end{aligned}$$

Then, by expanding square of the integrals and changing the variables, one obtains that

$$\begin{aligned} & EG_n^2(\tilde{X}_1, \tilde{X}_2) \\ &= n^{-4}h^{2d}\int\int\int\int B_h(x-w)B_h(z-x)B_h(z-v)B_h(v-w)\, d\varphi_{xzwv} \\ &= O(n^{-4}h^d). \end{aligned}$$

This proved (2.5.4). Similarly, one obtains

$$
\begin{aligned}
& EH_n^4(\tilde{X}_1, \tilde{X}_2) \\
= \;& n^{-4}h^{2d}E\left(\int K_h(x-X_1)K_h(x-X_2)\varepsilon_1\varepsilon_2\, d\varphi(x)\right)^4 \\
= \;& n^{-4}h^{2d}\int\int\int\int \left(EK_h(x-X_1)K_h(y-X_1)K_h(s-X_1)K_h(t-X_1)\sigma^4(X_1)\right)^2 \\
& \qquad\qquad d\varphi_{xyst} \\
= \;& O(n^{-4}h^{-d}),
\end{aligned}
$$

and that

$$
\begin{aligned}
& EH_n^2(\tilde{X}_1, \tilde{X}_2) \\
= \;& n^{-2}h^{d}E\int\int K_h(x-X_1)K_h(x-X_2)K_h(y-X_1)K_h(y-X_2)\varepsilon_1^2\varepsilon_2^2\; d\varphi(x)d\varphi(y) \\
= \;& n^{-2}h^{d}\int\int \left[EK_h(x-X_1)K_h(y-X_1)\sigma^2(X_1)\right]^2\, d\varphi(x)d\varphi(y) \\
= \;& O(n^{-2}),
\end{aligned}
$$

thereby verifying (2.5.5) and (2.5.6). This completes the proof of (2.5.1).

By (2.5.7),

$$
\begin{aligned}
& (1/2)n^2EH_n^2(\tilde{X}_1, \tilde{X}_2) \\
= \;& \frac{h^d}{2}\int\int\left(\int K(u)h^{-d}K(\frac{y-x}{h} + u)\sigma^2(x-uh)f(x-uh)\right)^2 d\varphi(x)d\varphi(y) \\
\longrightarrow \;& (1/2)\int(\sigma^2(x))^2 g(x)d\varphi(x)\int(\int K(u)K(v+u)du)^2\, dv
\end{aligned}
$$

by the continuity of σ^2 and f. This complete the proof of Lemma 2.5.1. □

Note that $C_n = n^{-1}E\int K_h^2(x-X_1)\,\varepsilon_1^2\,d\varphi(x)$. Let $e_n := E\int K_h^2(x-X_1)\,\varepsilon_1^2\,d\varphi(x)$. Then, by routine calculations,

$$
\begin{aligned}
& E\left(nh^{d/2}(\tilde{C}_n - C_n)\right)^2 \\
&= E\left(n^{-1}h^{d/2}\sum_{i=1}^{n}\left[\int K_h^2(x-X_i)\varepsilon_i^2\,d\varphi(x) - e_n\right]\right)^2 \\
&\le n^{-1}h^d E\left(\int K_h^2(x-X_1)\varepsilon_i^2\,d\varphi(x)\right)^2 \\
&= n^{-1}h^d E\left[\left(\int K_h^2(x-X_1)\,d\varphi(x)\right)^2 \varepsilon_1^4\right] \\
&= O((nh^d)^{-1}) = o(1).
\end{aligned}
$$

Combining this with the Lemma 2.5.1, one obtains that $nh^{d/2}(\tilde{M}_{hw}(\theta_0) - C_n)$ is $AN(0, \Gamma_n)$.

Proof of Lemma 2.5.2. Recall the definitions of U_n and Z_n from (2.3.1). To prove part (b), add and subtract $m_{\theta_0}(X_i)$ to the i^{th} summand inside the square integrand of $M_{hw}(\hat{\theta}_n)$, to obtain that

$$
\begin{aligned}
M_{hw}(\theta_0) - M_{hw}(\hat{\theta}_n) &= 2\int U_n(x)Z_n(x,\hat{\theta}_n)\,d\hat{\varphi}_w(x) - \int Z_n^2(x,\hat{\theta}_n)\,d\hat{\varphi}_w(x) \\
&= 2Q_1 - Q_2, \qquad \text{say.}
\end{aligned}
$$

We need to show that

$$
(2.5.7) \qquad (i)\quad nh^{d/2}Q_1 = o_p(1), \qquad (ii)\quad nh^{d/2}Q_2 = o_p(1).
$$

By subtracting and adding $(\hat{\theta}_n - \theta_0)^T \dot{m}_{\theta_0}(X_i)$ to the i^{th} summand of the second factor of integrand in Q_1, we can rewrite Q_1 as the sum of Q_{11} and Q_{12}, where

$$\begin{aligned} Q_{11} &= \int U_n(x) \left[n^{-1} \sum_{i=1}^{n} K_h(x - X_i) d_{ni} \right] d\hat{\varphi}_w(x), \\ Q_{12} &= (\hat{\theta}_n - \theta_0)^T \int U_n(x) \dot{\mu}_n(x, \theta_0) d\hat{\varphi}_w(x) \end{aligned}$$

where d_{ni} are as in (2.4.4). By (2.4.1), for every $\eta > 0$, there is a $k < \infty$, $N < \infty$, such that $P(A_n) \geq 1 - \eta$, for all $n > N$, where $A_n := \{(nh^d)^{1/2} \|\hat{\theta}_n - \theta_0\| < k\}$. By the Cauchy-Schwarz inequality, (2.2.2), (2.3.6) and the fact that

(2.5.8) $$\int \left(\bar{K}_n(x)\right)^2 d\hat{\varphi}_w(x) = O_p(1),$$

we obtain that on the event A_n, $nh^{d/2}|Q_{11}|$ is bounded above by

$$n^{1/2} \|\hat{\theta}_n - \theta_0\| (nh^d)^{1/2} \sup_{i, (nh^d)^{1/2} \|\theta - \theta_0\| < k} \frac{|d_{ni}|}{\|\hat{\theta}_n - \theta_0\|} O_p((nh^d)^{-1/2}).$$

This bound in turn is $o_p(1)$ by Theorem 2.4.1 and the assumption (m4). Hence to prove (2.5.7)(i), it remains to prove that $nh^{d/2}|Q_{12}| = o_p(1)$.

But Q_{12} can be rewritten as the sum of Q_{121} and Q_{122}, where

$$\begin{aligned} Q_{121} &= (\hat{\theta}_n - \theta_0)^T \int U_n(x) \dot{\mu}_n(x, \hat{\theta}_n) d\hat{\varphi}_w(x), \\ Q_{122} &= (\hat{\theta}_n - \theta_0)^T \int U_n(x) \left[\dot{\mu}_n(x, \hat{\theta}_n) - \dot{\mu}_n(x, \theta_0) \right] d\hat{\varphi}_w(x). \end{aligned}$$

Arguing as above, on the event A_n, $(nh^{d/2}|Q_{122}|)^2$ is bounded above by

$$n^2 h^d \|\hat{\theta}_n - \theta_0\|^2 \max_{1 \leq i \leq n} \|\dot{m}_{\hat{\theta}_n}(X_i) - \dot{m}_{\theta_0}(X_i)\|^2 O_p((nh^d)^{-1}) = o_p(1),$$

by (2.2.2), (2.3.6), (2.5.8), and assumptions (m5) and (h2).

Next, note that Q_{121} is the same as the expression in the left hand side of (2.4.6). Thus, it is equal to

$$
\begin{aligned}
(2.5.9) \quad & (\hat{\theta}_n - \theta_0)^T \int Z_n(x, \hat{\theta}_n)\mu_n(x, \hat{\theta}_n) d\hat{\varphi}_w(x) \\
&= (\hat{\theta}_n - \theta_0)^T \int Z_n(x, \hat{\theta}_n)\mu_n(x, \theta_0) d\hat{\varphi}_w(x) \\
&\quad +(\hat{\theta}_n - \theta_0)^T \int Z_n(x, \hat{\theta}_n) \left[\mu_n(x, \hat{\theta}_n) - \mu_n(x, \theta_0)\right] d\hat{\varphi}_w(x) \\
&= D_1 + D_2, \qquad \text{say,}
\end{aligned}
$$

But, by the Cauchy-Schwarz inequality, (2.2.2), (2.3.19), and (2.5.8), $nh^{d/2}|D_1|$ is bounded above by

$$nh^{d/2}\|\hat{\theta}_n - \theta_0\|^2 O_p(1) = o_p(1),$$

by Theorem 2.4.1 and the assumption (m5) and (h2). Similarly, one shows $nh^{d/2}|D_2|$ is bounded above by

$$nh^{d/2}\|\hat{\theta}_n - \theta_0\|^2 o_p(1) = o_p(1).$$

This completes the proof of (2.5.7)(i).

The proof of (2.5.7)(ii) similar. Details are left out for the sake of brevity. □

Proof of Lemma 2.5.3. Note that

$$
\begin{aligned}
nh^{d/2}|M_{hw}(\theta_0) - \tilde{M}_{hw}(\theta_0)| &\leq nh^{d/2} \int U_n^2(x) d\varphi(x) \sup_{x \in \mathcal{I}} |f^2(x)/\hat{f}_w^2(x) - 1| \\
&= nh^{d/2} O_p((nh^d)^{-1}) O_p((\log_k n)(\log n/n)^{\frac{d}{d+4}}) = o_p(1),
\end{aligned}
$$

by (2.3.5) and Lemma 2.4.3. Hence the lemma. □

Proof of Lemma 2.5.4. Let

$$t_i = m_{\hat{\theta}_n}(X_i) - m_{\theta_0}(X_i), \quad \Delta_n(x) := f^2(x)\left(\hat{f}_w^{-2}(x) - f^{-2}(x)\right).$$

Then,

$$
\begin{aligned}
\hat{C}_n &= \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(x-X_i)(\varepsilon_i-t_i)^2 d\hat{\varphi}_w(x) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(x-X_i)(\varepsilon_i-t_i)^2 d\varphi(x) \\
&\qquad + \frac{1}{n^2}\sum_{i=1}^{n}\int K_h^2(x-X_i)(\varepsilon_i-t_i)^2 \Delta_n(x) d\varphi(x) \\
&= A_{n1} + A_{n2}, \qquad \text{say.}
\end{aligned}
$$

In order to prove the lemma it suffices to prove that

$$
(2.5.10)\ (a)\quad nh^{d/2}(A_{n1}-\tilde{C}_n) = o_p(1), \quad \text{and} \quad (b)\quad nh^{d/2}A_{n2} = o_p(1).
$$

By expanding the quadratic term in the integrand, A_{n1} can be written as the sum of $\tilde{C}_n$, A_{n12}, and A_{n13}, where

$$
\begin{aligned}
A_{n12} &= n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)t_i^2 d\varphi(x), \\
A_{n13} &= -2n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i t_i d\varphi(x).
\end{aligned}
$$

But $|A_{n12}| \le \max_{1\le i\le n}|t_i|^2 n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)d\varphi(x)$. By (m4) and (2.4.1), one obtains that $\max_{i\le n}|t_i|^2 = O_p((nh^d)^{-1})$. Moreover, by the usual calculation, one obtains that

$$
n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)d\varphi(x) = O_p((nh^d)^{-1}).
$$

Hence,

$$
|A_{n12}| = O_p((nh^d)^{-1})O_p((nh^d)^{-1}) = O_p((nh^d)^{-2}).
$$

Similarly,

$$
\begin{aligned}
|A_{n13}| &\le 2\max_{i\le n}|t_i|n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)|\varepsilon_i|d\varphi(x)\\
&= O_p((nh^d)^{-1/2})O_p((nh^d)^{-1}) = O_p((nh^d)^{-3/2}).
\end{aligned}
$$

Hence

$$
\begin{aligned}
|nh^{d/2}(A_{n1}-\tilde{C}_n)| &= nh^{d/2}\left(O_p((nh^d)^{-2}) + O_p((nh^d)^{-3/2})\right)\\
&= O_p((nh^{-3d/2})^{-1}) + O_p((nh^{2d})^{-1/2}) = o_p(1).
\end{aligned}
$$

To prove the part (b) of (2.5.10), note that A_{n2} can be written as the sum of A_{n21}, A_{n22}, and A_{n23}, where

$$
\begin{aligned}
A_{n21} &= n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i^2\Delta_n(x)d\varphi(x),\\
A_{n22} &= n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)t_i^2\Delta_n(x)d\varphi(x),\\
A_{n23} &= -2n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i t_i\Delta_n(x)d\varphi(x).
\end{aligned}
$$

By taking the expected value and the usual calculation, one obtains that

$$
n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i^2 d\varphi(x) = O_p((nh^d)^{-1}).
$$

Hence

$$
\begin{aligned}
|nh^{d/2}A_{n21}| &\le \sup_{x\in\mathcal{I}}|\Delta_n(x)|n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X_i)\varepsilon_i^2 d\varphi(x)\\
&= nh^{d/2}O_p(\log_k n\,(\log n/n)^{\frac{2}{d+4}})O_p((nh^d)^{-1})\\
&= O_p(h^{-d/2}\log_k n\,(\log n/n)^{\frac{2}{d+4}}) = o_p(1),
\end{aligned}
$$

by Lemma 2.4.3 and (2.2.2). Similarly, one obtains that

$$
\begin{aligned}
|nh^{d/2}A_{n22}| &\leq \sup_{x\in\mathcal{I}}|\Delta_n(x)| \max_{1\leq i\leq n}|t_i|^2 n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X)i)d\varphi(x) \\
&= nh^{d/2}O_p(\log_k n\,(\log n/n)^{2/(d+4)})O_p((nh^d)^{-1})O_p((nh^d)^{-1}) \\
&= o_p((nh^{3d/2})^{-1}) = o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
|nh^{d/2}A_{n23}| &\leq 2\sup_{x\in\mathcal{I}}|\Delta_n(x)| \max_{i\leq n}|t_i| n^{-2}\sum_{i=1}^{n}\int K_h^2(x-X)i)|\varepsilon_i|d\varphi(x) \\
&= nh^{d/2}O_p(\log_k n\,(\log n/n)^{2/(d+4)})O_p((nh^d)^{-1/2})O_p((nh^d)^{-1}) \\
&= o_p((nh^{2d})^{-1/2}) = o_p(1),
\end{aligned}
$$

thereby completing the proof of the part (b) of (2.5.10), and hence that of the lemma. $\square$.

Proof of Lemma 2.5.5. Define

$$
\tilde{\Gamma}_n = h^d n^{-2}\sum_{i=1}^{m}\left(\int K_h(x-X_i)K_h(x-X_j)\varepsilon_i\varepsilon_j d\varphi(x)\right)^2 = \sum_{i=1}^{m} H_n^2(\tilde{X}_i,\tilde{X}_j).
$$

We shall first prove

$$
\hat{\Gamma}_n - \tilde{\Gamma}_n = o_p(1), \tag{2.5.11}
$$

$$
\tilde{\Gamma}_n - \Gamma_n = o_p(1). \tag{2.5.12}
$$

The claim of this lemma follows from these results and the fact that $\Gamma_n \longrightarrow \Gamma$.

For the sake of convenience, write $K_h(x - X_i)$ by $K_i(x)$. Now, rewrite $\hat{\Gamma}_n$ as the sum of the following terms:

$$
\begin{aligned}
B_1 &= h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) d\varphi(x) \right)^2, \\
B_2 &= h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) \Delta_n(x) d\varphi(x) \right)^2, \\
B_3 &= -2\frac{h^d}{n^2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) d\varphi(x) \right) \\
&\qquad \times \left(\int K_i(x) K_j(x) (\varepsilon_i - t_i)(\varepsilon_j - t_j) \Delta_n(x) d\varphi(x) \right).
\end{aligned}
$$

In order to prove (2.5.11), it suffices to prove that

$$
\text{(2.5.13)} \quad B_1 - \tilde{\Gamma}_n = o_p(1), \qquad B_2 = o_p(1), \qquad \text{and} \qquad B_3 = o_p(1).
$$

By taking the expected value, Fubini, and usual calculation one obtains that

$$
\text{(2.5.14)} \qquad h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) |\varepsilon_i| |\varepsilon_j| d\varphi(x) \right)^2 = O_p(1),
$$

$$
\text{(2.5.15)} \qquad h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) |\varepsilon_i| d\varphi(x) \right)^2 = O_p(1),
$$

$$
\text{(2.5.16)} \qquad h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) d\varphi(x) \right)^2 = O_p(1).
$$

Further more,

$$
\text{(2.5.17)} \qquad \sup_{x \in \mathcal{I}} \Delta_n(x) = o_p(1), \qquad \text{by (2.2.2)}
$$

$$
\text{(2.5.18)} \qquad \max_{i \le i \le n} |t_i| = o_p(1). \qquad \text{by (m4) and (2.4.1).}
$$

Note that by expanding $(\varepsilon_i - t_i)(\varepsilon_j - t_j)$ and the quadratic terms, $|B_1 - \tilde{\Gamma}_n|$ is bounded above by the sum of B_{12} and B_{13}, where

$$\begin{aligned} B_{12} &= h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) (|t_i t_j| + |\varepsilon_i t_i| + |t_i \varepsilon_j|) d\varphi(x) \right)^2, \\ B_{13} &= h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) |\varepsilon_i \varepsilon_j| d\varphi(x) \right) \\ &\qquad \times \left(\int K_i(x) K_j(x) (|t_i t_j| + |\varepsilon_i t_i| + |t_i \varepsilon_j|) d\varphi(x) \right). \end{aligned}$$

But $B_{12} = o_p(1)$ by (2.5.15), (2.5.16), (2.5.18), and the fact that $\{t_i\}$ are free of x. It further implies that $B_{13} = o_p(1)$ by (2.5.14) and applying the Cauchy-Schwarz inequality to the double sum. Hence $|B_1 - \tilde{\Gamma}_n| = o_p(1)$.

Note that

$$\begin{aligned} B_2 &\leq \sup_{x \in \mathcal{I}} |\Delta_n(x)| h^d n^{-2} \sum_{i=1}^{m} \left(\int K_i(x) K_j(x) |\varepsilon_i - t_i| |\varepsilon_j - t_j| d\varphi(x) \right)^2 \\ &= o_p(1) O_p(1) = o_p(1), \end{aligned}$$

by the inequality

$$|\varepsilon_i - t_i||\varepsilon_j - t_j| \leq |\varepsilon_i \varepsilon_j| + (|t_i t_j| + |\varepsilon_i t_i| + |t_i \varepsilon_j|),$$

and expanding the quadratic terms, and by (2.5.17), (2.5.14), and the result that B_{12} and B_{13} are both $o_p(1)$. Finally, again an application of the Cauchy-Schwarz inequality to the double sum yields $B_3 = o_p(1)$. This completes the proof of (2.5.13), and hence that of (2.5.11).

To proved (2.5.12), note that $\Gamma_n = E\tilde{\Gamma}_n$. Hence

$$\begin{aligned} E\left(\tilde{\Gamma}_n - \Gamma_n\right)^2 &\leq \sum_{i=1}^{m} EH_n^4(\tilde{X}_i, \tilde{X}_j) + c \sum_{i \neq j \neq k} EH_n^2(\tilde{X}_i, \tilde{X}_j) H_n^2(\tilde{X}_j, \tilde{X}_k) \\ &\leq (n^2 + cn^3) EH_n^4(\tilde{X}_1, \tilde{X}_2) \end{aligned}$$

for some constant c by expanding the quadratic terms and the fact that the variance is bounded above by the second moment. But by (2.5.5), this upper bound is $O((nh^d)^{-1}) = o(1)$. Hence (2.5.12) is proved, and so is the Lemma 2.5.5. $\square$

2.6 Simulations

The simulation study of the distribution of $\hat{\theta}_n$ and the the minimum distance $M_{hw}(\hat{\theta}_n)$ was conducted for a linear regression function family $\{m_\theta(x) = \theta x, \theta \in \mathbb{R}\}$ with various sample sizes. First we generated random sample $\{X_i\}_1^n$, $n = 50$, from uniform $[-1.1]$ distribution, and random sample $\{\varepsilon_i\}_1^n$ from normal distribution with mean zero and standard deviation 0.1. Then let $Y_i = m_{\theta_0}(X_i) + \varepsilon_i$, with $\theta_0 = 1$, $i = 1, ..., n$. The kernel functions we used to construct the test statistic are

$$K(u) = K^*(u) = 3/4(1 - u^2)I\{|u| \leq 1\}.$$

The bandwidth h is chosen to be $n^{-1/3}$ and w is chosen to be $n^{-1/5}$. The measure G is a measure with Lebeague density $g(x) = 1$ on $[-1.1]$.

