# Abstract


| | |
|---|---|
| Title of dissertation: | Two Goodness-of-Fit Tests for the Density Ratio Model |
| | Luquan Yu, Doctor of Philosophy, 2017 |
| Dissertation directed by: | Professor Benjamin Kedem<br>Department of Mathematics |



Under consideration are goodness-of-fit tests for the *density ratio model*. The model stipulates that the log-likelihood ratio of two unknown densities is of a known form which depends on finite dimensional parameters and a tilt function. We can derive the empirical distribution estimator $\tilde{G}$ from a reference sample, and the semi-parametric distribution estimator $\hat{G}$ under the density ratio model. Furthermore we can derive kernel density estimators $\tilde{g}$ and $\hat{g}$ corresponding to $\tilde{G}$ and $\hat{G}$ by choosing a bandwidth parameter. Goodness-of-fit test statistics can be constructed via the discrepancy between $\tilde{g}$ and $\hat{g}$ using Hellinger distance and a modification thereof. We propose two new test statistics by modifying the goodness-of-fit test statistics suggested by Bondell (2007)[4] and by Cheng and Chu (2004)[6]. Asymptotic results and limiting distributions are derived for both new test statistics, and the selections of the kernel and bandwidth are discussed. Monte-Carlo simulations show that the new test statistics improve the accuracy of the the goodness-of-fit test and that the limiting distributions of the new test statistics are more symmetric.

Key words: Density ratio model; Goodness-of-fit; Kernel density estimator; Hellinger distance; Bandwidth.

# Two Goodness-of-Fit Tests for the Density Ratio Model

by

Luquan Yu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professor Benjamin Kedem, Chair/Advisor
Professor Frank Alt, Dean's Representative
Professor Xin He
Dr. Myron Katzoff
Professor Paul Smith

# Dedication

To my parents

To my wife

# Acknowledgments

First and foremost, I'd like to express my deepest appreciation and respect to my academic advisor, Professor Benjamin Kedem, who have guided me through one of the most significant periods of my life. He has always made himself available for help and advice even during holiday breaks. His wisdom, knowledge and sense of humor inspired and motivated me all along. Only with his constant encouragement could I finish this dissertation. I benefit not only from the statistics I learned from him, but also from the way of analyzing and solving problems. He gave me the chance, and helped me out when I was in difficult and bad times. Thank you, Ben!!

I would also like to thank my committee members, Professor Francis Alt, Professor Xin He, Dr. Myron Katzoff and Professor Paul J.Smith for sparing their invaluable time to review the manuscript. I greatly appreciate their suggestions and advices to improve my work.

I would like to send my special thanks to Professor Eric Slud who gave great statistical courses including both the theoretical and the applied. These helped me building up my statistical background.

I would like to thank many people in the University of Maryland. Thanks are due to Ms. Regalado Celeste, Ms Hidalgo Haydee V. and Mr. William R. Schildknecht. from the Mathematics Department, and thanks are due to Jody Heckman-Bose from ISSS. They did provide a lot of help concerning many aspects during my study.

I would like to thank my friends, Xuhua He, Tingni Sun, Zhou Zhang and Peng

Zhao. They encouraged me and helped me at those times when I was low-spirited.

At last, I would like to acknowledge my families. I want to thank my parents and my sister for their unconditional love and support. And finally, I would like to thank my wife Qian Zhao. With her love and passion, I complete this milestone.

# Table of Contents

# List of Tables

List of Figures

# Chapter 1

# Literature Review

## 1.1 Introduction

Multiple-sample semiparametric problems have attracted attention over many years from both theoretical and practical points of view due to the fact that statisticians may collect data for the same or similar objects from different sources which refer to different distribution functions. Thus it is essential to build a system of distributions to analyze multiple-sample problems to derive the necessary statistical inference. To present the key ideas of this dissertation it is sufficient to focus on the two-sample system. Therefore we shall deal with systems of two distributions which represent samples from two sources. One distribution is called the ***reference*** distribution which refers to a reference sample, and the other one serves as the ***distortion*** or deviation from the reference distribution which refers to a distortion sample.

Since these two samples are observed and collected from the same or similar objects, it is believed that the log-likelihood ratio of the corresponding unknown densities, the ***reference*** and ***distortion***, is of a known form which depends on finite dimensional parameters and a tilt function. This form is called the ***Density Ratio Model*** (DRM).