Recall that $\hat{\theta}_n$ is the minimizer of $M_{hw}(\theta)$. By taking the derivative of $M_{hw}(\theta)$ in θ and solving the equation of $\partial M_{hw}(\theta)/\partial\theta = 0$, the minimum distance estimate of θ_0 is given by

$$\hat{\theta}_n = A_n/B_n,$$

where

$$
\begin{aligned}
A_n &= \int_{-1}^{1} (\sum_{i=1}^{n} K_h(x - X_i)Y_i)(\sum_{i=1}^{n} K_h(x - X_i)X_i)(\sum_{i=1}^{n} K_w(x - X_i))^{-2}dx \\
B_n &= \int_{-1}^{1} (\sum_{i=1}^{n} K_h(x - X_i)X_i)^2(\sum_{i=1}^{n} K_w(x - X_i))^{-2}dx.
\end{aligned}
$$

The normalized value of $\hat{\theta}_n$ then is calculated by $\sqrt{n}(\hat{\theta}_n - 1)$. The corresponding minimum distance and the estimate of its asymptotic mean are calculated by

$$
\begin{aligned}
M_{hw}(\hat{\theta}_n) &= \int_{-1}^{1} \left(\sum_{i=1}^{n} K_h(x - X_i)(Y_i - \hat{\theta}_n X_i)\right)^2 \left(\sum_{i=1}^{n} K_w(x - X_i)\right)^{-2} dx, \\
\hat{C}_n &= \int_{-1}^{1} \left(\sum_{i=1}^{n} K_h^2(x - X_i)(Y_i - \hat{\theta}_n X_i)^2\right) \left(\sum_{i=1}^{n} K_w(x - X_i)\right)^{-2} dx.
\end{aligned}
$$

The value of the test statstic is calculated by $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \tilde{C}_n)$. In order to plot a density curve, we repeated the above sampling and calculations for 1000 times. The density curves of normalized $\hat{\theta}_n$ and the test statistic are plot by using density plot command with Gussian kernel option in SPLUS2000. We also did the above simulation for $n = 100$ and $n = 200$.

The first three graphs in Figure 2.1 are the Monte Carlo density curves of $\sqrt{n}(\hat{\theta}_n - 1)$ from 1000 runs with sample size $n = 50$, $n = 100$, $n = 200$ respectively. The fourth graph is the $N(0, (0.173025)^2)$ density, the density curve of the limiting distribution of $\sqrt{n}(\hat{\theta}_n - 1)$ based on the theorem we obtained in section 4. The first three graphs in Figure 2 are the Monte Carlo density curves of $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \tilde{C}_n)$ from 1000 runs with sample size $n = 50$, $n = 100$, and $n = 200$ respectively. The fourth graph is the density curve of the limiting distribution of $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \tilde{C}_n)$ in Theorem 5.1. which is $N(0, (0.026344)^2)$ in the present case. The graphs show

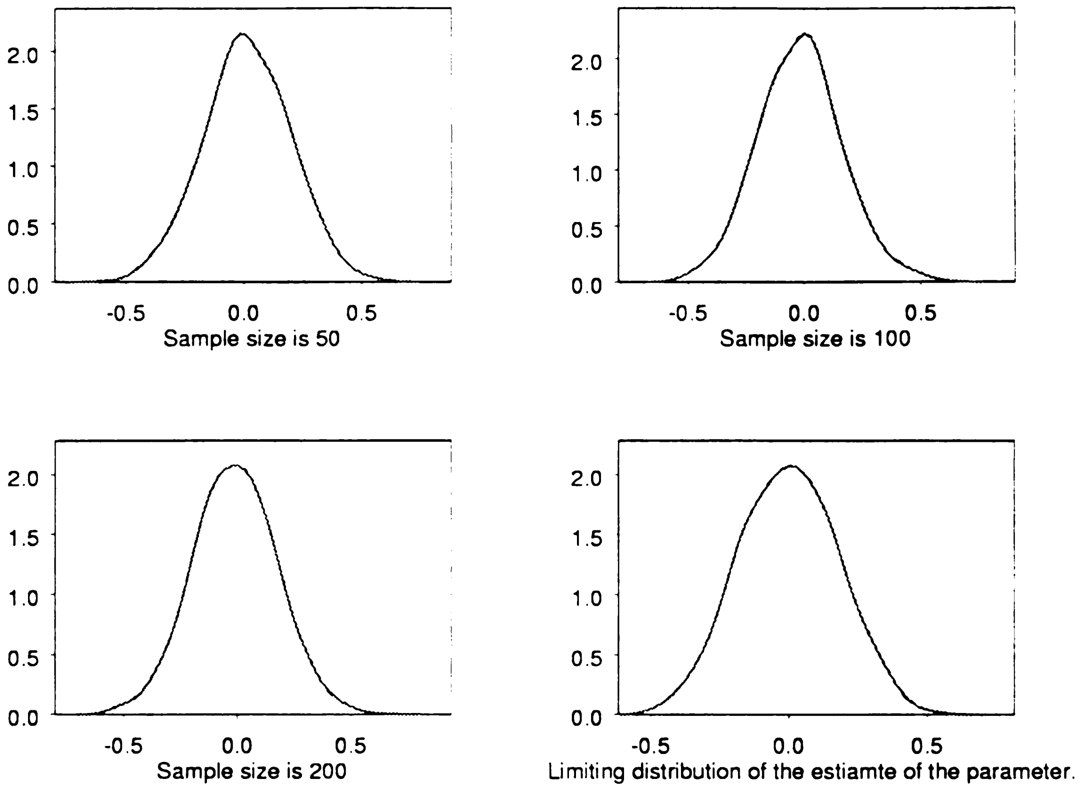


Figure 2.1: *The density curve of* $\sqrt{n}(\hat{\theta}_n - 1)$.

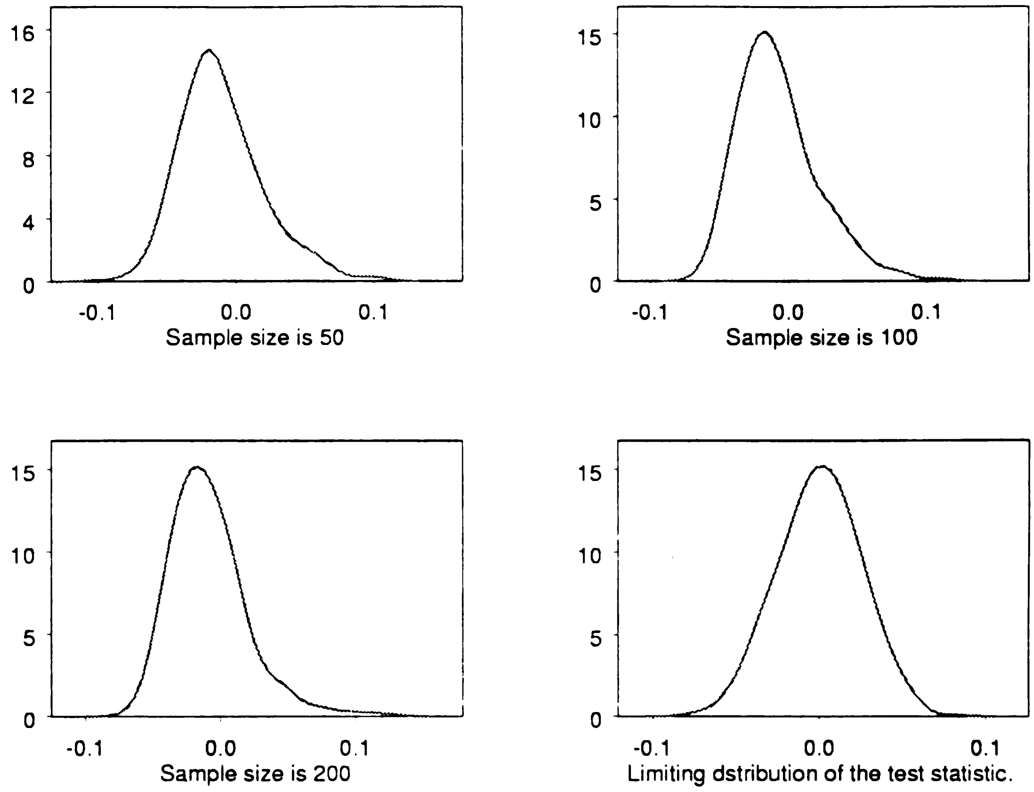


Figure 2.2: *The density curve of* $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \tilde{C}_n)$.

that the distribution of $\sqrt{n}(\hat{\theta}_n - 1)$ resembles the asymptotic normal distribution quite well even for sample size is 50. The distribution of $nh^{d/2}(M_{hw}(\hat{\theta}_n) - \tilde{C}_n)$ has a small negative bias compared with the asymptotic normal distribution for all three sample sizes. But the bias decreases as n increases.

A simulation for $d = 2$ and $m = 2$ was also conducted. The hypothesis to be tested is

$$H_0: \mu(x) = 0.5x_1 + 0.8x_2, \qquad vs. \quad H_1: H_0 \text{ is not true.}$$

The parametric model to be fitted is

$$\{m_\theta(x_1, x_2) = \theta_1 x_1 + \theta_2 x_2,\ \theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2\ x = (x_1, x_2)^T \in \mathbb{R}^2\}.$$

We chose the following five models to generate simulated data from:

model 0. $\quad Y_i = 0.5X_{1i} + 0.8X_{2i} + \varepsilon_i,$

model 1. $\quad Y_i = 0.5X_{1i} + 0.8X_{2i} + 0.3(X_{1i} - 0.5)(X_{2i} - 0.2) + \varepsilon_i,$

model 2. $\quad Y_i = 0.5X_{1i} + 0.8X_{2i} + 0.3X_{1i}X_{2i} - 0.5 + \varepsilon_i,$

model 3. $\quad Y_i = 0.5X_{1i} + 0.8X_{2i} + 1.4(exp\{-0.2X_{1i}^2\} - exp\{0.7X_{2i}^2\}) + \varepsilon_i,$

model 4. $\quad Y_i = I\{X_{2i} > 0.2\}X_{1i} + \varepsilon_i,$

The error distribution is $N(0, 0.3)$. X_{1i} are i.i.d $N(0, 0.7)$ and X_{2i} are i.i.d $N(0, 1)$. The sample sizes chosen are 30, 50, 100, and 200. The nominal level that is used to implement the test is $\alpha = 0.05$. There are 1000 replications for each combination of $(model, sample\, size)$. Data from model 0 are used to study the empirical size, and the data from models 1 to 4 are used to study the empirical power of the test. The empirical size (power) is computed by

$$Relative\, frequency\, of\ (\, value\, of\, the\, test\, statistic\ > F^{-1}(1-\alpha)),$$

where F is the asymptotic distribution of the test statistics under H_0.

The bandwidth h is chosen to be $n^{-1/4.5}$ and w is chosen to be $(\log n/n)^{1/(d+4)}$, the measure G is taken to be the uniform distribution on $[-1, 1]$.

The density curves of normalized θ_n and $M_{hw}(\theta_n)$ are plotted by using density plot command with Gussian kernel option in SPLUS2000 for one dimenstion and Surface-Spline Fine Grid for two dimension, where $\theta_n = (\theta_{1n}, \theta_{2n})^T$ and $\theta_0 = (0.5, 0.8)^T$.

The results of the power study are shown in the table. The tables gives the empirical sizes and powers for testing model 0 against models 1 to 4.

The simulation results of the densities of $\sqrt{n}(\theta_n - \theta_0)$, and the minimum distance test statistics are shown in Figure 2.3 to 2.9.

Figure 2.3 is the Monte Carlo density curves of $\sqrt{n}(\theta_{1n} - 0.5)$ from 1000 runs with sample size $n = 30$, $n = 50$, $n = 100$, $n = 200$ respectively. Figure 2.4 is the Monte Carlo density curves of $\sqrt{n}(\theta_{2n} - 0.8)$. Figure 2.5 is the Monte Carlo density surface of $\sqrt{n}(\theta_n - \theta_0)$ when $n = 30$. Figure 2.6 is the Monte Carlo density surface of $\sqrt{n}(\theta_n - \theta_0)$ when $n = 50$. Figure 2.7 is the Monte Carlo density surface of $\sqrt{n}(\theta_n - \theta_0)$ when $n = 100$. Figure 2.8 is the Monte Carlo density surface of $\sqrt{n}(\theta_n - \theta_0)$ when $n = 200$. Figure 2.9 is the Monte Carlo density of the test statistic under H_0 with sample size $n = 30$, $n = 50$, $n = 100$, $n = 200$.

In the following figures, "$\cdots\cdots$" is for $n = 30$, "$- \cdot -$" is for $n = 50$, "$- \ - \ -$" is for $n = 100$, and a heavy solid line is for $n = 200$.

Table 2.1: *Empirical sizes and powers for testing models 0 vs. model 1 to 4.*

	$n = 30$	n=50	n=100	n=200
model 0	0.005	0.022	0.036	0.049
model 1	0.003	0.062	0.670	0.895
model 2	0.931	0.999	1.000	1.000
model 3	0.461	0.975	1.000	1.000
model 4	0.035	0.368	0.977	1.000

Figure 2.3: *The density of* $\sqrt{n}(\theta_{1n} - 0.5)$.

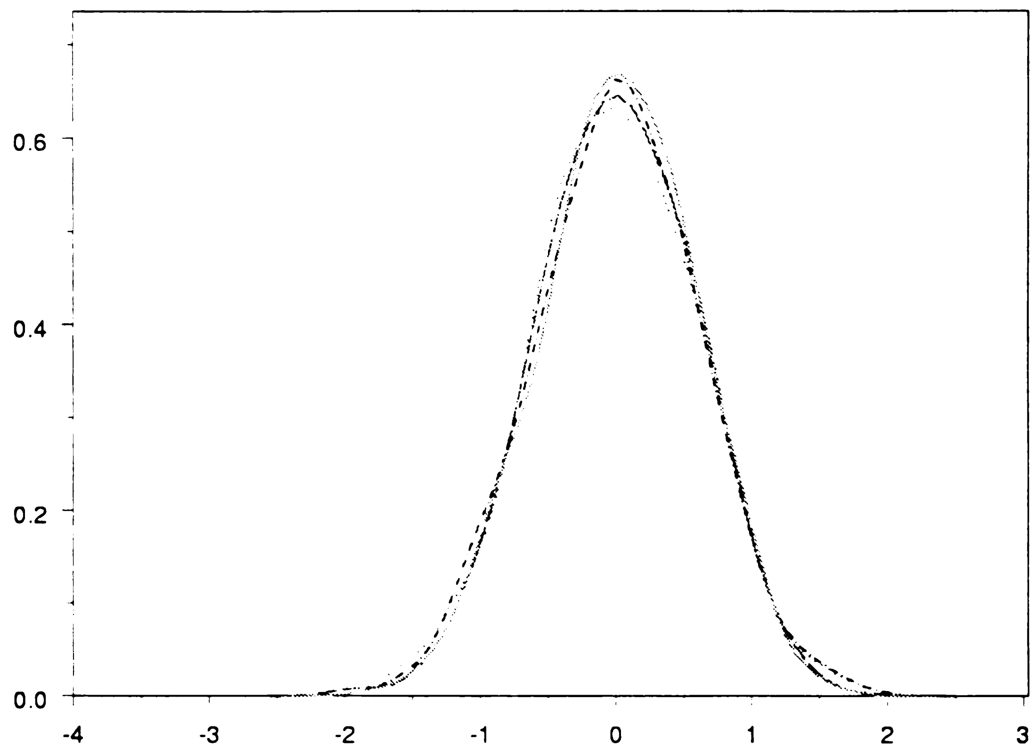


Figure 2.4: *The density of* $\sqrt{n}(\theta_{2n} - 0.8)$.

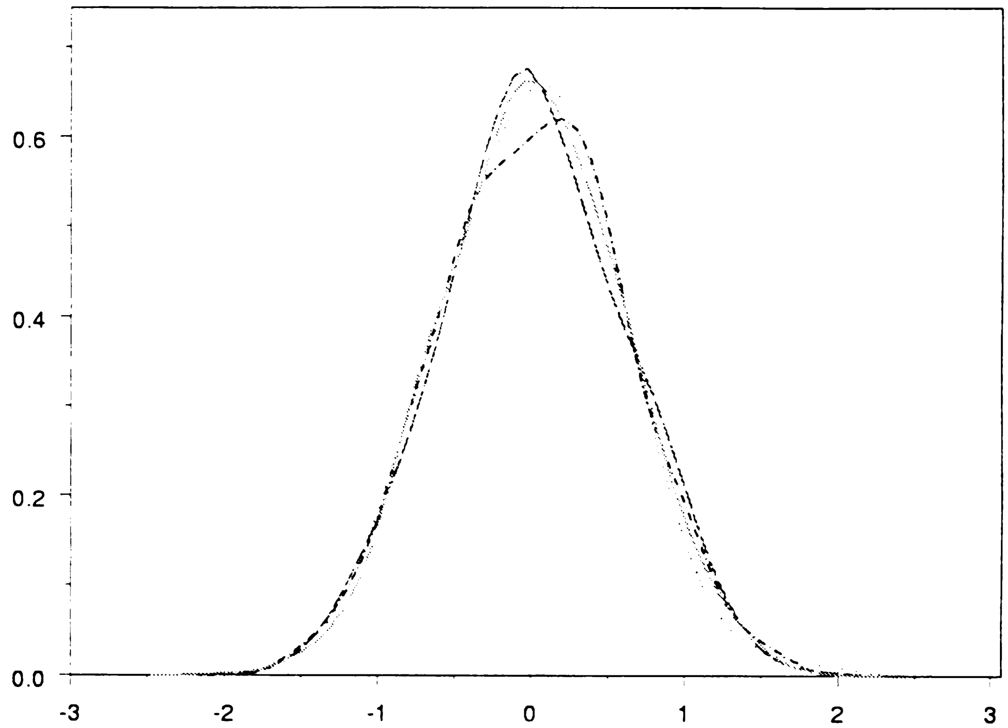


Figure 2.5: *The 2 dimentional density of* $\sqrt{n}(\theta_n - \theta_0)$ *when* $n = 30$.

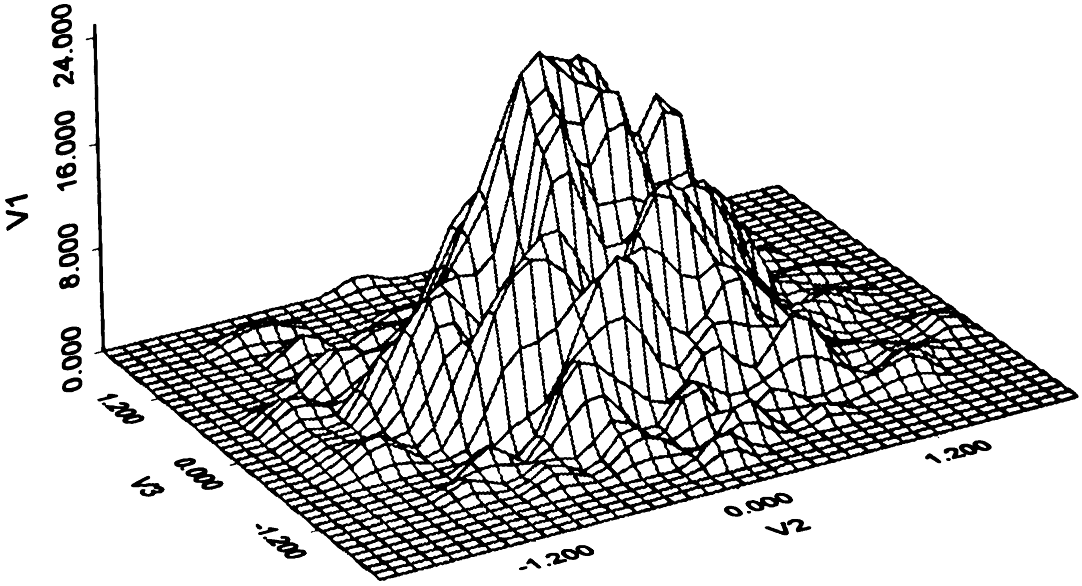


Figure 2.6: *The 2 dimentional density of* $\sqrt{n}(\theta_n - \theta_0)$ *when* $n = 50$.

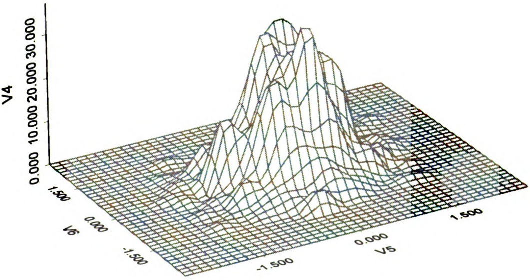


Figure 2.7: *The 2 dimentional density of* $\sqrt{n}(\theta_n - \theta_0)$ *when* $n = 100$.

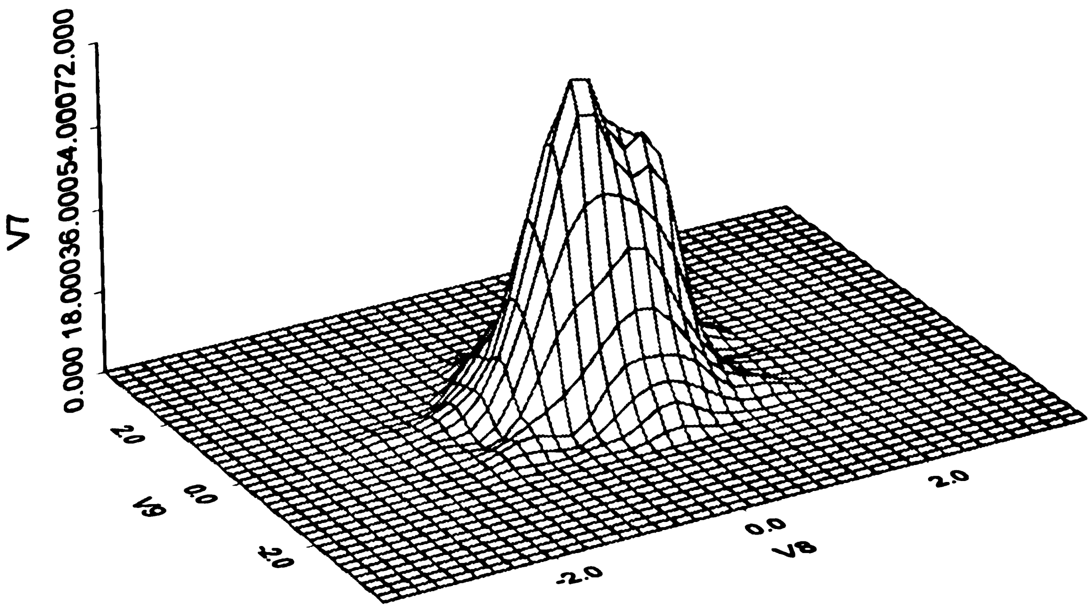


Figure 2.8: *The 2 dimentional density of* $\sqrt{n}(\theta_n - \theta_0)$ *when* $n = 200$.

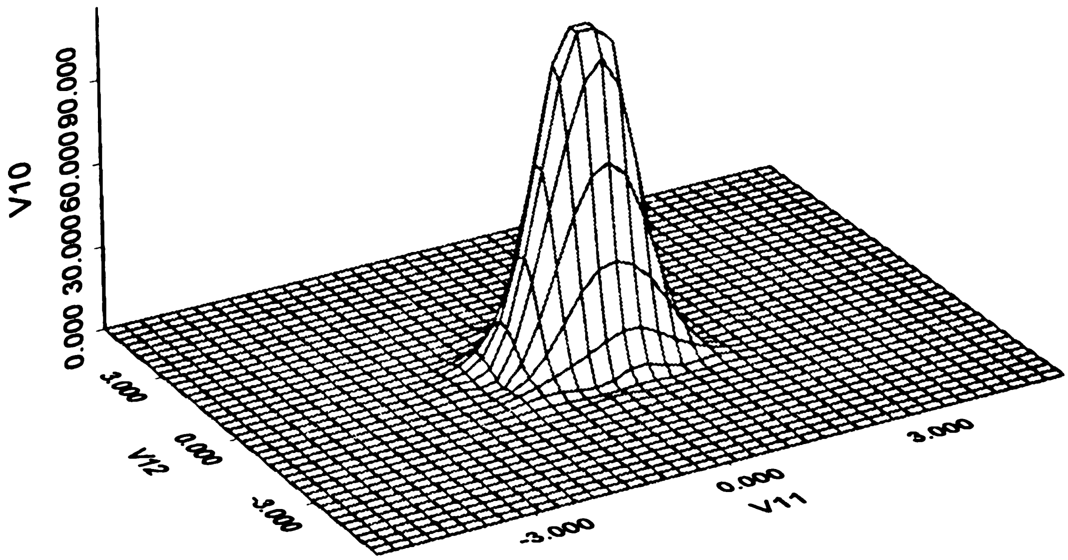
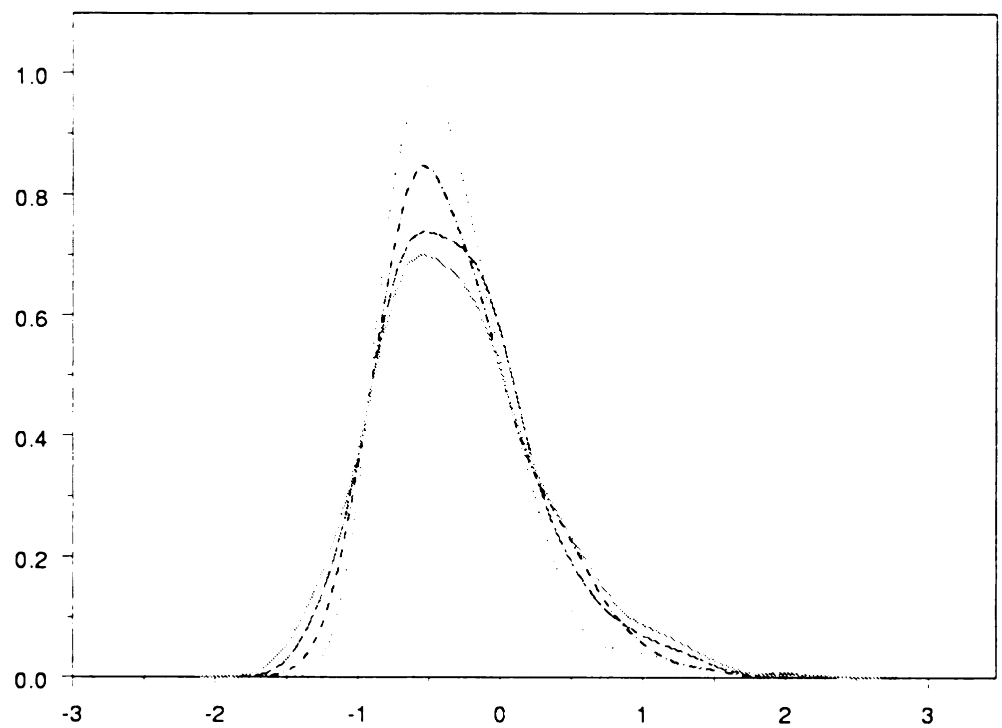


Figure 2.9: *The density of test statistics under H_0.*



Chapter 3

Minimum Distance Autoregressive Model Fitting

3.1 introduction

This chapter discusses application of minimum distance idea in fitting a parametric model to the autoregressive function. To be specific, let X_n be a real valued strictly stationary process having finite expectation. The autoregressive function is defined to be

$$\mu(x) = E(X_n | X_{n-1} = x), \ n \in \mathbb{Z}$$

Let $\{m_\theta(\cdot) : \theta \in \Theta\}, \Theta \subset \mathbb{R}^m$, Θ compact, be a given set of parametric functions. The statistical problem of interest here is to test the goodness-of-fit hypothesis

$$H_0 : \mu(x) = m_{\theta_0}(x), \quad \text{for some } \theta_0 \in \Theta, \text{ and for all } x \in \mathcal{I} \ vs. \ H_1 : H_0 \text{ is not true},$$

based on the sample $\{X_i : i \in \mathbb{Z},\}$ from the stochastic process, where $\mathcal{I}$ is a compact subset of $\mathbb{R}$.

In the context of regression fitting problem under the i.i.d set up, the asymptotic properties of minimum distance estimator of the parameter $\hat{\theta}_n$ are studied, where $\hat{\theta}_n$ is defined to be the argument that minimizes a transformation of the $L_2(G)$ distance between the nonparametric estimate of regression function μ and the parametric function m_θ. It has been shown that the so defined minimum distance estimator is consistent, asymptotically normally distributed with rate of $\sqrt{n}$. The corresponding minimized distance is also asymptotically normally distributed. Thus a class of tests can be constructed by using suitably standardized minimum distance. Encouraged by what have been shown in *i.i.d* case, we consider to apply the same idea to the autoregressive model checking.

when dealing with regression model fitting, to reduce the bias caused by $\hat{f}_h$ in $M_{hh}(\theta)$ defined in chapter 1, we used an optimal window width for the Nadaraya-Watson type estimation of f, i.e. $\hat{f}_{w_n}$. But it still causes bias. Hence in this chapter we consider using a slightly different L_2 distance defined as $M_h(\theta)$ of (1.0.8), which is actually the L_2-distance between $\widehat{m_\theta f}$ and a kernel estimator $m_\theta f$ defined as

$$\widehat{m_\theta f} = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})X_i,$$

where G is a σ-finite measure with bounded Lebesgue density g.

The estimate of the parametor is defined as (1.0.9). The test statistic T_n is defined to be

$$T_n = \frac{nh^{1/2}}{\hat{\Gamma}_n^2}\left(M_h(\theta_n) - \frac{1}{n^2}\sum_{i=1}^{n}\int_{cI} K_h^2(x - X_{i-1})dG(x)\hat{\varepsilon}_i^2\right),$$

where $\hat{\Gamma}_n^2$ is a consistent estimator of Γ,

$$\Gamma^2 = 4(\sigma^2)^2 \int_{\mathcal{I}} f^2(x)g^2(x)dx \int\left(\int K(u)K(u+v)du\right)^2 dv,$$

and $\hat{\varepsilon}_i = X_i - m_{\theta_n}(X_{i-1})$. Similar to the discussion in chapter 2, $\hat{\Gamma}_n^2$ can be chosen to be

$$4\int_{\mathcal{I}}\left(\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})\hat{\varepsilon}_i^2\right)^2 g^2(x)dx \int\left(\int K(u)K(u+v)du\right)^2 dv.$$

In this chapter a proof of the consistency of θ_n, the asymptotic normality of $\sqrt{n}(\theta_n - \theta_0)$, and asymptotic normality of the test statistic T_n are presented. A test of H_0 can be thus based on T_n.

3.2 Assumptions

Recall that the definition of "Geometrically Strongly Mixing" (GSM) from section 2.3 of Bosq (1998). $\{X_t\}$ is GSM if there exists $c_0 > 0$ and $\rho \in [0, 1)$ such that $\alpha(k) \le c_0\rho^k, k \ge 1$, where

$$\begin{aligned} \alpha(k) : &= \sup_{t\in\mathbb{Z}} \alpha(\sigma\{X_s, s \le t\}, \sigma\{X_s, s \ge t+k\}), \\ \alpha(\mathcal{A}, \mathcal{B}) : &= \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |P(A \bigcap B) - P(A)P(B)|, \end{aligned}$$

where $\sigma\{X_s, s \le t\}$ stands for the σ field generated by $\{X_s, s \le t\}$. It's also pointed out in Bosq (1998) that the usual linear processes are GSM.

Here we shall state the needed assumptions.

(M) The time series $\{X_i;\ X_i \in \mathbb{R},\ i \in \mathbb{Z}\}$, where $\mathbb{Z}$ stand for all integers, is strictly stationary satisfying GSM mixing condition and $X_i = \mu(X_{i-1}) + \varepsilon_i$.

McKeague and Zhang (1994) pointed out that it is easier to check geometric ergodicity, which implies strong mixing with a geometric mixing rate. From Tweedie (1983), one obtains that a sufficient (but by no means necessary) condition for geometric ergodicity of the nonlinear autoregressive process is that μ and σ are bounded on compact sets, where $\sigma^2 = E(\varepsilon_1^2|X_0)$.

About the errors and underlying design we assume the following:

(S1) The autoregressive function $\mu(\cdot)$ satisfies $\int \mu^2(x)dG(x) < \infty$, where G is a σ-finite measure on $\mathbb{R}$.

(S2) $\{\varepsilon_i\}$ are $i.i.d$ and ε_{i+1} is independent to X_j, $j = 0, .., i$, and $\sigma^2 := E\varepsilon_1^2$.

(S3) The density of X_0 is twice continuously differentiable Lebesgue density f that is bounded from below on $\mathcal{I}$. Denote the first and second derivatives of f by f' and f'', respectively. We also suppose that $\sup_{\{t_1,t_2,t_3,t_4\,t_5,t_6\}} \|f_{t_1,t_2,t_3,t_4,t_5,t_6}\|_\infty < \infty$, where $f_{t_1,t_2,t_3,t_4,t_5,t_6}$ is a joint density of X_{t_1}, X_{t_2}, X_{t_3}, X_{t_4}, X_{t_5}, and X_{t_6}.

About the kernel function K we shall assume the following:

Conditions (K), (A1), and (A2) are the same as those in chapter 2.

(A3) For each θ, $m_\theta(x)$ and $\dot{m}_{\theta_0}(x)$ are a.s continuous in x w.r.t the integrating measure G.

(A4) The function $\theta \mapsto m_\theta$ is continuous in $L_2(G)$: For any sequence $\theta_n, \theta \in \Theta$, $\|\theta_n - \theta\| \to 0$, implies $\rho(m_{\theta_n}, m_\theta) \to 0$.

(A5) For every $\varepsilon > 0$, there is an $N_\varepsilon < \infty$ such that for every $0 < k < \infty$,

$$\max_{1 \le i \le n, (nh)^{1/2}\|\theta - \theta_0\| \le k} h^{1/2} \|\dot{m}_\theta(X_i) - \dot{m}_{\theta_0}(X_i)\| = O_p(1).$$

About the bandwidth h we shall make the following assumptions:

(H) $h \sim n^{-a}$ for some $a > 0$, and there is a $\gamma > 0$ such that $nh^{2+\gamma} \longrightarrow \infty$.

In this chapter we will often use an inequality in Bosq (1998). We list it here as a lemma.

Lemma 3.2.1 *Let X and Y be real valued random variables such that $X \in L^q(P)$, $Y \in L^r(P)$, where $q, r > 1$ and $\frac{1}{q} + \frac{1}{r} = 1 - \frac{1}{p}$, then*

$$|Cov(X, Y)| \le 2p(2\alpha)^{1/p} \|X\|_q \|Y\|_r,$$

in particular

$$|Cov(X, Y)| \le 4\alpha \|X\|_\infty \|Y\|_\infty,$$

where

$$\alpha = \alpha(\sigma(X), \sigma(Y)),$$

$$\|X\|_\infty = \inf\{b : P(|X| > b) = 0\}.$$

Analog to the notations defined in section 2.3, we introduce some notations that will be needed in this chapter.

$$\begin{aligned} U_n(x,\theta) &= \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_{i-1})\left(X_i - m_\theta(X_{i-1})\right), \text{ and} \\ U_n(x) &= U_n(x,\theta_0) = \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_{i-1})\varepsilon_i, \end{aligned}$$

and $Z_n(x,\theta)$, $\mu_n(x,\theta)$, and $\dot{\mu}_n(x,\theta)$ are as defined in section 2.3 with X_i replaced by X_{i-1}. Note that

$$M_h(\theta_0) = \int_{\mathcal{I}} U_n^2(x)dG(x).$$

We also introduce the following notation,

$$\mathcal{J}_h := \{y \in \mathbb{R} : |x-y| \leq h, x \in \mathcal{I}\}.$$

3.3 Consistency of θ_n

The main result of this section is the consistency of θ_n. Similar to the proof of consistency of $\hat{\theta}_n$ in previous chapter, we will first prove the consistency of θ_n^* in Lemma 3.3.3, where now θ_n^* is defined to be

$$\begin{aligned} \theta_n^* &= \text{argmin}_{\theta\in\Theta} M_h^*(\theta), \quad \text{and} \\ M_h^*(\theta) : &= \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_{i-1})X_i - m_\theta(x)f(x)\right)^2 dG(x). \end{aligned}$$

This result will be in turn used to prove the consistency of θ_n in Theorem 3.1 .

Lemmas 3.3.1 and 3.3.2 list some results that will be needed in the proof of Lemma 3.3.3 and theorem.

Lemma 3.3.1 *Let ψ_n be a sequence of m-dimension vectors of real valued functions defined on $\mathbb{R}$, bounded on $\mathcal{J}_1$ uniformly in n. Then, under the condition (M), the following hold for $\forall x \in \mathcal{I}$ and $\forall\, 0 < a < 1$:*

$$(a)\ \left\| n^{-1} \sum_{i=1}^{n} (K_h(x - X_{i-1})\psi_n(X_{i-1}) - EK_h(x - X_0)\psi_n(X_0)) \right\| = o_p(\frac{1}{\sqrt{nh^{1+a}}})$$

$$(b)\ \left\| \int_{\mathcal{I}} n^{-1} \sum_{i=1}^{n} (K_h(x - X_{i-1})\psi_n(X_{i-1}) - EK_h(x - X_0)\psi_n(X_0))\, dG(x) \right\|$$
$$= O_p(\frac{1}{\sqrt{nh}}),$$

$$(c)\ E \int_{\mathcal{I}} \left\| n^{-1} \sum_{i=1}^{n} K_h(x - X_{i-1})\psi_n(X_{i-1}) - EK_h(x - X_0)\psi_n(X_0) \right\|^2 dG(x)$$
$$= o(\frac{1}{nh^{1+a}}),$$

where $\|\ \|$ stands for the usual L_2 norm defined on $\mathbb{R}^m$, i.e.

$$\|(a_1, ..., a_m)^T\| = \sqrt{a_1^2 + \cdots + a_m^2},\ \forall\, (a_1, ..., a_m)^T \in \mathbb{R}^m.$$

Proof. Note that the lemma holds for $\{\psi_n\}$ if and only if it holds for all j^{th} component of $\{\psi_n\}$, $1 \le j \le m$. Hence we only need to prove the lemma for the case of $m = 1$. Recall m is the dimension of Θ.

Let $a < b < 1$. For an $x \in \mathcal{I}$, let

$$\phi_n(X_i) := K_h(x - X_i)\psi_n(X_i) - EK_h(x - X_i)\psi_n(X_i).$$

Then $E\phi_n(X_i) = 0$. So

$$(3.3.1) \qquad E\left(n^{-1} \sum_{i=1}^{n} (K_h(x - X_{i-1})\psi_n(X_{i-1}) - EK_h(x - X_0)\psi_n(X_0)) \right)^2$$
$$= n^{-1} E\phi_n^2(X_0) + 2n^{-2} \sum_{i<j}^{n} \mathrm{Cov}(\phi_n(X_{i-1}), \phi_n(X_{j-1}))$$

The first term is $O(\frac{1}{nh})$ by the fact that the variance is bounded above by its second moment, boundedness of ψ_n and f. By Lemma 3.2.1, the second term is bounded above by

$$
\begin{aligned}
(3.3.2) \qquad & 2n^{-2}\sum_{i<j}^{n} 2b^{-1}(2\alpha(j-i))^{b}\|\phi_n(X_{i-1})\|_q\|\phi_n(X_{j-1})\|_r \\
= \; & 2^{2+b}n^{-2}\sum_{k=1}^{n-1}\sum_{i=1}^{n-k}\alpha^{b}(k)\|\phi_n(X_{i-1})\|_q\|\phi_n(X_{i+k-1})\|_r \\
\leq \; & 2^{2+b}n^{-1}\sum_{k=1}^{n-1}\alpha^{b}(k)\|\phi_n(X_0)\|_q^2,
\end{aligned}
$$

by taking $q = r = 2/(1-b)$ and $p = 1/b$. But note that

$$\|\phi_n(X_0)\|_q^2 = (E|\phi_n(X_0)|^q)^{2/q} = O(h^{-2(q-1)/q}) = O(h^{-(1+b)}) = o(h^{-(1+a)}).$$

Hence (3.3.1) is $o(\frac{1}{nh^{1+a}})$. Consequently (a) holds. Similarly, one may prove (b) and (c). □

Corollary 3.3.1 *Let $\psi(x)$ be a real valued continuous function on $\mathbb{R}$. Then under the conditions (M), (S1), (S2), and (S3),*

$$n^{-1}\sum_{i=1}^{n} K_h(x - X_{i-1})\psi(X_{i-1}) \longrightarrow \psi(x)f(x), \textit{ in probability}, \forall x \in \mathcal{I}.$$

Proof. Note that by the continuity of f and ψ,

$$EK_h(x - X_i)\psi(X_i) = \int K(u)\psi(x-uh)f(x-uh)du \longrightarrow \psi(x)f(x),$$

so the corollary follows by applying Lemma 3.3.1 (a) to $\psi_n = \psi$. □

Lemma 3.3.2 *Under the conditions (S1), (S2), and (K),*

$$E\int_{\mathcal{I}} Z_n^2(x, \theta_n)dG(x) \longrightarrow 0.$$

Proof. By adding and subtracting $K_h(x-X_{i-1})X_i$ to the i^{th} summand in $Z_n(x,\theta_n)$, and expanding the quadratic term, one obtains

$$\int_{\mathcal{I}} Z_n^2(x,\theta_n)dG(x) \le 2M_h(\theta_n) + 2M_h(\theta_0) \le 4M_h(\theta_0).$$

The second inequality follows from the definition of θ_n. Therefore, to prove the lemma it suffices to show that $EM_h(\theta_0) \to 0$. Note that by Fubini,

$$\begin{aligned} &EM_h(\theta_0)\\ &= \frac{\sigma^2}{n}\int_{\mathcal{I}} EK_h^2(x-X_0)dG(x) + \frac{2}{n}\sum_{i<j}^{n} E\int_{\mathcal{I}} K_h(x-X_{i-1})K_h(x-X_{j-1})dG(x)\varepsilon_i\varepsilon_j. \end{aligned}$$

The first term is $O((nh)^{-1})$ by direct calculation. The second term is 0 by taking conditional expectation on $\sigma\{X_s : s \le j\}$ first. Hence

$$EM_h(\theta_0) = E\int_{\mathcal{I}} U_n^2(x)dG(x) = O((nh)^{-1}). \tag{3.3.3}$$

So the lemma is proved. □

Lemma 3.3.3 *Under the conditions (S1), (S2), (S3), (K), (A1), and (A4),*

$$\theta_n^* \longrightarrow \theta_0, \quad \textit{in probability under } H_0.$$

Proof. The proof is similar to Corollary 3.1 in chapter 2. According to Lemma 3.1 in chapter 2, it suffices to show that

$$\int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x-X_{i-1})X_i - m_{\theta_0}(x)f(x)\right)^2 dG(x) = o_p(1). \tag{3.3.4}$$

Note that by plugging in $X_i = m_{\theta_0}(X_{i-1}) + \varepsilon_i$, adding and subtracting $EK_h(x-X_{i-1})m_{\theta_0}(X_{i-1})$ in the i^{th} summand of the integrand, the left hand side of

(3.3.4) is bounded above by the sum of the following three terms:

$$(a) \qquad \int_{\mathcal{I}} U_n^2(x)dG(x)$$

$$(b) \qquad \int_{\mathcal{I}} \left(\frac{1}{n} \sum_{i=1}^{n} [K_h(x - X_{i-1}) m_{\theta_0}(X_{i-1}) - EK_h(x - X_{i-1}) m_{\theta_0}(X_{i-1})] \right)^2 dG(x)$$

$$(c) \qquad \int_{\mathcal{I}} (EK_h(x - X_0) m_{\theta_0}(X_0) - m_{\theta_0}(x) f(x))^2 \, dG(x).$$

The term (a) is $O_p(1/(nh))$ by (3.3.3). The term (b) is $o_p(1)$ by Lemma 3.3.1 (c) with $\psi_n = m_{\theta_0}$. The term (c) is $o(1)$ because it is equal to

$$\int_{\mathcal{I}} \left(\int K(u)(m_{\theta_0}(x - uh) f(x - uh) - m_{\theta_0}(x) f(x)) du \right)^2 dG(x) = o(1)$$

by continuity of m_{θ_0} and f, compactness of $\mathcal{I}$. Hence (3.3.4) holds, so does the lemma. □

Now we are ready to present the main theorem of this section.