## 1.2 Density Ratio Model

Consider the following two independent random samples:

$$
\begin{aligned}
&u_1, ..., u_{n_0} \text{ from density } g(x).\\
&z_1, ..., z_{n_1} \text{ from density } g_1(x) = \exp(\alpha + \beta' h(x))g(x).
\end{aligned}
\tag{1.1}
$$

We consider $U = \{u_1, ..., u_{n_0}\}$ as the reference sample corresponding to an unknown distribution function $G(x)$ and density $g(x)$, and let $Z = \{z_1, ..., z_{n_1}\}$ be the distortion sample with unknown distribution function $G_1(x)$ and unknown density $g_1(x)$. $\alpha$ is an unknown scalar, $\beta$ is a $p \times 1$ unknown vector parameters, and $h(x)$, called the ***tilt*** function or distortion function, is a $p \times 1$ vector that consists of functions of $x$. We denote by $T = \{t_1, ..., t_n\} = \{u_1, ..., u_{n_0}, z_1, ..., z_{n_1}\}$, the combined or fused sample of both reference and distortion samples consisting of $n = n_0 + n_1$ observations.

Many researchers have made significant contributions to various aspects of the density ratio model, such as kernel density estimation (Fokianos 2004[10], Cheng and Chu 2004[5], Qin and Zhang 2005[28], Bondell 2005[3], Wu *et al.* 2010[34], Wu *et al.* 2012[35]), analysis of variance (Fokianos *et al.* 2001[9]), case-control studies (Prentice and Pyke 1979[26]), cluster detection (Wen and Kedem 2009[32]), regression analysis (Voulgaraki *et al.* 2012[31]), mortality rate prediction (Kedem *et al.* 2008[21]), out-of-sample fusion (Zhou 2013[41], Katzoff *et al.* 2014[18], Kedem *et al.* 2016[23], Kedem *et al.* 2017[20]) and goodness-of-fit tests (Cheng and Chen 2004[5], Cheng and Chu 2004[6], Bondell 2007[4], Zhang 1999, 2000, 2001, 2002 [37] [38] [39] [40], Xu and Wang 2011[36]).

## 1.3 Goodness-of-Fit Test for the Density Ratio Model

When we assume the two samples satisfy the density ratio model as in (1.1), goodness-of-fit tests are needed to justify and support the assumed model. The null hypothesis is:

$H_0$ : The two samples satisfy the DRM (1.1) with correctly specified tilt function.

Let $\hat{G}$ be the semiparametric estimate of underlying distribution $G$ obtained from the fused sample $T = \{t_1, ..., t_n\} = \{u_1, ..., u_{n_0}, z_1, ..., z_{n_1}\}$ under the DRM. Let $\tilde{G}$ be the empirical distribution function obtained from the reference sample $U = \{u_1, ..., u_{n_0}\}$ only. Qin and Zhang (1997)[27] suggested a Kolmogorov-Smirnov type goodness-of-fit test statistic for the logistic regression model based on case-control data, which is equivalent to density ratio model (1.1). Their statistic measures the discrepancy between $\hat{G}$ and $\tilde{G}$ by

$$\Delta_n = \sup_t \sqrt{n}|\hat{G}(t) - \tilde{G}(t)|. \tag{1.2}$$

Unfortunately there is no analytic expression for the distribution of $\Delta_n$. They presented a bootstrap procedure to approximate the critical values of this test statistic. Zhang (2000)[38] extended the Kolmogorov-Smirnov type test statistic to test the validity of a multiplicative-intercept risk model and presented a bootstrap procedure for approximating the $p$-value of the proposed test. Zhang (2002)[40] tested the validity of the generalized logit model by a weighted Kolmogorov-Smirnov type statistic and still needed a bootstrap procedure to approximate the $p$-values.

Zhang (1999)[37] proposed a chi-squared statistic to test the validity of the model. He distributed the fused sample data in a finite number of mutually exclusive intervals to derive a quadratic form in terms of the deviations between the cell probabilities obtained from the reference sample $U$ and from the fused sample $T$. The test statistic has an asymptotic chi-squared distribution and thus $p$-values could be obtained directly without employing the bootstrap method to evaluate the critical values. Zhang (2001)[39] constructed a Wald-type statistic by extending the information matrix test of White (1982)[33]. This Wald-type statistic is called the information-matrix-based goodness-of-fit statistic and it has an asymptotic chi-squared distribution. Both the chi-squared statistic and Wald-type statistic do not require the bootstrap procedure to derive $p$-values but they require a high-dimensional matrix inversion. Xu and Wang (2011)[36] developed a test procedure based on Zhang (2001)[39]. Their procedure can simultaneously test the validity of the model and also correct the bias of parameter estimators.

Bondell (2007)[4] presented a goodness-of-fit test by constructing a test statistic via the integrated discrepancy between two competing kernel density estimators in a bounded interval. He constructed the kernel density estimators in two ways:

$$\hat{g}(t) = \int K_b(t - x) d\hat{G}(x) \tag{1.3}$$

$$\tilde{g}(t) = \int K_b(t - x) d\tilde{G}(x) \tag{1.4}$$

where $K_b(\cdot)$ is a kernel with bandwidth $b$. Then he defined the test statistic as

$$I_n^B = n \int_{-L}^{L} \big(\hat{g}(t) - \tilde{g}(t)\big)^2 dt. \tag{1.5}$$

He proved that the test statistic $I_n^B$ tends in distribution to an infinite wighted sum of independent chi-square variables. However, it is not easy to derive the asymptotic distribution of $I_n^B$. Hence, he used a bootstrap method similar to that of Qin and Zhang (1997)[27].

Cheng and Chu (2004)[6] proposed a goodness-of-fit test statistic similar to (1.5)

$$J_n^C = \int \left(\hat{g}(t) - \tilde{g}(t)\right)^2 dt. \tag{1.6}$$

Furthermore, they proved that under certain conditions, the limiting distribution of $n\sqrt{b} \cdot J_n^C$ is normal. Therefore they constructed the goodness-of-fit test without the bootstrap procedure. In this dissertation, we shall study the Hellinger distance, a modification of (1.5) and (1.6), which has a certain advantage.

## 1.4 Hellinger Distance

In probability and statistics, the ***Hellinger distance*** is used to quantify the similarity between two probability distributions. Let $f(x)$ and $g(x)$ denote two probability density functions. Then the square of the Hellinger distance between $f$ and $g$ is defined as

$$H^2(f,g) = \frac{1}{2}\int (\sqrt{f(x)} - \sqrt{g(x)})^2 dx.$$

Since $H^2(f,g) \geq 0$ and

$$\begin{aligned} H^2(f,g) &= \frac{1}{2}\int (\sqrt{f(x)} - \sqrt{g(x)})^2 dx \\ &= 1 - \int \sqrt{f(x)g(x)} dx \leq 1, \end{aligned}$$

we have

$$0 \leq H^2(f,g) \leq 1.$$

The Hellinger distance measures the discrepancy between two probability density distributions. Wu, Krunamuni and Zhang (2010)[34] proposed to derive a parametric estimator in the density ratio model by minimizing the Hellinger distance between the semiparametric density estimator and nonparametric density estimator. Wu and Karunamni (2012)[35] investigated the asymptotic properties of the parametric estimators of the model, including consistency, asymptotic normality and efficiency. Zhu, Wu and Lu (2013)[42] extended the minimum Hellinger distance estimation to right-censored survival data.

## 1.5 Organization of this Dissertation

In this dissertation we propose two new goodness-of-fit test statistics based on the Hellinger distance. We begin with general semiparametric inference of the density ratio model, including the derivation of the maximum likelihood estimator of the underlying distribution $G$, and some of its asymptotic properties. We describe the two new goodness-of-fit test statistics, derive their asymptotic limiting distributions, study them by Monte Carlo simulation, and apply them in real data analysis.

This dissertation is organized as follows. Chapter 2 is an introduction to the semiparametric density ratio model, including the procedure to derive the maximum likelihood estimators of the parameters and underlying reference distribution function. At the end of Chapter 2, we introduce some asymptotic results, including the consistency and asymptotic distribution of the estimators. Chapter 3 proposes $I_n$, a new goodness-of-fit test statistic to test the validity of the density ratio model. The limiting distribution of the new test statistic under the null hypothesis is derived. A bootstrap procedure is applied in conjunction with Monte Carlo simulations. Chapter 4 proposes another new goodness-of-fit test statistic $J_n$. Under the null hypothesis which is model (1.1), the asymptotic limiting distribution of $J_n$ is derived. Furthermore the method to obtain the $p$-value corresponding to the test statistic $J_n$ is derived also. Chapter 5 provides simulation studies for $J_n$. Computer-generated samples from the normal distribution, the gamma distribution and the lognormal distribution are used to perform goodness-of-fit tests when the tilt function is both correctly specified and also misspecified. Chapter 6 applies the goodness-of-fit test developed in Chapter 3 and Chapter 4 in the analysis of experimental radar reflectivity data. Chapter 7 summarizes this dissertation.

# Chapter 2

# The Density Ratio Model

In this chapter we introduce a semiparametric statistical model which is called density ratio model. A profiling procedure, following Qin and Zhang (1997)[27], to derive the maximum likelihood estimation of parameters and underlying reference distribution, is introduced. The asymptotic results for the estimators which were derived by Qin and Zhang (1997)[27] and Lu (2007)[25] are given in the last section of this chapter.