Theorem 3.3.1 *Under the conditions (M), (S1), (S2), (S3), (K), (A1), and (A4),*

$$\theta_n \longrightarrow \theta_0, \qquad \textit{in probability under } H_0.$$

Proof. The proof of this theorem is similar to Theorem 3.1 in chapter 2. Here we only sketch the proof. Recall the definition of ρ from section 2.2 and note that $M_h^*(\theta) = \rho(r_n, m_\theta f)$, where $r_n(x) := n^{-1} \sum_{i=1}^n K_h(x - X_{i-1}) X_i$.

By the same argument as in the proof of Theorem 2.3.1 with M_{hw}^* and M_{hw} replaced by M_h^* and M_h, it suffices to prove the following result,

$$\sup_{\theta \in \Theta} \left| M_h(\theta) - M_h^*(\theta) \right| = o_p(1). \tag{3.3.5}$$

To prove (3.3.5), add and subtract $n^{-1}\sum_{i=1}^{n} K_h(x - X_{i-1})m_\theta(X_{i-1})$ inside the parenthesis of $M_h^*(\theta)$, expand the quadratic, and use the Cauchy-Schwarz inequality on the cross product, to obtain that the left hand side of (3.3.5) is bounded above by

$$\sup_{\theta\in\Theta} C_n(\theta) + 2\sup_{\theta\in\Theta}(C_n(\theta)M_h(\theta))^{1/2},$$

where

$$C_n(\theta) = \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})(m_\theta(X_{i-1}) - m_\theta(x)) \right)^2 dG(x).$$

To prove (3.3.5), it suffices to prove that

$$(a)\ \sup_{\theta\in\Theta} C_n(\theta) = o_p(1), \quad \text{and} \quad (b)\ \sup_{\theta\in\Theta} M_h(\theta) = O_p(1).$$

First to prove (a). Note that $K_h(x - X_i)$ is nonzero only if $X_i \in \mathcal{J}_1$ for large n such that $h \leq 1$, so $C_n(\theta)$ is bounded above by

$$\sup_{|y-x|\leq h, x,y\in\mathcal{J}_1} \left| m_\theta(y) - m_\theta(x) \right|^2 \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1}) \right)^2 dG(x).$$

As a consequence of Lemma 3.3.1 part (c) with $\psi_n = 1$,

$$\int_{\mathcal{I}} \left(n^{-1}\sum_{i=1}^{n} K_h(x - X_{i-1}) \right)^2 dG(x) = O_p(1). \tag{3.3.6}$$

And

$$\sup_{\theta\in\Theta} \sup_{|y-x|\leq h, x,y\in\mathcal{J}_1} \left| m_\theta(y) - m_\theta(x) \right|^2 = o(1)$$

because of the continuity of m and compactness of Θ and $\mathcal{J}_1$. Hence

$$\sup_{\theta\in\Theta} C_n(\theta) = o_p(1).$$

Next to prove (b). By plugging in $X_i = m_{\theta_0}(X_{i-1}) + \varepsilon_i$, one obtains that $M_h(\theta)$ is bounded above by:

$$2\int_{\mathcal{I}} U_n^2(x)dG(x) + 2\int_{\mathcal{I}} Z_n^2(x,\theta)dG(x).$$

The first term is $o_p(1)$ by (3.3.3). For large n such that $h \leq 1$, the second term is bounded above by

$$4 \sup_{\theta\in\Theta, y\in\mathcal{J}_1} m_\theta^2(y) \cdot \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})\right)^2 dG(x) = O_p(1)$$

by the continuity of m, the compactness of Θ and $\mathcal{J}_1$, and (3.3.6). Hence

$$\sup_{\theta\in\Theta} M_h(\theta) = O_p(1).$$

So (3.3.5) is proved, so is the theorem. □

3.4 Asymptotic distribution of $\sqrt{n}(\theta_n - \theta_0)$.

In this section we will prove the asymptotic normality of θ_n. Before that we introduce some notations that are going to be used in this section. Define

$$\begin{aligned}(3.4.1)\qquad \xi_n(x) &= EK_h(x - X_0)\dot{m}_{\theta_0}(X_0),\\ \eta_n(x) &= m_{\theta_n}(x) - m_{\theta_0}(x) - \dot{m}_{\theta_0}^T(x)(\theta_n - \theta_0),\end{aligned}$$

and $\xi(x)$, η^2 are as defined in (1.0.10) of Chapter 1.

Note that under the condition (M), θ_n is a solution to the equation $\partial M_h(\theta)/\partial\theta = 0$. i.e.

$$\int_{\mathcal{I}} U_n(x,\theta_n)\dot{\mu}_n(x,\theta_n)dG(x) = 0.$$

Plug in $X_i = m_{\theta_0}(X_{i-1}) + \varepsilon_i$ in $U_n(x, \theta_n)$ and rewrite the above equation in the following form:

$$(3.4.2) \qquad \int_{\mathcal{I}} U_n(x)\dot{\mu}_n(x, \theta_n)dG(x) = \int_{\mathcal{I}} Z_n(x, \theta_n)\dot{\mu}_n(x, \theta_n)dG(x)$$

As in the proof of Theorem 2.4.1, we will use Lemmas 3.4.1 and 3.4.2 below to show that the left hand side of the above equation is approximated by an average of martingale differences. Hence, by the martingale central limit theorem (M.G.C.L.T) converges in distribution to a normal random variable with rate $1/\sqrt{n}$. The right hand side can be written as $(\theta_n - \theta_0)$ times a random variable which, by Lemma 3.4.3, converges in probability to a positive constant. So the theorem about the asymptotic normality of θ_n follows.

Now we start with three lemmas.

Lemma 3.4.1 *Under the conditions (M), (S1), (S2), S(3), (K), (A1), and (A4), there is a function ξ such that the following hold:*

$$(a) \sup_{x \in \mathcal{I}} \left\| \dot{\mu}_n(x, \theta_0) - \xi(x) \right\| = o_p(1),$$

$$(b) \sup_{x \in \mathcal{I}} \left\| \dot{\mu}_n(x, \theta_n) - \xi(x) \right\| = o_p(1).$$

Lemma 3.4.2 *Let l be a real valued continuous function on $\mathcal{I}$. Under the conditions (M), (S1), (S2), (S3), (K), (A1), (A3), and (A4),*

$$\sqrt{n} \int \left(\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_{i-1}) l(X_{i-1}) \varepsilon_i \xi_n(x) \right) dG(x)$$

converges in distribution to a normal random vector with mean zero and covariance matrix given by

$$\sigma^2 \int_{\mathcal{I}} \dot{m}_{\theta_0}(x)\dot{m}_{\theta_0}^T(x)l^2(x)g^2(x)f(x)dx.$$

The result is also true when ξ_n are replaced by ξ.

Lemma 3.4.3 *Under the conditions (M), (S1), (S2), (S3), (K), (A1), (A2), (A3), and (A4),*

$$(3.4.3)\, \|\theta_n - \theta_0\|^{-1} \int \left(\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_{i-1})\eta_n(X_{i-1}) \right) \xi(x)dG(x) = o_p(1).$$

We will state the main theorem of this section.

Theorem 3.4.1 *Under the conditions (M), (S1), (S2), (S3), (K), (A1), (A2), (A3), and (A4), $\sqrt{n}(\theta_n - \theta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix $\Sigma_0^{-1}\boldsymbol{\eta}^2\Sigma_0^{-1}$, where Σ_0, $\boldsymbol{\eta}^2$ are as defined in (1.0.10).*

Proof. Note that the right hand side of (3.4.2) can be written as $(\theta_n - \theta_0)R_n$, where R_n is a sum of following terms:

$$\begin{aligned} R_{n1} &= \int_{\mathcal{I}} \dot{\mu}_n(x, \theta_0)\dot{\mu}_n^T(x, \theta_n)dG(x) \\ R_{n2} &= \int_{\mathcal{I}} \left(\frac{1}{n} \sum_{i=1}^{n} K_h(x - X_{i-1}) \frac{\eta_n(X_{i-1})}{\|\theta_n - \theta_0\|} \right) \dot{\mu}_n(x, \theta_n)dG(x). \end{aligned}$$

By Lemma (3.4.1), $R_{n1} = \int_{\mathcal{I}} \xi(x)\xi^T(x)dG(x) + o_p(1)$. By Lemma (3.4.3) and Lemma (3.4.1), $R_{n2} = o_p(1)$. Hence R_n converges in probability to Σ_0.

Note that by adding and subtracting $K_h(x - X_{i-1})\xi(x)$ to the i^{th} summand, the left hand side of (3.4.2) can be written as a sum of the following two terms:

$$\begin{aligned} L_{n1}^1 &= \int_{\mathcal{I}} U_n(x)\xi(x)dG(x), \\ L_{n2}^1 &= \int_{\mathcal{I}} U_n(x)\left[\dot{\mu}_n(x,\theta_n) - \xi(x)\right]dG(x). \end{aligned}$$

By Lemma 3.4.2 with $l = 1$ and $\xi_n = \xi$, $\sqrt{n}L_{n1}^1$ converges in distribution to a normal random vector. By Lemma 3.4.1 and (3.3.3), the term $L_{n2}^1 = o_p((nh)^{-1/2})$.

So the left hand side of (3.4.2) is $o_p((nh)^{-1/2})$ and the right hand side of (3.4.2) is $(\theta_n - \theta_0)R_n$ where R_n converges to Σ in probability. Hence

$$(\theta_n - \theta_0) = o_p((nh)^{-1/2}). \tag{3.4.4}$$

Next we shall show that $(\theta_n - \theta_0)$ is actually $O_p(1/\sqrt{n})$. Note that by adding and subtracting $K_h(x - X_{i-1})\dot{m}_{\theta_0}(X_{i-1}) - \xi_n(x)$ to the i^{th} summand, the left hand side of (3.4.2) can also be written as the sum of the following three terms:

$$\begin{aligned} L_{n1}^2 &= \int_{\mathcal{I}} U_n(x)\xi_n(x)dG(x), \\ L_{n2}^2 &= \int_{\mathcal{I}} U_n(x)\left(\dot{\mu}_n(x,\theta_0) - \xi_n(x)\right)dG(x), \\ L_{n3}^2 &= \int_{\mathcal{I}} U_n(x)Z_n(x,\theta_n)dG(x). \end{aligned}$$

By Lemma 3.4.2 with $l = 1$, $\sqrt{n}L_{n1}^2$ converges in distribution to a normal random vector. The term $L_{n2}^2 = o_p(1/\sqrt{n})$ by the Cauchy-Schwarz inequality, Fubini, Lemma 3.3.1 (c), and (3.3.3). The term $L_{n3}^2 = o_p(1/\sqrt{n})$ is by (3.4.4), (3.3.3), and the assumption (A5).

Combine the above discussion to conclude that

$$\sqrt{n}(\theta_n - \theta_0)R_n = \sqrt{n}L_{n2}^1 + o_p(1).$$

Hence the Theorem holds by Lemma 3.4.2. □.

Next we are going to prove the three lemmas.

Proof of Lemma 3.4.1.

We will prove a bit more general form of this lemma. i.e. for any continuous function l on $\mathcal{I}$,

$$(3.4.5) \qquad \sup_{x\in\mathcal{I}} \Big|\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})l(X_{i-1}) - l(x)f(x)\Big| \longrightarrow 0$$

where f is the density function of X_0.

Because $l(x)f(x)$ is continuous on compact set $\mathcal{I}$, so it is bounded on $\mathcal{I}$. So

$$\begin{aligned}
& \sup_{x\in\mathcal{I}} \Big|E\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})l(X_{i-1}) - l(x)f(x)\Big| \\
= & \sup_{x\in\mathcal{I}} \Big|\int K(u)[l(x - uh)f(x - uh) - l(x)f(x)]du\Big| \\
\le & \sup_{|y-x|\le h, x\in\mathcal{I}, y\in\mathcal{J}_1} \Big|l(y)f(y) - l(x)f(x)\Big| \longrightarrow 0.
\end{aligned}$$

In order to complete the proof of (3.4.5), we still need to show that

$$(3.4.6) \quad \sup_{x\in\mathcal{I}} \Big|\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})l(X_{i-1}) - EK_h(x - X_0)l(X_0)\Big| = o_p(1).$$

Let $\zeta_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_{i-1})l(X_{i-1})$. Consider covering compact set $\mathcal{I} \in B = \{x \in \mathbb{R} : |x| \le b\}$ for some $b < \infty$ by ν_n closed set: $B_{jn} = \{x : |x - x_{jn}| < b/\nu_n\}$, where $1 \le j \le \nu_n$ such that $B_{jn}^0 \bigcap B_{kn}^0 = \phi$ for $j \ne k$.

By the assumption that K is Lipschitz, there is a finite positive number β such that

$$\begin{aligned}
\left|\zeta_n(x)-\zeta_n(x_{jn})\right| &\leq \frac{\beta}{h^2}|x-x_{jn}|\frac{1}{n}\sum_{i=1}^{n}|l(X_{i-1})| \leq \frac{b\beta}{h^2\nu_n}\frac{1}{n}\sum_{i=1}^{n}|l(X_{i-1})|, \\
\left|E\zeta_n(x)-E\zeta_n(x_{jn})\right| &\leq \frac{b\beta}{h^2\nu_n}E|l(X_0)|, \quad \forall x\in B_{jn}, \quad 1\leq j\leq n.
\end{aligned}$$

So

(3.4.7)

$$\begin{aligned}
&\sup_{x\in\mathcal{I}}\left|\zeta_n(x)-E\zeta_n(x)\right| \\
&\quad\leq \sup_{1\leq j\leq\nu_n}\sup_{x\in B_{jn}}\left|\zeta_n(x)-E\zeta_n(x)\right| \\
&\quad\leq \sup_{1\leq j\leq\nu_n}\sup_{x\in B_{jn}}\left(|\zeta_n(x)-\zeta_n(x_{jn})|+|\zeta_n(x_{jn})-E\zeta_n(x_{jn})|+|E\zeta_n(x_{jn})-E\zeta_n(x)|\right) \\
&\quad= \frac{2b\beta}{h^2\nu_n}O_p(1)+\sup_{1\leq j\leq\nu_n}\left|\zeta_n(x_{jn})-E\zeta_n(x_{jn})\right|.
\end{aligned}$$

Note that $|K_h(x-X_{i-1})l(X_{i-1})|$ is zero unless $X_{i-1}\in\mathcal{J}_1$ for large n such that $h\leq 1$, and it is bounded by $c/(2h)$ when $X_{i-1}\in\mathcal{J}_1$ for some constant c. By Theorem 1.3, part (1) of Bosq (1998), for any $1<q<(n/2)$,

$$P(|\zeta_n(x_{jn})-E\zeta_n(x_{jn})|>\varepsilon)\leq 4\,exp\left(-\frac{\varepsilon^2}{8c^2}\cdot qh^2\right)+22q\left(1+\frac{4c}{\varepsilon}\frac{1}{h}\right)^{1/2}\alpha[\frac{n}{2q}].$$

Choose $\nu_n=n$, and $q=\sqrt{n}/h$, then

$$\begin{aligned}
&P(\sup_{1\leq j\leq\nu_n}|\zeta_n(x_{jn})-E\zeta_n(x_{jn})|>\varepsilon) \\
&\leq 4\nu_n exp\left(-\frac{\varepsilon^2}{8c^2}\cdot nh^{2+\gamma}\right)+22\nu_n q\left(1+\frac{4c}{\varepsilon}\frac{1}{h}\right)^{1/2}\alpha[h^{-\gamma}/2] \\
&\leq c_1\,ne^{-c_2\sqrt{nh^2}}+c_3n^{3/2}h^{-1}\rho_0^{-\sqrt{nh^2}/2}=o(1),
\end{aligned}$$

for some positive constants c_1, c_2, and c_3 by conditions (M) and (H) . Hence (3.4.7) is $o_p(1)$, so is (3.4.6). this also completes the proof of (3.4.5).

By taking $l = m_{\theta_0}$ in (3.4.5), then part (a) of the lemma is proved. To prove part (b) of the lemma, it suffices to prove that

$$\sup_{x\in\mathcal{I}} \left\|\dot{\mu}(x,\theta_n) - \dot{\mu}(x,\theta_0)\right\| = o_p(1). \tag{3.4.8}$$

Because for large n such that $h \leq 1$, $K_h(x - X_{i-1})$ is nonzero only if $X_{i-1} \in \mathcal{J}_1$, so by the continuity of $\dot{m}_\theta$, compactness of Θ and $\mathcal{J}_1$, and the consistency of θ_n,

$$\sup_{y\in\mathcal{J}_1} \|\dot{m}_{\theta_n}(y) - \dot{m}_{\theta_0}(y)\| = o_p(1).$$

Apply (3.4.5) with $l(x) = 1$, one obtains that

$$\sup_{x\in\mathcal{I}} n^{-1}\sum_{i=1}^{n} K_h(x - X_{i-1}) = O_p(1).$$

Hence (3.4.8) is bounded above by

$$\sup_{y\in\mathcal{J}_1} \|\dot{m}_{\theta_n}(y) - \dot{m}_{\theta_0}(y)\| \sup_{x\in\mathcal{I}} n^{-1}\sum_{i=1}^{n} K_h(x - X_{i-1}) = o_p(1).$$

That completes the proof of the part (b) of the lemma. □

We are going to apply the Martingale Central Limit theorem, i.e. Corollary 3.1 of Hall and Heyde (1989) to prove Lemma 3.4.2. For the sake of completeness, we state the corollary here as a lemma:

Lemma 3.4.4 *Suppose $S_{nk_n} = \sum_{i=1}^{n} X_{ni}$, and $(S_{ni}, \mathcal{F}_{n,i})$ is a zero-mean, square integrable martingale array with differences X_{ni}, and η^2 is an a.s finite random* *variable. If $\{X_{ni}\}$ satisfy the following conditions:*

$$(a) \qquad \forall \varepsilon > 0, \sum_{i=1}^{n} E[X_{ni}^2 I_{\{|X_{ni}|>\varepsilon\}}|\mathcal{F}_{n,i-1}] \to 0, \text{ in probability.}$$

$$(b) \qquad V_n^2 = \sum_{i=1}^{n} E(X_{ni}^2|\mathcal{F}_{n,i-1}) \to \eta^2, \text{ in probability.}$$

$$(c) \qquad \mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}, \text{ for } i \leq i \leq k_n, n \geq 1.$$

Then, S_{nk_n} converges in distribution to a normal random variable with mean zero and variance η^2.

Proof of Lemma 3.4.2. W.L.O.G, here only gives the proof for the case that θ is one dimension. We will construct a martingale array and verify the three conditions of the Lemma 3.4.4. Define

$$\begin{aligned} S_{nj} &= \sum_{i=0}^{j} n^{-1/2} \int_{\mathcal{I}} K_h(x - X_{i-1}) l(X_{i-1}) \xi_n(x) dG(x) \varepsilon_i, \\ \mathcal{F}_{n,j} &= \sigma\{X_0, X_1, ..., X_j, \varepsilon_1, ..., \varepsilon_j\}. \end{aligned}$$

Then $\{S_{ni}, \mathcal{F}_{n,i}\}$ is a zero mean, square integrable martingale array, and $\mathcal{F}_{n,i} \subseteq \mathcal{F}_{n+1,i}$, $X_{ni} = n^{-1/2} \int_{\mathcal{I}} K_h(x - X_{i-1}) l(X_{i-1}) \xi_n(x) dG(x) \varepsilon_i$. So the condition (c) holds.