## 2.1 Introduction

Model (1.1) is a biased sampling model with weights depending on parameters. Vardi (1982)[30] studied a related nonparametric two-sample estimation problem. Consider a sample from a distribution function $G$ and another sample from $F_G$, the ***length-biased distribution*** of $G$, such that

$$F_G(y) = \frac{1}{\mu} \int_0^y x dG(x) \tag{2.1}$$

where $\mu = \int_0^\infty x dG(x) < \infty$. The nonparametric maximum likelihood estimation (NPMLE) for $G$ and its asymptotic properties were discussed in Vardi (1982)[30]. Gill, Vardi and Wellner(1988)[15] proved the consistency and asymptotic normality of Vardi's NPMLE.

Gilbert *et al.*(1999) extended the biased sampling problem model (2.1) to a semiparametric problem with $G$ unspecified. The model is

$$F_G(y, \theta) = \frac{1}{W(\theta, G)} \int_{-\infty}^{y} w(x, \theta) dG(x) \tag{2.2}$$

where $W(\theta, G) = \int_{-\infty}^{\infty} w(x, \theta) dG(x) < \infty$. They showed that under the condition that the biasing function $w(x, \theta)$ is known, the semiparametric biased sampling problem is identifiable. The large sample behavior of the semiparametric MLE was investigated by Gilbert (2000)[12], and goodness-of-fit test statistics of Cramér-von Mises type, Anderson-Darling type, and Kolmogorov-Smirnov type were studied by Gilbert (2004)[13].

## 2.2 Semiparametric Density Ratio Model

We are considering semiparametric inference from DRM by assuming that the log-likelihood ratio of two unknown densities has a known linear form which depends on unknown finite dimensional parameters. The DRM is equivalent to a generalized logistic regression model in case-control sampling (Qin and Zhang (1997)[27]).

Consider model (1.1):

$u_1, ..., u_{n_0}$ is a random sample with reference density $g(x)$.

$z_1, ..., z_{n_1}$ is a random sample with density $g_1(x) = \exp(\alpha + \beta' h(x)) g(x)$.

Here $g(x)$ and $g_1(x)$ are unknown probability density functions, $\alpha$ is an unknown scalar, $\beta$ is a $p \times 1$ vector parameter, and $h(x)$, which is called the distortion or tilt function, is a $p \times 1$ vector that consists of known functions of $x$. Let $U = \{u_1, ..., u_{n_0}\}$ is the reference sample corresponding to an unknown distribution function $G(x)$, the cdf corresponding to $g(x)$, and let $Z = \{z_1, ..., z_{n_1}\}$ is the distortion sample with unknown density $g_1(x)$. Let $T = \{t_1, ..., t_n\} = \{u_1, ..., u_{n_0}, z_1, ..., z_{n_1}\}$ be the combined or fused sample consisting of both the reference and distortion samples of $n = n_0 + n_1$ observations.

Consider the following well-known examples of DRM.

**Example 1** (Normal distribution)**.** Assume that $U \sim N(\mu_1, \sigma_1^2)$ with density $g(\cdot)$ and $Z \sim N(\mu_2, \sigma_2^2)$ with density $g_1(\cdot)$. Then ratio of the densities is

$$\frac{g_1(x)}{g(x)} = \exp\left\{ \log\left(\frac{\sigma_1}{\sigma_2}\right) + \left(\frac{\mu_1^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2}\right) + \left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\ ,\ \frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right)\left(\begin{array}{c} x \\ \\ x^2 \end{array}\right)\right\}.$$

This is a special case of model (1.1) with tilt function $h(x) = (x, x^2)'$ and parameters

$$\begin{aligned} \alpha &= \log\left(\frac{\sigma_1}{\sigma_2}\right) + \left(\frac{\mu_1^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2}\right) \\ \beta &= \left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\ ,\ \frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right)'. \end{aligned}$$

Note that when $\mu_1 = \mu_2 = 0$, $h(x) = x^2$.

**Example 2** (Gamma distribution)**.** Assume that $U \sim Gamma(\alpha_1, \beta_1)$ with density $g(\cdot)$ and $Z \sim Gamma(\alpha_2, \beta_2)$ with density $g_1(\cdot)$. Then ratio of the densities is

$$\frac{g_1(x)}{g(x)} = \exp\left\{ \log\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_2)} + (\alpha_1 \log\beta_1 - \alpha_2 \log\beta_2) + \left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\ ,\ \alpha_2 - \alpha_1\right)\left(\begin{array}{c} x \\ \\ \log x \end{array}\right)\right\}.$$

Again this is a special case of model (1.1) with tilt function $h(x) = (x, \log x)'$, and parameters

$$\begin{aligned} \alpha &= \log\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_2)} + (\alpha_1 \log\beta_1 - \alpha_2 \log\beta_2) \\ \beta &= \left(\frac{1}{\beta_1} - \frac{1}{\beta_2}\ ,\ \alpha_2 - \alpha_1\right)'. \end{aligned}$$

Note that when $\alpha_1 = \alpha_2$, $h(x) = x$, and when $\beta_1 = \beta_2$, $h(x) = \log x$.

**Example 3** (Lognormal distribution)**.** Assume that $U \sim Lognormal(\mu_1, \sigma_1^2)$ with density $g(\cdot)$ and $Z \sim Lognormal(\mu_2, \sigma_2^2)$ with density $g_1(\cdot)$. Then ratio of the densities is

$$\frac{g_1(x)}{g(x)} = \exp\left\{ \log\left(\frac{\sigma_1}{\sigma_2}\right) + \left(\frac{\mu_1^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2}\right) + \left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\ ,\ \frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right) \begin{pmatrix} \log x \\ \\ (\log x)^2 \end{pmatrix} \right\}.$$

This is a special case of model (1.1) with tilt function $h(x) = (\log x, (\log x)^2)'$, and the parameters

$$\alpha = \log\left(\frac{\sigma_1}{\sigma_2}\right) + \left(\frac{\mu_1^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2}\right)$$
$$\beta = \left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\ ,\ \frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right)'.$$

Note that when $\mu_1 = \mu_2 = 0$, $h(x) = (\log x)^2$, and when $\sigma_1 = \sigma_2$, $h(x) = \log x$.

In Chapter 3 and Chapter 5, Monte Carlo simulation studies for our test statistics are applied by using computer-generated samples from normal, gamma and lognormal distributions. Correctly specified tilt functions in the previous three examples will be used in goodness-of-fit tests. And based on these examples, we also can intentionally choose misspecified tilt functions to test our goodness-of-fit procedures.

## 2.3 Maximum Likelihood Estimation

In this chapter we follow the main results from Qin and Zhang (1997)[27] and Lu (2007)[25] and our contributions will be in the following chapters, and the difference being that we deal with pdf's whereas Qin and Zhang (1997)[27] and Lu (2007)[25] deal with cdf's.

A maximum likelihood estimator of $G$ can be obtained by maximizing the empirical likelihood over the class of discrete cumulative distribution functions with jumps at all the observed values $\{t_1, ..., t_n\} = \{u_1, ..., u_{n_0}, z_1, ..., z_{n_1}\}$ from the combined sample. Let $p_i = dG(t_i)$ denote the size of the jump at the observed value $t_i$. Then the empirical likelihood is defined as as follows:

$$\begin{aligned} \mathscr{L}(\alpha, \beta, G) &= \prod_{i=1}^{n_0} dG(u_i) \prod_{j=1}^{n_1} \exp\big(\alpha + \beta' h(z_j)\big) dG(z_j) \\ &= \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \exp\big(\alpha + \beta' h(z_j)\big) \\ &= \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} w(z_j) \end{aligned} \tag{2.3}$$

where

$$w(t) = \exp\big(\alpha + \beta' h(t)\big). \tag{2.4}$$

To maximize the empirical likelihood, we follow a profiling procedure by which first the $p_i$ are optimized in terms of $\alpha, \beta$, and then the $p_i$ are substituted back into the likelihood to obtain a function of $\alpha, \beta$ only.

When $\alpha, \beta$ are fixed, the empirical likelihood (2.3) is optimized by maximizing the product term $\prod_{i=1}^{n} p_i$ subject to the following constraints

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i[w(t_i) - 1] = 0. \tag{2.5}$$

The constraints (2.5) simply express the fact that the discrete reference probability masses and their distortions sum each to 1.

Lagrange multipliers are used in the maximization of $\prod_{i=1}^{n} p_i$. This is equivalent to maximizing

$$\sum_{i=1}^{n} \log p_i + \lambda \left[ \sum_{i=1}^{n} p_i - 1 \right] - \lambda_1 \left[ \sum_{i=1}^{n} p_i (w(t_i) - 1) \right]$$

where $\lambda$, $\lambda_1$ are Lagrange multipliers. Differentiating with respect to $p_i$ and equating to 0 gives,