For any $\lambda > 0$ and $c > 0$,

$$(3.4.9) \qquad \sum_{i=1}^{n} E[X_{ni}^2 I\{|X_{ni}| > \lambda\}|\mathcal{F}_{n,i-1}] \leq \frac{1}{\lambda^c} \sum_{i=1}^{n} E[|X_{ni}|^{2+c}|\mathcal{F}_{n,i-1}].$$

Because $\xi_n(x) = EK_h(x - X_0)\dot{m}_{\theta_0}(X_0) = \int K(u)\dot{m}_{\theta_0}(x - uh)f(x - uh)du$, the kernel function K has bounded support, and continuity of $\dot{m}_{\theta_0}$ and f, so ξ_n is bounded uniformly in n at $x \in \mathcal{I}$, and suppose the bound is B_ξ. Furthermore, note that

(3.4.10)

$$\int_{\mathcal{I}} K_h(x - X_i)\|\xi_n(x)\|dG(x) \leq B_\xi \int K(u)g(X_i + uh)du \leq B_\xi \sup_y g(y) < \infty.$$

So by stationary of X_i and definition of $\mathcal{F}_{n,i-1}$, (3.4.9) is bounded above by

$$\begin{aligned} & \lambda^{-c} n^{-c/2}(B_\xi \sup_y g(y))^{2+c} \frac{1}{n}\sum_{i=1}^n E[|\varepsilon_i|^{2+c}|\mathcal{F}_{n,i-1}] \\ = \; & \lambda^{-c} n^{-c/2} C = o_p(1), \end{aligned}$$

for some constant C. Hence the condition (a) holds.

For the condition (b), note that

$$\begin{aligned} EV_n^2 &= \frac{1}{n}\sum_{i=1}^n E\left((\int_{\mathcal{I}} K_h(x - X_{i-1})l(X_{i-1})\xi_n(x)dG(x))^2\sigma^2 \right) \\ &= E(\int_{\mathcal{I}} K_h(x - X_0)l(X_0)\xi_n(x)dG(x))^2\sigma^2 \\ &= \int_{\mathcal{I}}\int_{\mathcal{I}} (EK_h(x - X_0)K_h(y - X_0)l^2(X_0))\, \xi_n(x)\xi_n(y)dG(x)dG(y)\cdot\sigma^2 \\ &\longrightarrow \sigma^2 \int_{\mathcal{I}} \dot{m}_{\theta_0}^2(x)l^2(x)g^2(x)f^3(x)dx. \end{aligned}$$

Let V_{ni} denote $(\int_{\mathcal{I}} K_h(x - X_i)l(X_{i-1})\xi_n(x)dG(x))^2\sigma^2$. Note that, V_{ni} is bounded uniformly in ni. Then

(3.4.11)

$$\begin{aligned} Var(V_n^2) &= E\left(\frac{1}{n}\sum_{i=1}^n (V_{ni} - EV_{ni})\right)^2 \\ &= \frac{1}{n}Var(V_{n1}) + \frac{2}{n^2}\sum_{i<j}^n \mathrm{Cov}(V_{ni}, V_{nj}). \end{aligned}$$

By (3.4.10), and the fact that variance is bounded above by its second moment, the first term on the right hand side of the last equality sign of (3.4.11) is bounded above by $c\,(nh)^{-1}$ for some constant c. Hence it converges to zero.

By Lemma 3.2.1 and (3.4.10), the second term on the right hand side of the last equality sign of (3.4.11) is bounded above by

$$\frac{1}{n^2}\sum_{k=1}^{n-1}\sum_{j-i=k} 4\|V_{ni}\|_\infty^2 \alpha(k) \le C_1\frac{1}{n} = O(\frac{1}{n})$$

for some constant C_1. So (3.4.11) tends to zero. This proves that condition (b) holds. So by Lemma 3.4.4, $\sqrt{n}\int\left(\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)l(X_i)\varepsilon_{i+1}\right)\xi_n(x)dG(x)$ converges in distribution to a normal random variable with mean zero and variance given by

$$\sigma^2\int_{\mathcal{I}} \dot{m}_{\theta_0}^2(x)l^2(x)g^2(x)f^3(x)dx,$$

in particular when $l = 1$, the variance is

$$\eta^2 = \sigma^2\int_{\mathcal{I}} \dot{m}_{\theta_0}^2(x)g^2(x)f^3(x)dx.$$

$\square$

Proof of Lemma 3.4.3.

By (A2) and consistency of θ_n, one obtains

$$\max_{0\le i\le n-1}|\eta_n(X_i)|/\|\theta_n-\theta_0\| = o_p(1). \tag{3.4.12}$$

Similar to the proof of (3.4.10), $\int_{\mathcal{I}} K_h(x-X_i)|\xi(x)|dG(x)$ are bounded uniformly in i and $h \le 1$, hence

$$\int_{\mathcal{I}} K_h(x-X_i)|\xi(x)|dG(x)|\eta_n(X_i)|/\|\theta_n-\theta_0\| = o_p(1) \text{ uniform in } i.$$

So (3.4.3) is also a $o_p(1)$. $\square$

3.5 Asymptotic behavior of the minimum distance.

In chapter 2, it has been proved that the standardized minimum distance is asymptotically normally distributed with rate $nh^{1/2}$ under the *i.i.d* setup. In this section we will show that the same result is also true if the observations are from a stochastic process satisfying a GSM condition. This result can be seen from the following three propositions.

Before present the propositions, Define

$$\hat{\varepsilon}_i = X_i - m_{\theta_n}(X_{i-1}) \quad i = 1, ..., n.$$

Proposition 3.5.1 *Under the conditions (M) to (H),*

$$M_h(\theta_n) - M_h(\theta_0) = o_p(\frac{1}{nh^{1/2}}).$$

Proposition 3.5.2 *Under the conditions (M) to (H),*

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{I}} K_h^2(x - X_{i-1})dG(x)(\hat{\varepsilon}_i^2 - \varepsilon_i^2) = o_p(\frac{1}{nh^{1/2}}).$$

Proposition 3.5.3 *Under the conditions (M) to (H),*

$$n^{-2}\sum_{i<j}^{n}\int_{\mathcal{I}} K_h(x - X_{i-1})K_h(x - X_{j-1})dG(x)\varepsilon_i\varepsilon_j$$

is asymptotically normally distributed with mean zero and asymptotic variance σ_n^2, where σ_n^2 is specified in the condition (C1) of the proof of Proposition 3.5.3.

A natural consequence of these three propositions is the following theorem.

Theorem 3.5.1

$$\tilde{T}_n = \frac{nh^{1/2}}{\Gamma^2}\left(M_h(\theta_n) - \frac{1}{n^2}\sum_{i=1}^n \int_{cI} K_h^2(x - X_{i-1})dG(x)\hat{\varepsilon}_i^2\right)$$

converges to a standard normal random variable as n *tends to infinity, where*

$$\Gamma^2 = 4(\sigma^2)^2 \int_I f^2(x)g^2(x)dx \int \left(\int K(u)K(u+v)du\right)^2 dv.$$

Similar to the discussion in chapter 2, one obtains that Γ_n^2 defined in the introduction section of this chapter is a consistent estimator of Γ^2. Hence based on this theorem, we can therefore conduct the goodness-of-fit test by using T_n as a test statistic.

Next we will prove the propositions and the theorems. But before that we will present some lemmas first.

Lemma 3.5.1 *Under the conditions (M) to (H),*

$$\int_I \left(n^{-1}\sum_{i=1}^n K_h(x - X_{i-1})\eta_n(X_{i-1})/\|\theta_n - \theta_0\|\right)^2 dG(x) = o_p(1).$$

Proof. Let $|\eta_n(X_i)|/\|\theta_n - \theta_0\| = \gamma_{ni}$. Note that the left hand side of the lemma can be expanded as a sum of two terms:

$$(3.5.1) \qquad \begin{aligned} &\frac{1}{n^2}\sum_{i=1}^n \int_I K_h^2(x - X_{i-1})\gamma_{n(i-1)}^2 dG(x), \text{ and} \\ &\frac{2}{n^2}\sum_{i<j}^n \int_I K_h(x - X_{i-1})K_h(x - X_{j-1})dG(x)\gamma_{n(i-1)}\gamma_{n(j-1)}. \end{aligned}$$

The first term of (3.5.1) is equal to

$$\frac{1}{nh}\cdot\frac{1}{n}\sum_{i=1}^n \int K^2(u)g(X_{i-1} + uh)du\, \gamma_{n(i-1)}^2 = o_p(\frac{1}{nh})$$

by the boundedness of g and (3.4.12). Note that

$$\begin{aligned}
& E\int_{\mathcal{I}} K_h(x-X_i)K_h(x-X_j)dG(x) \\
= \;& \int_{\mathcal{I}}\int K(u)K(v)f_{ij}(x-uh, x-vh)dudvdG(x) \\
\leq \;& \sup_{x,y\in\mathcal{J}_1} f_{ij}(x,y)\, G(\mathcal{I}) < \infty.
\end{aligned}$$

Hence the second term of (3.5.1) is $o_p(1)$. This completes the proof. □

Lemma 3.5.2 *Under the conditions (M) to (H),*

$$\int_{\mathcal{I}} Z_n^2(x,\theta_n)dG(x) = O_p(n^{-1}).$$

Proof. By plugging in $m_{\theta_n}(X_{i-1}) - m_{\theta_0}(X_{i-1}) = \dot{m}_{\theta_0}^T(X_{i-1})(\theta_n - \theta_0) + \eta_n(X_{i-1})$ to the i^{th} summand in $Z_n(x,\theta_n)$, and basic inequality, the left hand side of the lemma is bounded above by a sum of following two terms:

$$\begin{aligned}
& 2\|\theta_n - \theta_0\|^2 \int_{\mathcal{I}} \|\dot{\mu}_n(\theta_0)\|^2 dG(x), \\
& 2\|\theta_n - \theta_0\|^2 \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^n K_h(x-X_{i-1})\Big(\frac{\eta_n(X_{i-1})}{\|\theta_n-\theta_0\|}\Big)\right)^2 dG(x).
\end{aligned}$$

The first term is $O_p(\frac{1}{n})$ by Lemma 3.3.1 part (c) with $\psi_n = m_{\theta_0}$ and Theorem 3.4.1. The second term is $o_p(\frac{1}{n})$ by Lemma 3.5.1 and the Theorem 3.4.1. Hence the lemma is proved. □

Now we are ready to prove the propositions.

Proof of proposition 3.5.1. By plugging in $X_i = m_{\theta_0}(X_{i-1}) + \varepsilon_i$ to the i^{th} summand in $M_h(\theta_n)$ and $M_h(\theta_0)$, and expanding the quadratic, the left hand side of proposition 3.5.1 is equal to a sum of following two terms:

$$(a) \quad \int_{\mathcal{I}} Z_n^2(x,\theta_n)^2 dG(x), \quad \text{and} \quad (b)\, 2\int_{\mathcal{I}} Z_n(x,\theta_n)U_n(x)dG(x).$$

(a) is $O_p(\frac{1}{n})$ by Lemma 3.5.2. By plugging in

$$m_{\theta_n}(X_{i-1}) - m_{\theta_0}(X_{i-1}) = \dot{m}_{\theta_0}^T(X_{i-1})(\theta_n - \theta_0) + \eta_n(X_{i-1})$$

to the i^{th} summand in $Z_n(x,\theta_n)$, the term (b) can be written as a sum of following two terms:

$$(b_1) \qquad (\theta_n - \theta_0)^T \int_{\mathcal{I}} \dot{\mu}_n(x,\theta_0)U_n(x)dG(x),$$

$$(b_2) \qquad (\theta_n - \theta_0)^T \frac{(\theta_n - \theta_0)}{\|\theta_n - \theta_0\|} \int_{\mathcal{I}} \left(\frac{1}{n}\sum_{i=1}^n K_h(x - X_{i-1})(\frac{\eta_n(X_{i-1})}{\|\theta_n - \theta_0\|}) \right) U_n(x)dG(x).$$

The term (b_1) is $o_p(\frac{1}{nh^{1/2}})$ by Lemma 3.4.1 (a) , Lemma 3.4.2 with $l = 1$ and ξ_n replaced by ξ, (3.3.3), and Theorem 3.4.1. The term (b_2) is $o_p(\frac{1}{nh^{1/2}})$ by Cauchy-Schwarz inequality, Lemma 3.5.1, (3.3.3), and Theorem 3.4.1. Hence the proposition follows. □

Proof of proposition 3.5.2. By plugging in $\hat{\varepsilon}_i = X_i - m_{\theta_n}(X_{i-1})$ and $X_i = m_{\theta_0}(X_{i-1}) + \varepsilon_i$ to the i^{th} summand, and expanding the quadratic, the left hand side of this proposition can be written as a sum of following two terms:

$$(a) \qquad \frac{1}{n^2}\sum_{i=1}^n \int_{\mathcal{I}} K_h^2(x - X_{i-1})(m_{\theta_n}(X_{i-1}) - m_{\theta_0}(X_{i-1}))^2 dG(x),$$

$$(b) \qquad \frac{-2}{n^2}\sum_{i=1}^n \int_{\mathcal{I}} K_h^2(x - X_{i-1})\varepsilon_i\,(m_{\theta_n}(X_{i-1}) - m_{\theta_0}(X_{i-1}))dG(x).$$

To prove the proposition, it suffices to show that

$$(3.5.2) \qquad (a) = o_p(\frac{1}{nh^{1/2}}) \qquad \text{and} \qquad (b) = o_p(\frac{1}{nh^{1/2}}).$$

Note that by plugging in $m_{\theta_n}(X_{i-1})-m_{\theta_0}(X_{i-1}) = \dot{m}_{\theta_0}^T(X_{i-1})(\theta_n-\theta_0)+\eta_n(X_{i-1})$ to the i^{th} summand in (a), expanding the quadratic, term (a) is bounded above by the sum of following terms

$$\|\theta_n - \theta_0\|^2 \frac{2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{I}} K_h^2(x - X_{i-1})\dot{m}_{\theta_0}^2(X_{i-1})dG(x), \qquad \text{and}$$
$$\|\theta_n - \theta_0\|^2 \frac{2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{I}} K_h^2(x - X_{i-1})(\frac{\eta_n^2(X_{i-1})}{\|\theta_n - \theta_0\|^2})dG(x).$$

The first term of (a) is $O_p(\frac{1}{n^2h})$ by taking the expectation of the summation and Theorem 3.4.1. Similar to the argument of first term in (3.5.1), the second term of (a) is $O_p(\frac{1}{n^2h})$. Hence the first part of (3.5.2) holds.

Using the same skill and similar argument as above, one obtains that term (b) can be written as a sum of following two terms

$$(\theta_n - \theta_0)^T \frac{1}{nh}\int_{\mathcal{I}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K^2(\frac{x - X_{i-1}}{h})\dot{m}_{\theta_0}(X_{i-1})\varepsilon_i\right)dG(x),$$
$$(\theta_n - \theta_0)^T \frac{(\theta_n - \theta_0)}{\|\theta_n - \theta_0\|}\frac{1}{nh}\int_{\mathcal{I}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K^2(\frac{x - X_{i-1}}{h})(\frac{\eta_n(X_{i-1})}{\|\theta_n - \theta_0\|})\varepsilon_i\right)dG(x).$$

By taking the expectation of the absolute value of the integration, and Theorem 3.4.1, the first term is $o_p(\frac{1}{nh^{1/2}})$. Similarly the second term is also $o_p(\frac{1}{nh^{1/2}})$. Hence the second part of (3.5.2) holds, so does the proposition. □

Now we prove the proposition 3.5.3.

Proof of proposition 3.5.3. Define

$$\phi_{ij} = \int_{\mathcal{I}} K_h(x - X_{i-1})K_h(x - X_{j-1})dG(x)\varepsilon_i\varepsilon_j,$$
$$V_{nj} = \frac{1}{n^2}\sum_{i=1}^{j}\phi_{ij}, \qquad U_n = \sum_{j=1}^{n}V_{nj}.$$

Then, U_n is a sum of Martingale differences. In order to apply the M.G.C.L.T to prove the proposition 3.5.3, one needs to check the following three conditions:

$$(C1) \qquad Var(U_n) = \sum_{j=1}^{n-1} EV_{nj}^2 = \sigma_n^2,$$

$$(C2) \qquad \frac{1}{\sigma_n^2} \sum_{j=1}^{n} V_{nj}^2 \longrightarrow 1 \text{ in probability},$$

$$(C3) \qquad \frac{1}{\sigma_n^2} \sum_{j=1}^{n} E\left\{V_{nj}^2 I_{\{|V_{nj}|>\varepsilon\sigma_n\}} | \mathcal{F}_{n,j-1}\right\} \longrightarrow 0 \text{ in probability}, \forall \varepsilon > 0.$$

The proof of proposition 3.5.3 is broken down into four lemmas.

Lemma 3.5.3 *Under the conditions (M) to (H),*

$$(n^2h)^2 E\left(n^{-4} \sum_{i<l<j} \phi_{ij}\phi_{lj}\right)^2 = o(1).$$

Lemma 3.5.4 *Under the conditions (M) to (H),*

$$Var(U_n) = \sum_{j=1}^{n} EV_{nj}^2 = \sigma_n^2 = O(n^2h).$$

Lemma 3.5.5 *Under the conditions (M) to (H),*

$$\sigma_n^{-4} \sum_{j=1}^{n} EV_{nj}^4 = o(1).$$

Lemma 3.5.6 *Under the conditions (M) to (H),*

$$\sigma_n^{-2} n^{-4} \sum_{i<j}^{n} (\phi_{ij}^2 - E\phi_{ij}^2) = o_p(1).$$

The condition (C1) and (C3) are the direct consequences of lemma 3.5.4 and Lemma 3.5.5 respectively. Note that

$$V_{nj}^2 = \frac{1}{n^4}\left(\sum_{i=1}^{j} \phi_{ij}^2 + 2\sum_{i<l<j} \phi_{ij}\phi_{lj}\right).$$

So,

$$\sum_{j=1}^{n}(V_{nj}^2 - EV_{nj}^2)/\sigma_n^2$$

$$= \frac{1}{\sigma_n^2 n^4}\sum_{i<j}^{n}(\phi_{ij}^2 - E\phi_{ij}^2) + \frac{2}{\sigma_n^2 n^4}\sum_{i<l<j}(\phi_{ij}\phi_{lj} - E\phi_{ij}\phi_{lj}) = o_p(1)$$

by Lemma 3.5.3, Lemma 3.5.4, and lLemma 3.5.6. Hence the condition (C2) holds by Lemma 3.5.4. Therefore the proposition is proved. □

Next we will focus on the proof of the four lemmas.

Proof of Lemma 3.5.3. Note that left hand side of lemma is equal to

$$\frac{h^2}{n^4}\sum_{(i<l<j),(i'<l'<j')} E\phi_{ij}\phi_{lj}\phi_{i'j'}\phi_{l'j'} \tag{3.5.3}$$

where the summand $E\phi_{ij}\phi_{lj}\phi_{i'j'}\phi_{l'j'}$ is

$$E\int_{\mathcal{I}\times\mathcal{I}\times\mathcal{I}\times\mathcal{I}} K_h(x-X_{i-1})K_h(x-X_{j-1})K_h(y-X_{l-1})K_h(y-X_{j-1})$$
$$K_h(s-X_{i'-1})K_h(s-X_{j'-1})K_h(t-X_{l'-1})K_h(t-X_{j'-1})\varepsilon_i\varepsilon_l\varepsilon_j^2\varepsilon_{i'}\varepsilon_{l'}\varepsilon_{j'}^2 dG_{xyst},$$

and $dG_{xyst} := dG(x)dG(y)dG(s)dG(t)$.

Define

$$F_\nu = \{(i<l<j, i'<l'<j') : \textit{there are } \nu \textit{ distinct values in } i,l,j,i',l',j'.\}.$$

Then on F_ν, (3.5.4) is bounded above by

$$const\cdot\frac{1}{h^{8-\nu-3}} = O(\frac{1}{h^{5-\nu}}), \quad \nu\le 4. \tag{3.5.4}$$

Denote the ν distinct indices defined in F_ν by $(i_1 < i_2 <, ..., < i_\nu)$. Define d_j be the j^{th} largest difference among $i_{l+1} - i_l$, $i = 1, ..., \nu - 1$. Also define

$$K_{j_1, j_2, \dots j_k}(\cdot,\cdot,\cdot,\cdot) = \prod_{l\in(j_1,j_2,\dots j_k)} K_h(\cdot - X_{l-1})\varepsilon_l^{p_l},$$

where p_l is either 1 or 2. Then (3.5.3) can be written as

$$(3.5.5)\quad \frac{h^2}{n^4}\sum_{\nu=3}^{6} c_\nu \sum_{F_\nu} E\phi_{ij}\phi_{lj}\phi_{i'j'}\phi_{l'j'} = c_3A_3 + c_4A_4 + c_5A_5 + c_6A_6, \qquad \text{say,}$$

for some constants c_3, c_4, c_5, and c_6. In order to prove the lemma, it suffices to show that $A_\nu = o(1)$, for $\nu = 3, 4, 5, 6$.