$$\frac{1}{p_i} + \lambda_0 - \lambda_1 (w(t_i) - 1) = 0 \tag{2.6}$$

or

$$1 + \lambda_0 p_i - \lambda_1 p_i [w(t_i) - 1] = 0.$$

Sum up over $i = 1, ..., n$ and apply the constraints (2.5). Then we have

$$n + \lambda_0 = 0.$$

Substitute $\lambda_0 = -n$ into (2.6), and defining $\rho = n_1/n_0$, we derive

$$p_i = \frac{1}{n_0} \cdot \frac{1}{1 + \rho \exp(\alpha + \beta' h(t_i))}.$$

Substituting the $p_i$'s back into the likelihood, we obtain the log-likelihood function

$$\ell(\alpha, \beta) = \sum_{j=1}^{n_1} (\alpha + \beta' h(z_j)) - \sum_{i=1}^{n} \log(1 + \rho \exp(\alpha + \beta' h(t_i))). \tag{2.7}$$

We get $(\hat{\alpha}, \hat{\beta})$ by solving two score equations,

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= n_1 - \sum_{i=1}^{n} \frac{\rho \exp(\alpha + \beta' h(t_i))}{1 + \rho \exp(\alpha + \beta' h(t_i))} = 0 \\ \frac{\partial \ell}{\partial \beta} &= \sum_{j=1}^{n_1} h(z_j) - \sum_{i=1}^{n} \frac{\rho \exp(\alpha + \beta' h(t_i)) h(t_i)}{1 + \rho \exp(\alpha + \beta' h(t_i))} = 0, \end{aligned} \tag{2.8}$$

and therefore,

$$\hat{p}_i = \frac{1}{n_0} \cdot \frac{1}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))}. \tag{2.9}$$

Consequently, the MLE of the distribution function from the combined data $t_1, ..., t_n$ under DRM is

$$\hat{G}(t) = \frac{1}{n_0} \sum_{i=1}^{n} \frac{I(t_i \leq t)}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))}. \tag{2.10}$$

## 2.4 Asymptotic Results

Assume that the true parameters are $(\alpha_0, \beta_0)$.

Define

$$A = \begin{pmatrix} \int \frac{\exp(\alpha_0 + \beta_0' h(y))}{1 + \rho \exp(\alpha_0 + \beta_0' h(y))} dG(y) & \int \frac{h(y) \exp(\alpha_0 + \beta_0' h(y))}{1 + \rho \exp(\alpha_0 + \beta_0' h(y))} dG(y) \\ \int \frac{h(y) \exp(\alpha_0 + \beta_0' h(y))}{1 + \rho \exp(\alpha_0 + \beta_0' h(y))} dG(y) & \int \frac{h(y)' h(y) \exp(\alpha_0 + \beta_0' h(y))}{1 + \rho \exp(\alpha_0 + \beta_0' h(y))} dG(y) \end{pmatrix} \tag{2.11}$$

and

$$\mathbf{S} = \frac{\rho}{1 + \rho} A \tag{2.12}$$

The asymptotic properties of $(\alpha, \beta)$ are derived by Qin and Zhang (1997)[27] and Lu (2007)[25] as follows:

**Lemma 1** (Qin & Zhang (1997), Lu (2007))**.** *Under certain regularity conditions, if model (1.1) is true, then the asymptotic expansion of* $(\hat{\alpha}, \hat{\beta})$ *is*

$$\begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} = \frac{1}{n} \mathbf{S}^{-1} \begin{pmatrix} \frac{\partial l(\alpha_0, \beta_0)}{\partial \alpha} \\ \frac{\partial l(\alpha_0, \beta_0)}{\partial \beta} \end{pmatrix} + o_p(\frac{1}{\sqrt{n}}) \tag{2.13}$$

*and*

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \xrightarrow{d} \mathbf{N}(0, \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}) \tag{2.14}$$

*where*

$$\mathbf{V} = \frac{\rho}{1+\rho}A - \rho \begin{pmatrix} \int \frac{\exp(\alpha_0 + \beta_0' h(y))}{1+\rho\exp(\alpha_0+\beta_0' h(y))} dG(y) \\ \int \frac{h(y)\exp(\alpha_0 + \beta_0' h(y))}{1+\rho\exp(\alpha_0+\beta_0' h(y))} dG(y) \end{pmatrix}$$
$$\times \begin{pmatrix} \int \frac{\exp(\alpha_0 + \beta_0' h(y))}{1+\rho\exp(\alpha_0+\beta_0' h(y))} dG(y) & \int \frac{h(y)\exp(\alpha_0 + \beta_0' h(y))}{1+\rho\exp(\alpha_0+\beta_0' h(y))} dG(y) \end{pmatrix}. \tag{2.15}$$

# Chapter 3

# A Test Statistic

In this chapter we propose a test statistic constructed from the Hellinger distance, which is used to measure the discrepancy between two competing probability densities. We start by defining these two competing density estimators as kernel density estimators. One kernel density estimator is derived from the reference sample $U$ and another one is derived from the fused sample $T$. Next our test statistic is decomposed into two components. We derive the limiting distribution of one component and render the value of the other component relatively small by choosing an appropriate bandwidth. Therefore this gives a numerical approximation for our test statistic. A bootstrap procedure is used to obtain an approximation to this distribution of the test statistic.

## 3.1 Kernel Density Estimator

For a given kernel, $K(x) \geq 0$, with $\int K(x)dx = 1$ and $\int K^2(x)dx < \infty$, by using the empirical distribution $G_n$, we can construct a kernel density estimate as a convolution

$$\int K_b(x-y)dG_n(y)$$

where $b$ is the bandwidth and $K_b(\cdot) = (1/b)K(\cdot/b)$.

The empirical distribution of the reference sample $u_1, ..., u_{n_0}$ is

$$\tilde{G}(t) = \frac{1}{n_0} \sum_{i=1}^{n_0} I(u_i \leq t). \tag{3.1}$$

From (2.10) and (3.1), we get two competing estimators for $G(t)$: $\hat{G}(t)$ as derived from the combined sample $T = \{t_1, ..., t_n\} = \{u_1, ..., u_{n_0}, z_1, ..., z_{n_1}\}$, and $\tilde{G}(t)$ as derived from the reference sample $\{u_1, ..., u_{n_0}\}$ only.

We can now construct density estimators by using these two competing estimators as follows:

$$\begin{aligned} \hat{g}(t) &= \int K_b(t-y) d\hat{G}(y) \\ \tilde{g}(t) &= \int K_b(t-y) d\tilde{G}(y). \end{aligned} \tag{3.2}$$

**Lemma 2.** *Let $\hat{g}(x)$ and $\tilde{g}(x)$ be the kernel density estimators defined in (3.2). Then*

$$\begin{aligned} \hat{g}(t) &= \frac{1}{n_0} \sum_{i=1}^{n} \frac{1}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))} K_b(t - t_i) \\ \tilde{g}(t) &= \frac{1}{n_0} \sum_{i=1}^{n_0} K_b(t - u_i). \end{aligned} \tag{3.3}$$

*Proof.* Let $\delta(x)$ be the Dirac delta function. So, for any function $f(x)$

$$\int f(x) \delta(x - t) dx = f(t). \tag{3.4}$$

By (2.10) and (3.1), we have

$$\begin{aligned} \hat{G}(t) &= \frac{1}{n_0} \sum_{i=1}^{n} \frac{I(t_i \leq t)}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))} \\ \tilde{G}(t) &= \frac{1}{n_0} \sum_{i=1}^{n_0} I(u_i \leq t)). \end{aligned}$$

So

$$\begin{aligned}
\hat{g}(t) &= \int K_b(t-x)d\hat{G}(x) \\
&= \int K_b(t-x)\sum_{i=1}^{n}\hat{p}_i\delta(x-t_i)dx \\
&= \frac{1}{n_0}\sum_{i=1}^{n}\frac{K_b(t-t_i)}{1+\rho\exp(\hat{\alpha}+\hat{\beta}'h(t_i))}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{g}(t) &= \int K_b(t-x)d\tilde{G}(x) \\
&= \int K_b(t-x)\sum_{i=1}^{n_0}\frac{1}{n_0}\delta(x-u_i)dx \\
&= \frac{1}{n_0}\sum_{i=1}^{n_0}K_b(t-u_i).
\end{aligned}$$

□

## 3.2 A New Test Statistic

Bondell (2007)[4] presented a goodness-of-fit test by constructing a test statistic via the discrepancy between two competing kernel density estimators. He constructed the kernel density estimators $\hat{g}$ as (1.3) and $\tilde{g}$ as (1.4) and fixing the bandwidth at $b = 1$. Then he defined the test statistic $I_n^B$ as (1.5). He proved that under $H_0$ the test statistic $I_n^B$ tends in distribution to an infinite weighted linear combination of independent chi-square variables with one degree of freedom each.

In this chapter we define a new test statistic as the Hellinger distance which measures the discrepancy between the two competing density estimators, $\hat{g}$ and $\tilde{g}$. The associated integrated squared error is obtained in a closed and bounded interval

$[-L, L]$. Through this dissertation, we use many integrals with certain limits. We assume that our data are such that the integrals beyond the limits are negligible.