But when $\nu = 3$ or 4, by (3.5.4) and (3.5.5),

$$A_3 = O(\frac{h^2}{n^4}\cdot n^3 \cdot \frac{1}{h^{5-3}}) = O(\frac{1}{n}) = o(1),$$
$$A_4 = O(\frac{h^2}{n^4}\cdot n^4 \cdot \frac{1}{h^{5-4}}) = O(h) = o(1).$$

So we only need to show $A_5 = o(1)$ and $A_6 = o(1)$.

Define

$$\tau = min\left\{j \le 5 : d_j = i_{l+1} - i_l,\ and\, EK_{i_1,\ldots,i_l}(\cdot,\cdot,\cdot,\cdot) = 0\, for\ some\ l.\right\}.$$

It is seen from this definition that on F_ν

$$(3.5.6)\qquad \tau \le 8 - \nu.$$

Next we will show that

$$(3.5.7)\qquad A_\nu = O(\frac{1}{n^{4-\tau}h^2}), \quad \nu = 5, 6.$$

Suppose $d_\tau = i_{l+1} - i_l$ for some l. On F_ν, $\nu = 5, 6$, (3.5.3) is equal to

$$\int_{I\times I\times I\times I} \mathrm{Cov}\left(K_{i_1,\ldots,i_l}(x,y,s,t), K_{i_{l+1},\ldots,i_\nu}(x,y,s,t)\right) dG(x)dG(y)dG(s)dG(t).$$

By Lemma 3.2.1, the above term is bounded above by

$$(3.5.8)\quad \int_{I\times I\times I\times I} 2p[2\alpha(d_\tau)]^{1/p}\|K_{i_1,\ldots,i_l}\|_q \|K_{i_{l+1},\ldots,i_\nu}\|_r dG(x)dG(y)dG(s)dG(t),$$

for any $p, q, r > 1$, and $\frac{1}{q} + \frac{1}{q} + \frac{1}{r} = 1$, where

$$\|K_{i_1,\dots,i_l}\|_q = (E|K_{i_1,\dots,i_l}(x, y, s, t)|^q)^{1/q}.$$

By an usual calculation, (3.5.8) is bounded above by

$$const \cdot p[\alpha(d_\tau)]^{1/p} \frac{1}{h^{8-(\frac{l}{q}+\frac{\nu-l}{r})-3}}.$$

By taking $q = r = \nu$, A_ν is bounded above by

$$const \cdot \frac{h^2}{n^4} n^\tau \sum_{d_\tau} d_\tau^{\nu-1-\tau} [\alpha(d_\tau)]^{3/5} \frac{1}{h^4} = O(\frac{1}{n^{4-\tau} h^2}), \quad \nu = 5, 6,$$

by condition (M). So, (3.5.7) is proved. In view of (3.5.6),

$$A_5 = O(\frac{1}{nh^2}) = o(1), \qquad A_6 = O(\frac{1}{n^2 h^2}) = o(1).$$

The lemma therefore is proved. □

Proof of Lemma 3.5.4. Because $E\phi_{ij} = 0$ for any $i \neq j$, so $EV_{nj}V_{nl} = 0$ for $j \neq l$, and $EU_n = 0$. Hence

$$\begin{aligned} Var(U_n) &= EU_n^2 = E\left(\sum_{j=1}^{n-1} V_{nj}\right)^2 = \sum_{j=1}^{n-1} EV_{nj}^2 \\ &= \frac{1}{n^4} \sum_{i<j}^{n} E\phi_{ij}^2 + \frac{1}{n^4} \sum_{i<l<j} E\phi_{ij}\phi_{lj}. \end{aligned} \tag{3.5.9}$$

By Lemma 3.5.3 the second term of right hand side of the above equation is $o((n^2h)^{-1})$. Note that $E\phi_{ij}^2$ can be written as a sum of two terms:

$$(a) \quad \int_{I\times I} \text{Cov}\left(K_i(x, y)\varepsilon_i^2, K_j(x, y)\varepsilon_j^2\right) dG(x)dG(y),$$

$$(b) \quad \int_{I\times I} \left(EK_1(x, y)\varepsilon_1^2\right)^2 dG(x)dG(y),$$

where

$$K_i(x,y) = K_h(x - X_{i-1})K_h(y - X_{i-1}).$$

By Lemma 3.2.1, the term (a) is bounded above by

$$\text{(3.5.10)} \quad const \cdot 2p[2\alpha(j-i)]^{1/p} \int_{\mathcal{I}\times\mathcal{I}} \|K_i(x,y)\|_q \|K_j(x,y)\|_r dG(x)dG(y)$$

for any $p, q, r > 1$ and $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$. Take $p = q = r = 3$, then (3.5.10) is bounded above by

$$\text{(3.5.11)} \qquad const \cdot [\alpha(j-i)]^{1/3} \frac{1}{h^{2\frac{1}{3}}}.$$

Hence (a) is bounded above by

$$const \frac{1}{n^4 h^{2/3}} \sum_{k=1}^{n-1} \sum_{j-i=k} (\lambda(k))^{1/3} = o((n^2 h)^{-1}).$$

On the other hand, by the direct calculation,

$$\begin{aligned} h \cdot \int_{\mathcal{I}\times\mathcal{I}} & \left(EK_1(x,y)\varepsilon_1^2\right)^2 dG(x)dG(y) \\ & \longrightarrow (\sigma^2)^2 \int_{\mathcal{I}} f^2(x) g^2(x) dx \cdot \int \left(\int K(u)K(u+v)du \right)^2 dv =: \Gamma^2. \end{aligned}$$

This together with (3.5.9) as well as Lemma 3.5.3 implies that

$$n^2 h \cdot V(U_n) = n^2 h \cdot \sigma_n^2 = n^2 h \cdot \sum_{j=1}^{n-1} EV_{nj}^2 \longrightarrow \Gamma^2.$$

So Lemma 3.5.4 is proved. □

Proof of Lemma 3.5.5.

Note that

$$(3.5.12)\ n^4h^2\sum_j EV_{nj}^2 = c_1\frac{h^2}{n^4}\sum_{i<j}E\phi_{ij}^4 + c_2\frac{h^2}{n^4}\sum_{i\neq l<j}E\phi_{ij}^2\phi_{lj}^2 + c_3\frac{h^2}{n^4}\sum_{i\neq l<j}E\phi_{ij}\phi_{lj}^3$$
$$+c_4\frac{h^2}{n^4}\sum_{i\neq l\neq k<j}E\phi_{ij}\phi_{lj}\phi_{kj}^2 + c_5\frac{h^2}{n^4}\sum_{i\neq l\neq s\neq t<j}E\phi_{ij}\phi_{lj}\phi_{sj}\phi_{tj}$$

for some constants c_1, c_2, c_3, c_4 and c_5.

By direct calculation, the first term on the right hand side of (3.5.12) is bounded above by

$$const\cdot\frac{h^2}{n^4}\cdot\sum_{i<j}\frac{1}{h^3} = O(\frac{1}{n^2h}) = o(1).$$

Similarly, the second and the third terms are bounded above by

$$const\cdot\frac{h^2}{n^4}\sum_{i\neq l<j}\frac{1}{h^2} = O(\frac{1}{n}) = o(1),$$

and the fourth term is bounded above by

$$const\cdot\frac{h^2}{n^4}\sum_{i\neq l\neq k<j}\frac{1}{h} = O(h) = o(1).$$

So in order to prove the lemma, it suffices to show that the fifth term on right hand side of (3.5.12) is $o(1)$. But note that the fifth term is actually a special case of (3.5.3), hence it is $o(1)$ by Lemma 3.5.3. The lemma is proved. □

Now we are about to prove Lemma 3.5.6. But before that we first prove an inequality. This inequality is an extension of Theorem 1.3 part (1) of Bosq (1998). The inequality is given in the following theorem.

Theorem 3.5.2 *Suppose $\{\mathcal{F}_n\}$ is a family of σ-field satisfying $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$. ξ_{ij} is measurable w.r.t $\mathcal{F}_{i\bigvee j}$, $E\xi_{ij} = 0$ for $i \neq j$, and $|\xi_{ij}| \leq b$. Then for any $\varepsilon > 0$,*

$$P\left(|\sum_{i<j}^{n}\xi_{ij}| > n^2\varepsilon\right) \leq c_1 q exp\left\{-\frac{q\varepsilon^2}{8b^2}\right\} + c_2 q^2\left(1+\frac{8b}{\varepsilon}\right)^{1/2} \cdot \alpha\left([\frac{n}{2q}]\right),$$

for some constants $c_1, c_2 > 0$ and $1 \leq q \leq [\frac{n}{2}]$.

Proof.

First consider blocking. Let p be an integer between 1 and n. Let $q = [\frac{n}{2p}] + 1$.

Define

$$\begin{aligned}
V_1^{1(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = 1 \sim p,\ j = 1 \sim p, \ldots\ldots \\
V_l^{1(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = 2lp+1 \sim (2l+1)p,\ j = 2lp+1, \sim (2l+1)p, \\
&\quad \ldots\ldots \\
V_1^{2(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = 1 \sim p,\ j = p+1 \sim 2p, \ldots\ldots \\
V_l^{2(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = 2lp+1 \sim (2l+1)p,\ j = (2l+1)p+1 \sim 2(l+1)p, \\
&\quad \ldots\ldots \\
V_1^{3(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = p+1 \sim 2p,\ j = p+1 \sim 2p, \ldots\ldots \\
V_l^{3(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = (2l+1)p+1 \sim 2(l+1)p,\ j = (2l+1)p+1 \sim 2(l+1)p, \\
&\quad \ldots\ldots \\
V_1^{4(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = p+1 \sim 2p,\ j = 2p+1 \sim 3p, \ldots\ldots \\
V_l^{4(1)} &= \sum_{i<j}^{n}\xi_{ij},\ i = (2l+1)p+1 \sim 2(l+1)p, \\
&\qquad j = (2l+1)p+1 \sim (2(l+2)+1)p, \ldots\ldots
\end{aligned}$$

Define

$$A_k = \{(i,j) : (k-1)p < j - i \le kp,\ i = 1 \sim n,\ j = 1 \sim n\}.$$

Then

$$\sum_{i<j}^{n} \xi_{ij} = \sum_{k=1}^{2q} \sum_{(i,j)\in A_k} \xi_{ij}.$$

So for any $\varepsilon > 0$,

$$P\left(|\sum_{i<j}^{n} \xi_{ij}| > n^2\varepsilon\right) \le \sum_{k=1}^{2q} P\left(|\sum_{(i,j)\in A_k} \xi_{ij}| > \frac{n^2\varepsilon}{2q}\right). \tag{3.5.13}$$

On each A_k, define $V_l^{i(k)}$ $i = 1 \sim 4$ as above, then

$$\sum_{(i,j)\in A_k} \xi_{ij} = \sum_l V_l^{1(k)} + \sum_l V_l^{2(k)} + \sum_l V_l^{2(k)} + \sum_l V_l^{4(k)}. \tag{3.5.14}$$

Hence

$$P\left(|\sum_{(i,j)\in A_k} \xi_{ij}| > \frac{n^2\varepsilon}{2q}\right) \le \sum_{i=1}^{4} P\left(|\sum_l V_l^{i(k)}| > \frac{n^2\varepsilon}{8q}\right).$$

But by recursively using Bradley's lemma 1.2 in Bosq (1998), there are independent random variables $W_1^{i(k)}, \ldots, W_l^{i(k)}, \ldots$ such that $P_{W_l^{i(k)}} = P_{V_l^{i(k)}}$, and

$$P\left(|W_l^{i(k)} - V_l^{i(k)}|\right) \le 11 \cdot \left(\frac{\|V_l^{i(k)} + c\|_\infty}{\lambda}\right)^{1/2} \alpha(p). \tag{3.5.15}$$

for any $0 < \lambda \le \|V_l^{i(k)} + c\|_\infty$. Hence

$$\begin{aligned}
&\left(|\sum_l V_l^{i(k)}| > \frac{n^2\varepsilon}{8q}\right) \\
= \;& P\left(|\sum_l V_l^{i(k)}| > \frac{n^2\varepsilon}{8q}, |W_l^{i(k)} - V_l^{i(k)}| \le \lambda,\ \forall l\right) \\
&\quad + P\left(\bigcup_l \{|W_l^{i(k)} - V_l^{i(k)}| > \lambda\}\right) \\
\le \;& P\left(|\sum_l W_l^{i(k)}| > \frac{n^2\varepsilon}{8q} - q\lambda\right) + \sum_{l=1}^{q} P\left(|W_l^{i(k)} - V_l^{i(k)}| > \lambda\right).
\end{aligned} \tag{3.5.16}$$

Choose

$$\lambda = min\left(\frac{n^2\varepsilon}{16q^2}, (\delta-1)bp^2\right),$$

then

$$\begin{aligned} \|V_l^{i(k)} + c\|_\infty &\leq \|V_l^{i(k)}\|_\infty + c \leq (\delta+1)bp^2. \\ \|V_l^{i(k)} + c\|_\infty &\geq c - \|V_l^{i(k)}\|_\infty \geq (\delta-1)bp^2 > 0. \end{aligned}$$

So, $0 < \lambda \leq \|V_l^{1(k)} + c\|_\infty$. Hence in view of (3.5.15), (3.5.16) is bounded above by

(3.5.17)

$$P\left(|\sum_l W_l^{i(k)}| > \frac{n^2\varepsilon}{16q}\right) + q \cdot 11 \cdot \left(max(\frac{\delta+1}{\delta-1}, \frac{16q^2p^2b(\delta+1)}{n^2\varepsilon})\right)^{1/2} \cdot \alpha(p)$$

Choose δ such that

$$\frac{1}{\delta-1} = \frac{16q^2p^2b}{n^2\varepsilon} = \frac{4b}{\varepsilon},$$

then (3.5.17) is bounded above by

$$P\left(|\sum_l W_l^{i(k)}| > \frac{n^2\varepsilon}{16q}\right) + 11 \cdot q \cdot \left(1 + \frac{8b}{\varepsilon}\right)^{1/2} \cdot \alpha(p). \tag{3.5.18}$$

But by applying Hoeffding's inequality to $W_l^{1(k)}$, one may obtain that

$$P\left(|\sum_l W_l^{i(k)}| > \frac{n^2\varepsilon}{16q}\right) \leq 2exp\left\{-\frac{2(\frac{n^2\varepsilon}{16q})^2}{q(bp^2)^2}\right\} \leq 2exp\left\{-\frac{n\varepsilon^2}{16b^2p}\right\}.$$

Hence (3.5.18) is bounded above by

$$2exp\left\{-\frac{n\varepsilon^2}{16b^2p}\right\} + 11 \cdot q \cdot \left(1 + \frac{8b}{\varepsilon}\right)^{1/2} \cdot \alpha(p).$$

So (3.5.13) is bounded above by

$$16 \cdot q \cdot exp\left\{-\frac{n\varepsilon^2}{16b^2 p}\right\} + 88 \cdot q^2 \left(1 + \frac{8b}{\varepsilon}\right)^{1/2} \cdot \alpha(p).$$

The theorem is thus proved. □

We will apply the above inequality to prove Lemma 3.5.5.

Proof of Lemma 3.5.6.

We will show that for any $\lambda > 0$,

$$P\left(\left|\frac{h}{n^2}\sum_{i<j}^{n}(\phi_{ij}^2 - E\phi_{ij}^2)\right| > \lambda\right) = o(1).$$

Apply Theorem 3.5.2 with $\xi_{ij} = \phi_{ij}^2 - E\phi_{ij}^2$, By boundedness of g and kernel function K, one obtains that ξ_{ij} is bounded above by c/h^2 for some constant c that doesn't depend on i, j. For any positive number λ,

$$(3.5.19) \quad P\left(\left|\frac{h}{n^2}\sum_{i<j}^{n}(\phi_{ij}^2 - E\phi_{ij}^2)\right| > \lambda\right) \le P\left(\left|\sum_{i<j}^{n}(\phi_{ij}^2 - E\phi_{ij}^2)\right| > n^2\frac{\lambda}{h}\right).$$

By Theorem 3.5.2 with $\varepsilon = \lambda/h$ and $b = c/h^2$, one obtains that (3.5.19) is bounded above by

$$(3.5.20) \quad c_1 \cdot q \cdot exp\left\{-\frac{\varepsilon^2}{8c^2} \cdot qh^2\right\} + c_2 \cdot q^2 \cdot \left(1 + \frac{8c}{\varepsilon} \cdot \frac{1}{h}\right)^{1/2} \cdot \alpha([\frac{n}{2q}]).$$

Choose $q = n^{\frac{2}{2+\gamma}}$, then $qh^2 \longrightarrow \infty$ and $n/(2q) = O(n^{\frac{\gamma}{2+\gamma}}) \longrightarrow \infty$ as n tends to ∞. Hence both two terms in (3.5.20) tend to 0 by condition (M). therefore the proof of the Lemma 3.5.6 is complete. □

Chapter 4

Simulations

This chapter contains a simulation study comparing three tests. More precisely, let $\{X_t,\, t = 0, \pm 1 \pm 2, ...\}$ be a stationary stochastic process satisfying

$$X_t = \mu(X_{t-1}) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ are i.i.d. r.v.'s with mean zero and ε_t is independent of X_{t-1}, for all t. The parametric family of functions to be fitted to μ is chosen to be $m_\theta(x) = \theta\, x,\ x \in \mathbb{R},\ \theta \in \mathbb{R}$ with $\theta_0 = 0.8$. That is, the hypothesis to be tested is

$$H_0 : \mu(x) = 0.8x, \qquad vs. \quad H_1 : H_0 \text{ is not true.}$$

We chose the following three models to generate simulated data from:

$$\textit{model}\,1. \qquad X_{t+1} = 0.8X_t + \varepsilon_{t+1},$$

$$\textit{model}\,2. \qquad X_{t+1} = 0.8X_t - 1.2exp(-X_t^2))X_t + \varepsilon_{t+1} + 0.1,$$

$$\textit{model}\,3. \qquad X_{t+1} = 0.8X_t + 0.5(X_t - 0.5)^2 - 0.3(X_t - 0.5)^3 + \varepsilon_{t+1}.$$

The error distribution is either $N(0, 0.1)$ or double exponential. The sample sizes chosen are 50, 100, 200, and 500. The three different tests are those of Koul and Stute (1999) denoted by KS, An and Cheng (1991) denoted by AC, and the minimum distance test of Chapter 3 denoted by MD. The nominal level that is used to implement the test is $\alpha = 0.05$. There are 1000 replications for each combination of $(model, sample\, size, error\, distribution)$. Data from model 1 are used to study the empirical size, and the data from models 2 and 3 are used to study the empirical power of these tests. The empirical size (power) is computed by

$$Relative\, frequency\, of\, (\, value\, of\, the\, test\, statistic\, > F^{-1}(1-\alpha)),$$

where F is the asymptotic distribution of the test statistics under H_0.

The steps to compute the test statistics are as follows: Let $X_{(0)} \leq, ..., \leq X_{(n)}$ denote the ordered $X_0, X_1, ..., X_n$.

1. **Koul and Stute test**:

Step 1: Compute the least square estimate of θ_0 under H_0:

$$\theta_{lse} = \sum_{i=1}^{n} X_{i-1}X_i / \sum_{i=1}^{n} X_{i-1}^2.$$

Step 2: Compute $V_n(X_i)$, $i = 1, 2, ...n$, where

$$V_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \theta_{lse}X_{i-1})\, I(X_{i-1} \leq x), \quad x \in \mathbb{R}.$$

Step 3: Compute $A_n(X_{(i)})$, $X_{(i)} \leq x_0$, where x_0 is the 99^{th} percentile of the sample

$X_0, ..., X_n$, and

$$A_n(x) = \int y^2 I(y \geq x) G_n(dy) = \frac{1}{n}\sum_{i=1}^{n} X_{i-1}^2 I(X_{i-1} \geq x),$$

$$G_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_{i-1} \leq x), \quad x \in \mathbb{R}.$$

Step 4: Compute the estimate of the error variance:

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \theta_{lse} X_{i-1})^2 .$$

Step 5: Compute $T_n V_n(X_i)$, $i = 1, 2, ..., n$, where

$$T_n V_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[I(X_{i-1} \leq x) - \frac{1}{n}\sum_{j=1}^{n} X_{j-1} A_n^{-1}(X_{j-1}) X_{i-1} \times I(X_{j-1} \leq X_{i-1} \wedge x) \Big] (X_i - \theta_{lse} X_{i-1}).$$

Step 6: compute the test statistic

$$T_n^1 = \sup_{x \leq x_0} \frac{|T_n V_n(x)|}{\sigma_n \sqrt{G_n(x_0)}} = \sup_{X_i \leq x_0} \frac{|T_n V_n(X_i)|}{\sigma_n \sqrt{G_n(x_0)}}.$$

The limiting distribution of the test statistic is $\sup_{0<t<1} |B(t)|$, where $B(t)$ is a standard Brownian motion. The 95^{th} percentile of this distribution is approximately equal to 2.2414 obtained from the formula

$$P\Big(\sup_{0<t<1} |B(t)| < b\Big) = P(|B(1)| < b) + 2\sum_{i=1}^{\infty} (-1)^i P\Big((2i-1)b < B(1) < (2i+1)b\Big),$$

given on page 553 of the book by Resnick (1992). The number 2.2414 is such that

$$\left| P\Big(\sup_{0<t<1} |B(t)| < 2.2414\Big) - 0.95 \right| \leq (0.1)^6.$$

The same cut off value was also used by An and Cheng (1991).

2. **An and Cheng test**:

Step 1: Compute the sample covariance:

$$\hat{\gamma}=\frac{1}{n}\sum_{i=1}^{n} X_{i-1}^2, \quad \text{and} \quad \hat{\gamma}_1=\frac{1}{n}\sum_{i=1}^{n} X_i X_{i-1}.$$

Step 2: Compute

$$\hat{\rho}=\frac{\hat{\gamma}_1}{\hat{\gamma}_0}.$$

Step 3: Compute $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \hat{\rho}X_{i-1})^2$.

Step 4: Compute $\hat{e}_i = X_i - \hat{\rho}X_{i-1}$.

Step 5: Compute $\hat{K}_n(X_i)$, $i = 1, 2, ..., n$, where

$$\hat{K}_n(t)=\frac{1}{\sqrt{m\hat{\sigma}^2}}\sum_{i=2}^{m}\hat{e}_i I(X_{i-1}<t),$$

where $m = m(n)$ is a subsequence of n for which m/n is about 0.75.

Step 7: Compute the test statistic $\hat{K}_n = \sup_t |\hat{K}_n(t)| = \sup_i |\hat{K}_n(X_i)|$.

The limiting distribution of the test statistic is the same as that of the KS test.

3. **Minimum distance test**:

Choose the kernel function $K(u) = \frac{3}{4}(1-u^2)I(|u| \le 1)$, the compact set $\mathcal{I} = [-1, 1]$, $h = n^{-1/4}$ and $G(x) = x$ on $[-1, 1]$.

Step 1: Compute the minimum distance estimate of θ_0,

$$\theta_n = \frac{\int_{\mathcal{I}}\left(\sum_{i=1}^{n} K_h(x-X_{i-1})X_i\right)\left(\sum_{i=1}^{n} K_h(x-X_{i-1})X_{i-1}\right)dx}{\int_{\mathcal{I}}\left(\sum_{i=1}^{n} K_h(x-X_{i-1})X_{i-1}\right)^2 dx}.$$

Step 2: Compute the residuals by

$$\hat{\varepsilon}_i = X_i - \theta_n X_{i-1},\ i = 1, 2, ..., n.$$

Step 3: Compute the test statistic

$$T_n = \frac{nh^{1/2}}{\hat{\Gamma}_n^2}\Big[\int_{\mathcal{I}}\Big(\frac{1}{n}\sum_{i=1}^{n}K_h(x-X_{i-1})\hat{\varepsilon}_i\Big)^2 dx - \frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{I}}K_h^2(x-X_{i-1})\hat{\varepsilon}_i^2 dx\Big].$$

where

$$\Gamma^2 = 4\int_{\mathcal{I}}\Big(\frac{1}{n}\sum_{i=1}^{n}K_h(x-X_{i-1})\hat{\varepsilon}_i^2\Big)^2 dx \int\Big(\int K(u)K(u+v)du\Big)^2 dv.$$

The simulation programming was done using C language. To generate a time series of size n from a given error distribution, first a random sample was generated from uniform $[0,1]$, then by calling rnorm function from R the errors from $N(0,0.1)$ were generated, while the errors from the double exponential were generated by inverting the distribution function. Then a series of $(101+n)$ r.v.'s are generated based on models 1, 2, 3, and the errors. Finally, $X_0, \ldots, X_n$ are taken to be the last $(n+1)$ observations.

The sizes and powers of the three tests were simulated for the sample sizes $n=50$, $n=100$, $n=200$ and $n=500$, each repeated 1000 times. The density curves of normalized θ_n and $M_h(\theta_n)$ are plotted by using density plot command with Gussian kernel option in SPLUS2000.

The results of the simulation study are shown in the Tables 4.1 to 4.3. The Tables 4.1 and 4.2 give the empirical sizes and powers of the three tests for testing model 1 against model 2, and for the error distributions double exponential and $N(0,0.1)$, respectively. Table 4.3 gives a similar data when testing model 1 against model 3 with the error distribution $N(0,0.1)$ only. From these tables, one sees that all three tests have equally good power performance for the sample size 500.

But KS and MD tests have better empirical sizes than the AC test for most of all remaining sample sizes. Compared to the other two tests, the MD test performs better for testing model 1 against model 3, and also has good performance in testing model 1 against model 2 when the error distribution is $N(0, 0.1)$ and the sample size 100 or more, but its power is not as good as that of the other two tests when the sample size is as small as 50. Overall, AC test has good power performance but it seems the empirical size is not good for testing model 1 against model 2 with double exponential errors. KS test seems to have better performance than AC and MD in testing model 1 against model 2 with normal error distribution, but not as good as the other two for testing model 1 against model 3 with $N(0, 0.1)$.

Table 4.1: *Tests for model 1 v.s. model 2 with double exponential errors.*

	$n = 50$		n=100		n=200		n=500	
tests	size	power	size	power	size	power	size	power
AC	0.059	0.027	0.308	0.209	0.132	0.789	0.128	0.934
KS	0.072	0.137	0.064	0.446	0.054	0.837	0.051	0.990
MD	0.012	0.111	0.043	0.406	0.045	0.824	0.050	0.999

Table 4.2: *Tests for model 1 vs. model 2 with $N(0, 0.1)$ errors.*

	$n = 50$		n=100		n=200		n=500	
tests	size	power	size	power	size	power	size	power
AC	0.010	0.867	0.022	0.999	0.025	1.000	0.034	1.000
KS	0.029	0.998	0.036	1.000	0.042	1.000	0.049	1.000
MD	0.011	0.659	0.019	1.000	0.023	1.000	0.044	1.000

Table 4.3: *Tests for model 1 vs. model 3 with the $N(0, 0.1)$ errors.*

	$n = 50$		n=100		n=200		n=500	
tests	size	power	size	power	size	power	size	power
AC	0.019	0.372	0.019	0.947	0.025	1.000	0.029	1.000
KS	0.013	0.554	0.071	0.777	0.059	0.871	0.046	1.000
MD	0.011	0.421	0.021	0.982	0.035	1.000	0.049	1.000

Tables 4.4 and 4.5 below list the mean and standard deviation of θ_n under H_0 with double exponential and $N(0, 0.1)$ errors, respectively. From the tables one can see that θ_n converges to $\theta_0 = 0.8$ as sample sizes change from 50 to 500, and the standard deviation tends to be smaller as sample size tends to be larger.

The simulation results of the densities of $\sqrt{n}(\theta_n - 0.8)$, the minimum distance test statistics, and suitably scaled minimized distances are shown in Figure 4.1 to Figure 4.12.

Table 4.4: *Mean and s.d.(θ_n) under model 1 with double exponential errors.*

sample size	n=50	n=100	n=200	n=500
mean	0.82	0.809	0.807	0.802
stdev	0.0963	0.0777	0.0533	0.0339

Table 4.5: *Mean and s.d(θ_n) under model 1 with normal errors.*

sample size	n=50	n=100	n=200	n=500
mean	0.845	0.821	0.813	0.807
stdev	0.0957	0.0682	0.0475	0.0306

Figure 4.1 is the Monte Carlo density curves of $\sqrt{n}(\theta_n - 0.8)$ from 1000 runs with sample size $n = 50$, $n = 100$, $n = 200$, $n = 500$ respectively when the error distribution is double exponential. Figure 4.2 is the Monte Carlo density curves of $\sqrt{n}(\theta_n - 0.8)$ when the error distribution is $N(0, 0.1)$. The graphs show that the distribution of $\sqrt{n}(\theta_n - 0.8)$ converges to its asymptotic normal distribution very quickly.

Figure 4.3 is the Monte Carlo density of T_n under H_0 when the error distribution is double exponential. Figure 4.4 is the Monte Carlo density of T_n under H_0 when the error distribution is double exponential. Figure 4.5 is the Monte Carlo density of T_n under H_0 when the error distribution is $N(0, 0.1)$. Figure 4.6 is the Monte Carlo density of T_n under H_0 when the error distribution is $N(0, 0.1)$. Figure 4.7

is the Monte Carlo density of T_n under H_0 when the error distribution is $N(0, 0.1)$. From these graphs, it can be seen that eventually the density of the test statistics converge to a standard normal density under H_0, and to a normal density with unit variance and a positive mean under the alternatives.

Figure 4.8 to Figure 4.12 are the Monte Carlo density of $nh^{1/2}M_h(\theta_n)$ with sample size $n = 50$, $n = 100$, $n = 200$ under models 1, 2, and 3, when the error distributions are double exponentail and normal. From the graphs we find that the densities under models 2 and 3 are approximately their counterparts under model 1 with positive shifts.

In the following figures, "$\cdots\cdots$" is for $n = 50$, "$- \cdot -$" is for $n = 100$, a heavy solid line is for $n = 200$, and a light solid line is for standard normal distribution.

Figure 4.1: *The density of $\sqrt{n}(\theta_n - 0.8)$ when the errors are double exponential.*

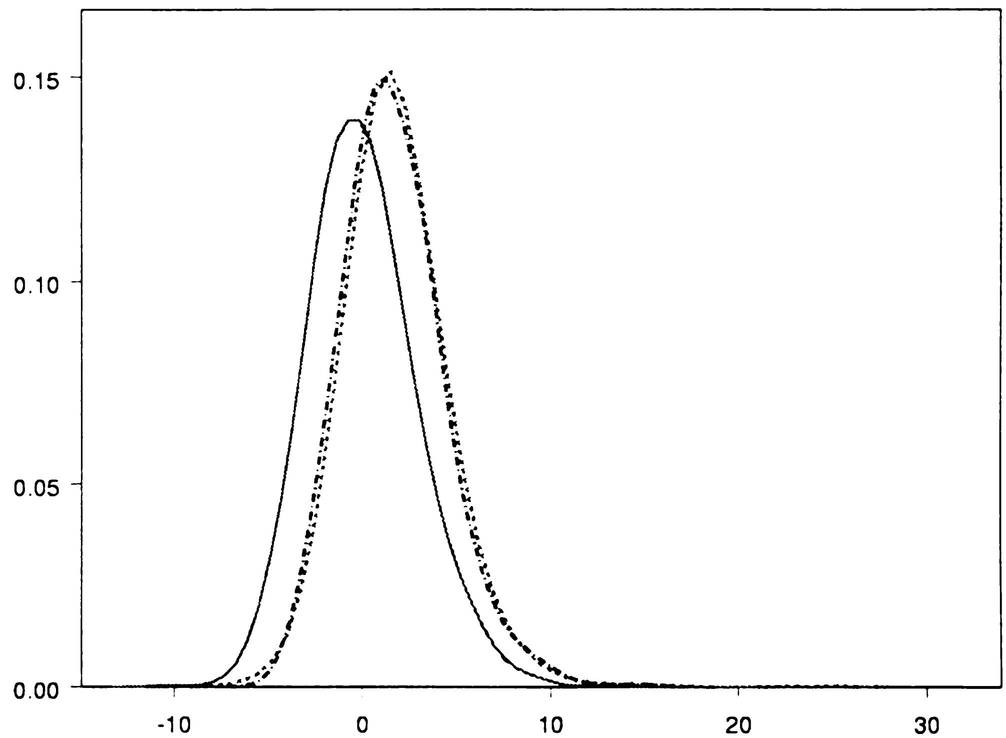


Figure 4.2: *The density of* $\sqrt{n}(\theta_n - 0.8)$ *when the errors are* $N(0, 0.1)$.

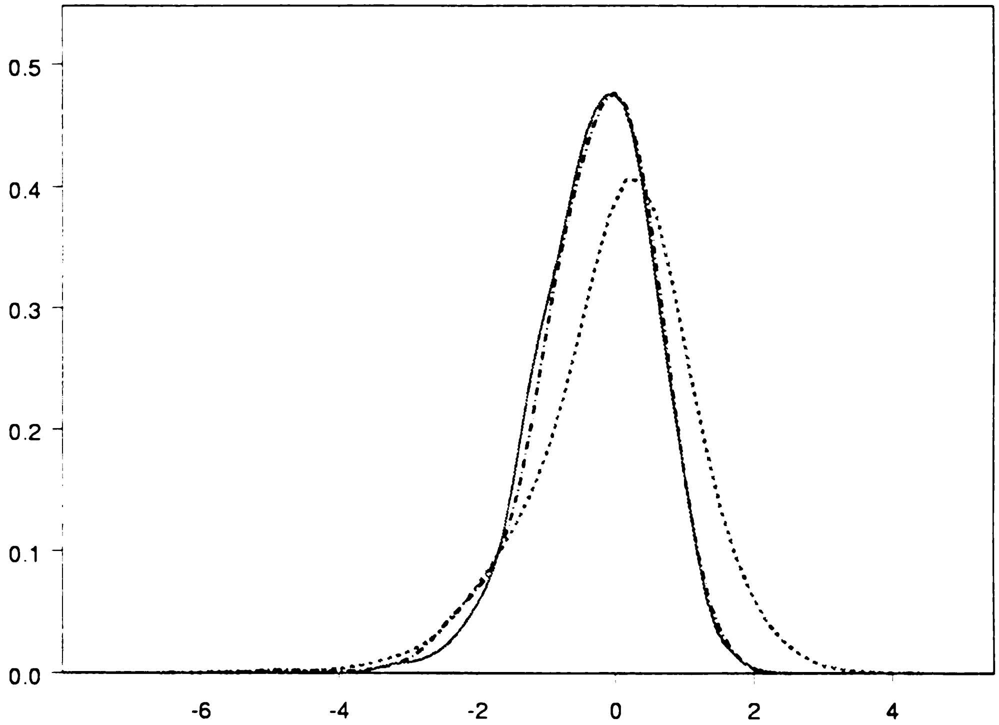


Figure 4.3: *The density of $T_n(\theta_n)$ under model 1 with double exponential errors.*

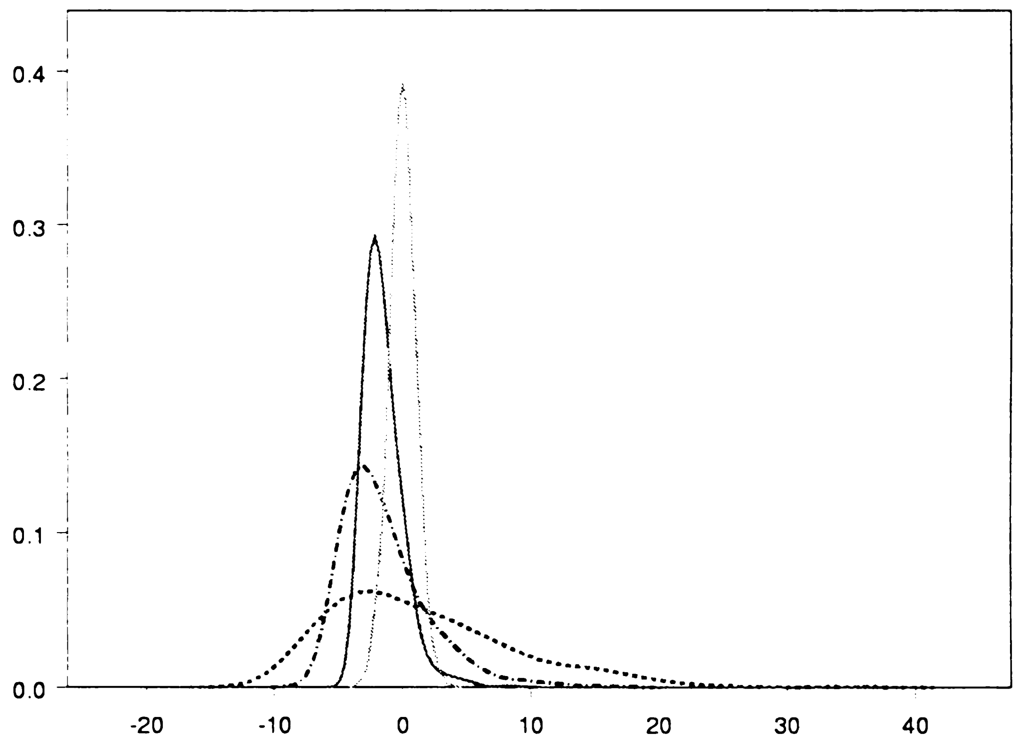


Figure 4.4: *The density of $T_n(\theta_n)$ under model 2 with double exponential errors.*

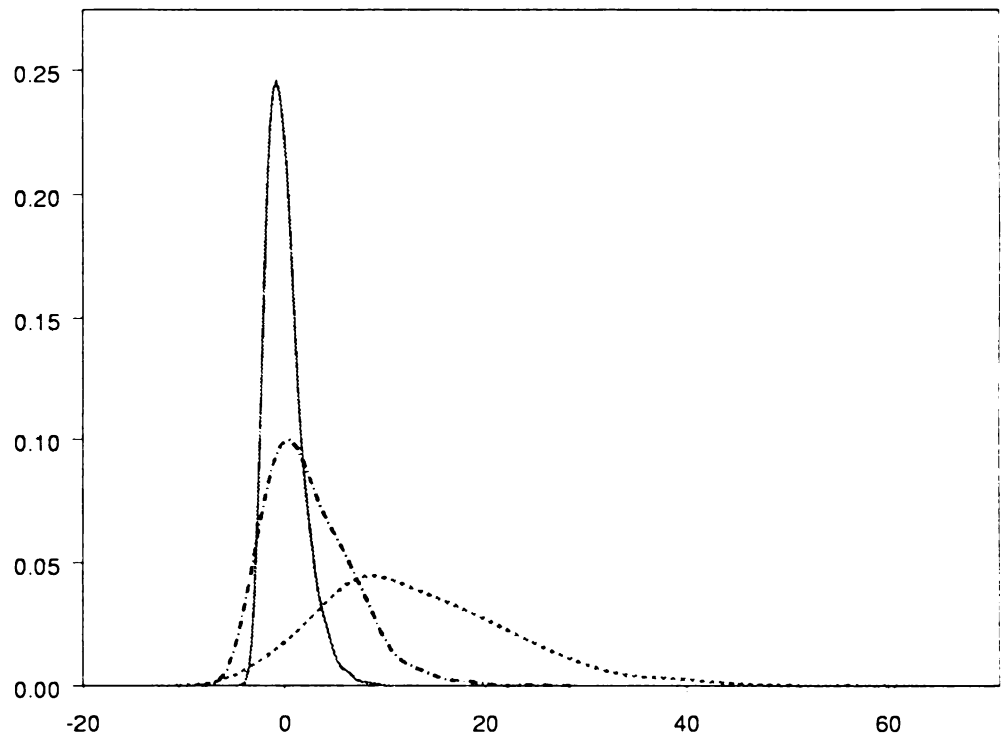


Figure 4.5: *The density of $T_n(\theta_n)$ under model 1 with $N(0, 0.1)$ errors.*

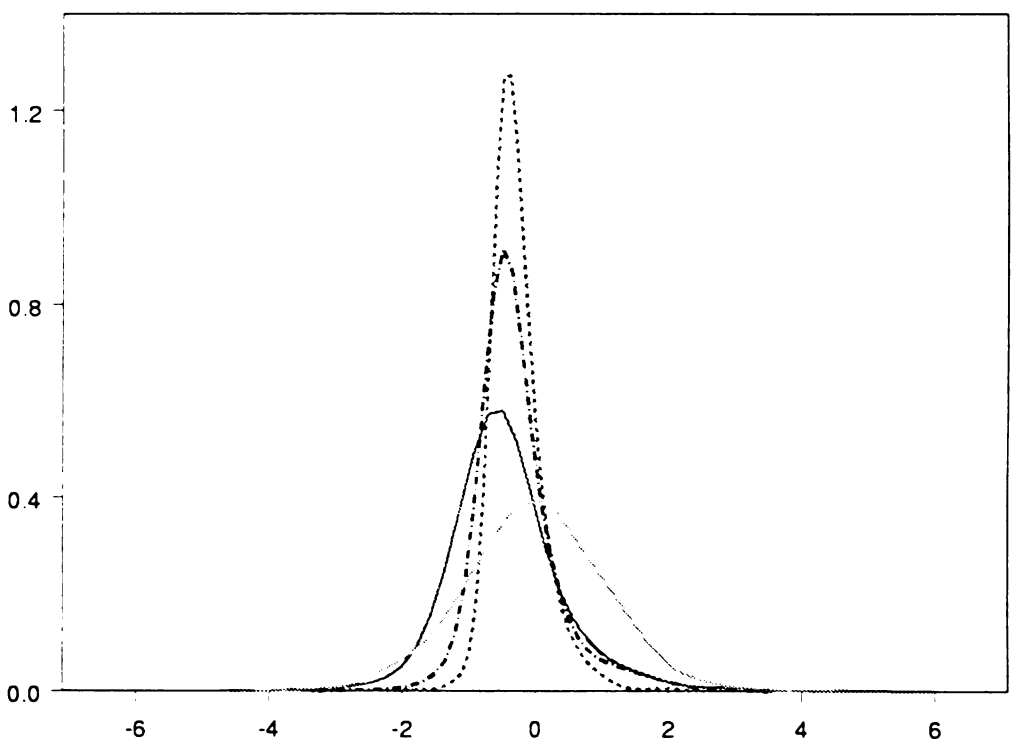


Figure 4.6: *The density of $T_n(\theta_n)$ under model 2 with $N(0, 0.1)$ errors.*

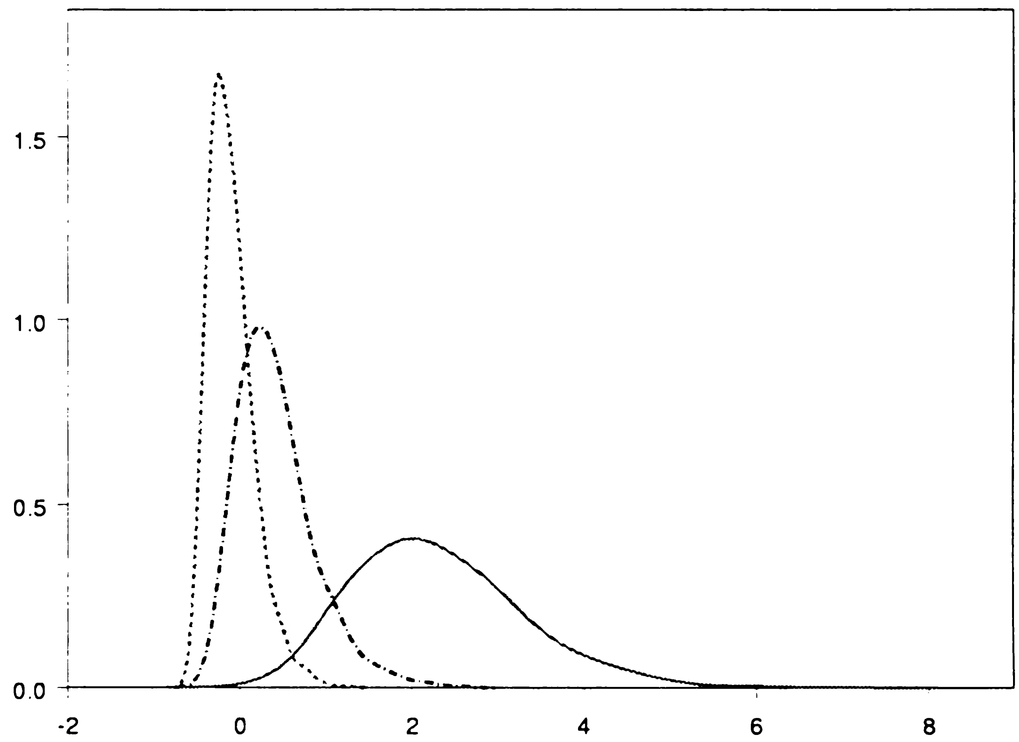


Figure 4.7: *The density of $T_n(\theta_n)$ under model 3 with $N(0, 0.1)$ errors.*

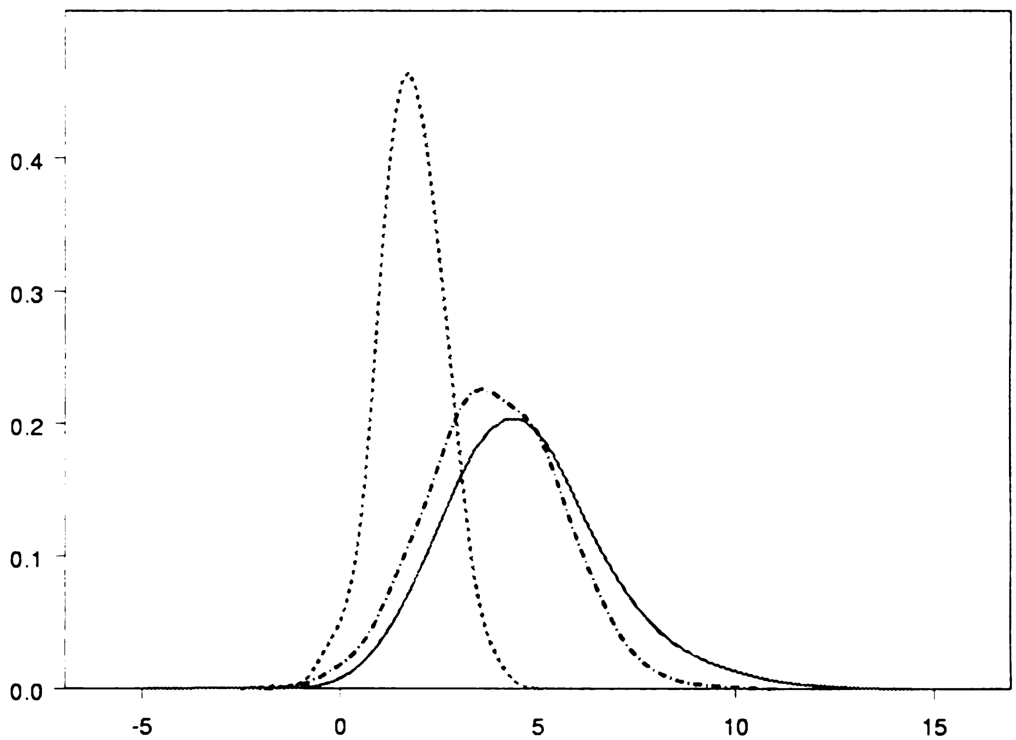


Figure 4.8: *The density of the suitably scaled minimized distance under model 1 with double exponential errors.*

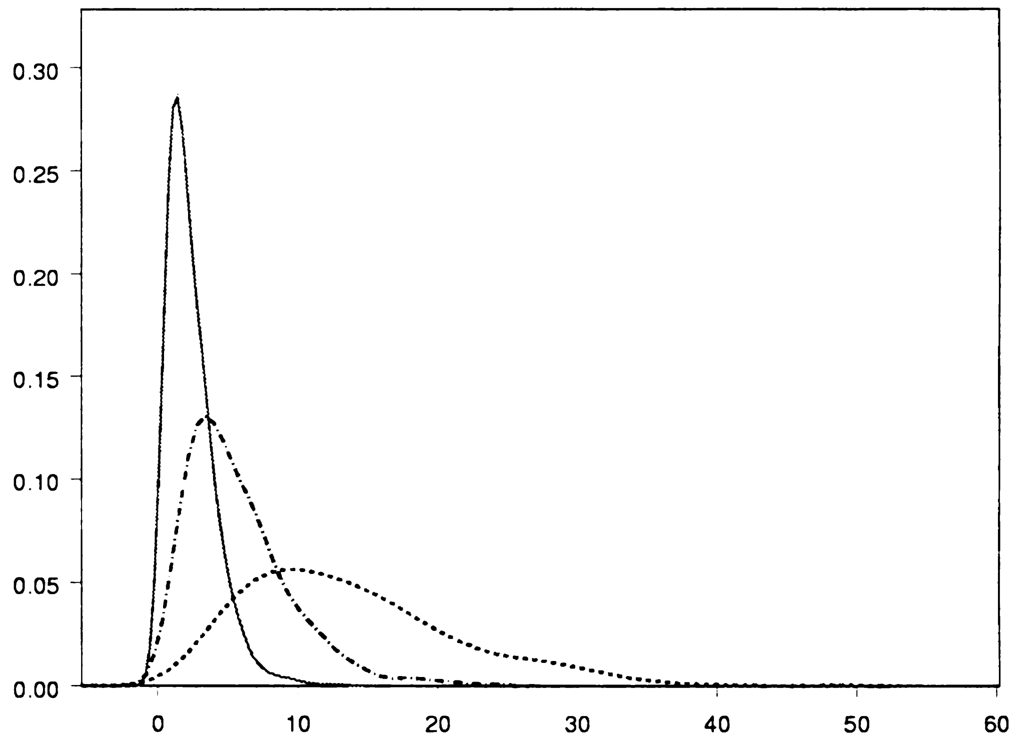


Figure 4.9: *The density of the suitably scaled minimized distance under model 2 with double exponential errors.*

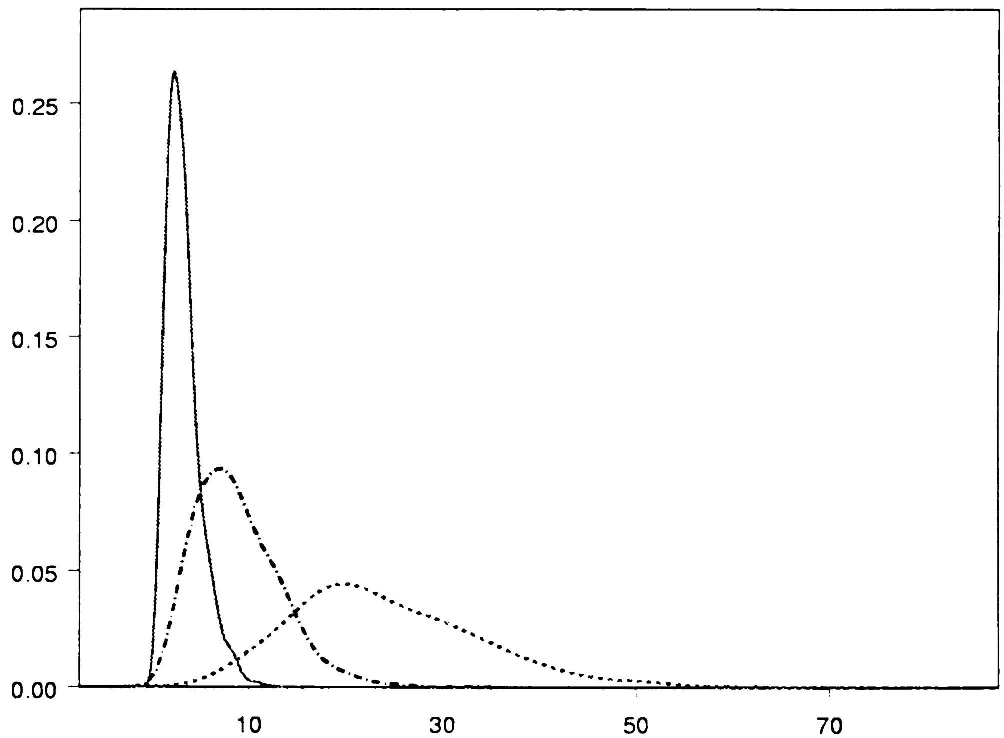


Figure 4.10: *The density of the suitably scaled minimized distance under model 1 with $N(0, 0.1)$ errors.*

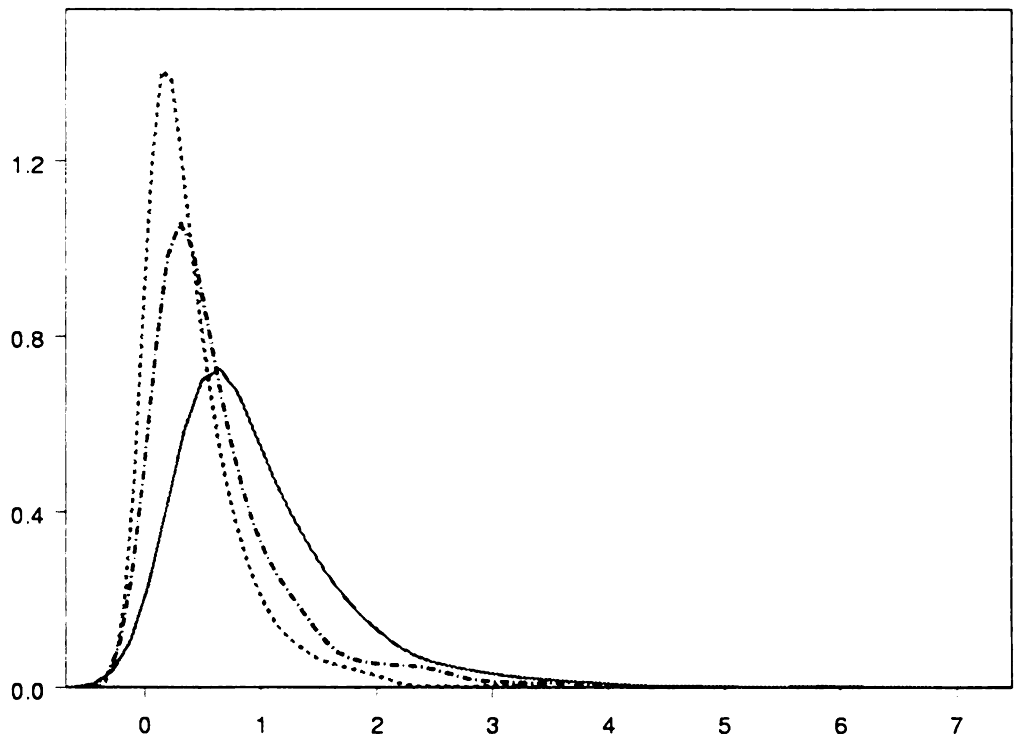


Figure 4.11: *The density of the suitably scaled minimized distance under model 2 with $N(0, 0.1)$ errors.*

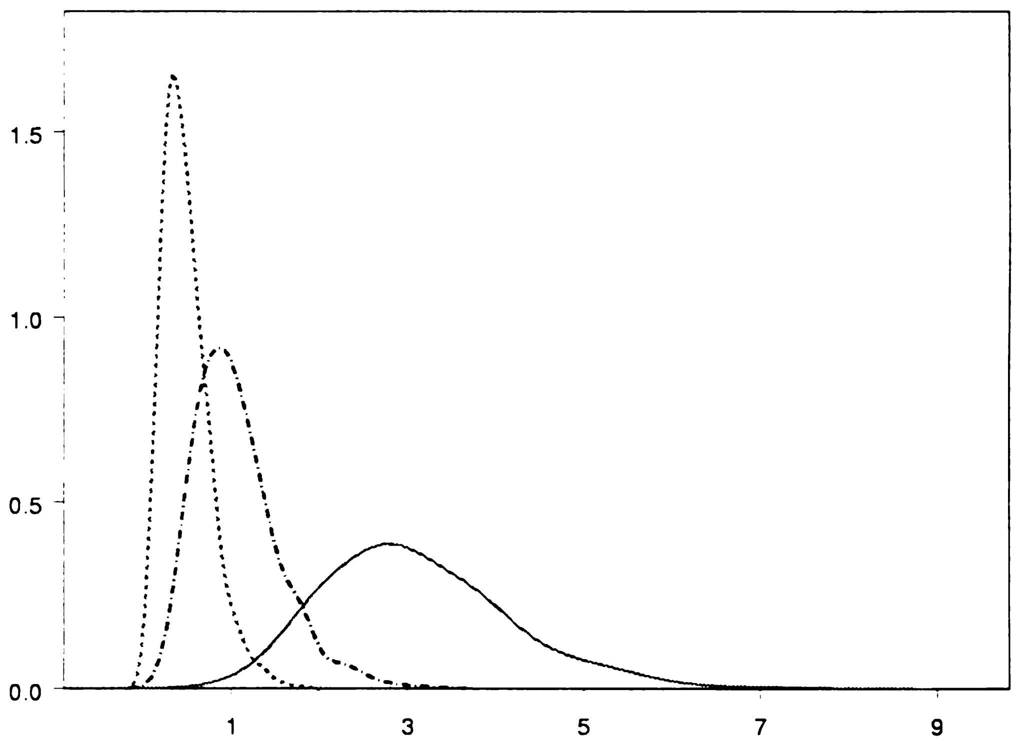
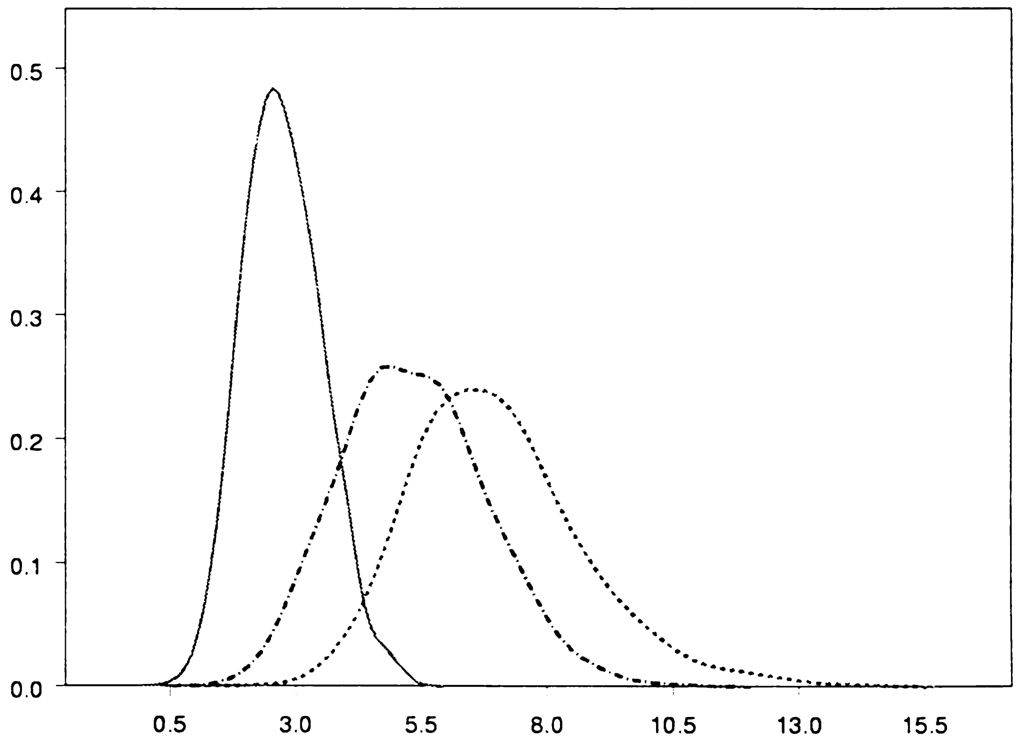


Figure 4.12: *The density of the suitably scaled minimized distance under model 3 with* $N(0, 0.1)$ *errors.*



BIBLIOGRAPHY

An Hong-zhi and Cheng Bing (1991). A Kolmogorov-Smirnov type Stgatitic with application to test for nonlinearity in time series , International Statistical Review, **59**, 287-307.

Beran, R.J. (1977). Minimum Hellinger distance estimates for parametric models, *Ann. Statist.*, **5**, 445-463.

Beran, R.J. (1978). An effiecient and adaptive robust estimator of location. *Ann. Statist.*, **6**, 292-313.

Bosq, D (1998). Nonparametric statistics for stochastic processes, Second Edition, Springer.

Cox, D., Koh, E., Wahba, G., and B. S. Yandell. (1988). Testing the parametric null model hypothesis in semiparametric partial and generalized spline models. *Ann. Statist.*, **16**, 113-119.

Donoho, D.L. and R. C. Liu. (1988a). The "automatic" robustness of minimum distance functionals. *Ann. Statist.*, **16**, 552-586.

Donoho, D.L. and R. C. Liu. (1988b). Pathologies of some minimum distance estimators. *Ann. Statist.*, **16**, 587-608.

Eubank, R.L. and Hart, J.D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.*, **20**, 1412-1425.

Eubank, R.L. and Hart, J.D. (1993). Commonality of CUSUM, von Neumann and smooth-ing based goodness-of-fit tests. *Biometrika*, **80**, 89-98.

Eubank, R.L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85**, no. 410, 387-392.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Mult. Analysis.*, **14**, 1-16.

Härdle, W. and E. Mammen. (1993). Comparing nonparametric versus parametric regression Fits. *Ann. Statist.*, **21**, 1926-1947.

Hjellvik, V, Yao, Q, and Tjϕstheim, Dag. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Inference.* **68**, no. 2, 295-321.

Koul, H and , Stute, W. (1999). Nonparametric model checks for time series. *Ann. Statist.*, **27**, 204-236.

Mack, Y.P. and Silverman, B.W. (1993). Weak and strong uniform consistency of kernel regression estimates. Z. Wahrscheinichkeitstheorie verw. Gebiete, 405-415.

McKeague, I. and Zhang, M. (1993). Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.*, **22**, 495-514.

Resnick, S. (1992). Adventures in stochastic processes, Birkhuser Boston, Inc., Boston, MA.

Stute, W. (1996). Nonparametric model checks for regression, *Ann. Statist.*, **25**, 613-641.

Stute, W., Thies, S. & L. Zhu. (1997). Model checks for regression: An innovation process approach. *Ann. Statist.*, **26**, 1916-1934.

Tweedie, R.L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing theory. In Papers in Probability, Statistics and Analysis (J.F.C. Kingman and G.E.H.Reuter, eds.)260-276. Cambridge Univ. Press.

Wolfowitz, J. (1953). Estimation by the minimum distance method. *Ann. Inst. Statist. Math.*, Tokyo **5**, 9–23.

Wolfowitz, J. (1954). Estimation by the minimum distance method in nonparametric stochastic difference equations. *Ann. Math. Statist.*, **25**, 203–217.

Wolfowitz, J. (1957). The minimum distance method. *Ann. Math. Statist.*, **28**, 75-88.