We define a new goodness-of-fit statistic in terms of the Hellinger distance

$$I_n \equiv nb \int_{-L}^{L} \left( \sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)} \right)^2 dt. \tag{3.5}$$

It is convenient to define

$$W_n(t) \equiv \frac{\sqrt{nb}}{2\sqrt{g(t)}} \Big( \hat{g}(t) - \tilde{g}(t) \Big) \tag{3.6}$$

and

$$\begin{aligned}
D_n(t) \equiv\ & nb\Big( \sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)} \Big)^2 - W_n(t)^2 \\
=\ & \frac{nb}{\big(\sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}\big)^2} \big(\hat{g}(t) - \tilde{g}(t)\big)^2 - \frac{nb}{4g(t)} \big(\hat{g}(t) - \tilde{g}(t)\big)^2 \\
=\ & nb\big(\hat{g}(t) - \tilde{g}(t)\big)^2 \frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)})^2(\sqrt{g(t)} + \sqrt{\tilde{g}(t)})} (g(t) - \tilde{g}(t)) \\
& + nb\big(\hat{g}(t) - \tilde{g}(t)\big)^2 \frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)})^2(\sqrt{g(t)} + \sqrt{\hat{g}(t)})} (g(t) - \hat{g}(t)),
\end{aligned} \tag{3.7}$$

then

$$I_n = \int_{-L}^{L} W_n(t)^2 dt + \int_{-L}^{L} D_n(t) dt. \tag{3.8}$$

In the following sections we assume that the bandwidth $b$ is fixed at some properly chosen values. A similar fixed bandwidth approach can be found in Anderson et al. (1994)[1]. We will prove that $W_n(t)$ converges weakly to $W(t)$, a Gaussian process with mean 0. In order to approximate $W_n(t)$, we need to derive an approximation of $\hat{g}(t)$ first. Later we shall show that $\int_{-L}^{L} D_n(t)dt$ is very small.

## 3.3 An Approximation of $\hat{g}(t)$

We would like to obtain the Taylor expansion of $\hat{g}(t)$. For this purpose, define

$$H_1(t;\alpha,\beta) = \frac{1}{n_0}\sum_{i=1}^{n}\frac{K_b(t-t_i)}{1+\rho\exp(\alpha+\beta' h(t_i))} \tag{3.9}$$

Differentiate $H_1(t;\alpha,\beta)$ with respect to $\alpha$ and $\beta$ respectively.

$$\begin{aligned}\frac{\partial H_1(t;\alpha,\beta)}{\partial\alpha} &= -\frac{1}{n_0}\sum_{i=1}^{n}\frac{\rho\exp(\alpha+\beta' h(t_i))K_b(t-t_i)}{\big(1+\rho\exp(\alpha+\beta' h(t_i))\big)^2}\\ \frac{\partial H_1(t;\alpha,\beta)}{\partial\beta} &= -\frac{1}{n_0}\sum_{i=1}^{n}\frac{\rho\exp(\alpha+\beta' h(t_i))h(t_i)K_b(t-t_i)}{\big(1+\rho\exp(\alpha+\beta' h(t_i))\big)^2}.\end{aligned} \tag{3.10}$$

Taking the expectation at $(\alpha_0,\beta_0)$, we have

$$\begin{aligned}E\Big(\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial\alpha}\Big) &= -\frac{1}{n_0}E\Big(\sum_{i=1}^{n}\frac{\rho\exp(\alpha_0+\beta_0' h(t_i))K_b(t-t_i)}{\big(1+\rho\exp(\alpha_0+\beta_0' h(t_i))\big)^2}\Big)\\ &= -\frac{1}{n_0}\Big[n_0\int\frac{\rho\exp(\alpha_0+\beta_0' h(y))K_b(t-y)}{\big(1+\rho\exp(\alpha_0+\beta_0' h(y))\big)^2}dG(y)\\ &\qquad +n_1\int\frac{\rho\big(\exp(\alpha_0+\beta_0' h(y))\big)^2K_b(t-y)}{\big(1+\rho\exp(\alpha_0+\beta_0' h(y))\big)^2}dG(y)\Big]\\ &= -\int\frac{\rho\exp(\alpha_0+\beta_0' h(y))K_b(t-y)}{1+\rho\exp(\alpha_0+\beta_0' h(y))}dG(y)\\ &= -\rho A(t)\end{aligned} \tag{3.11}$$

where

$$A(t)\equiv\int\frac{\exp\big(\alpha_0+\beta_0' h(y)\big)K_b(t-y)}{1+\rho\exp\big(\alpha_0+\beta_0' h(y)\big)}dG(y)$$

and

$$
\begin{aligned}
E\Big(\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial\beta}\Big) &= -\frac{1}{n_0}E\Big(\sum_{i=1}^{n}\frac{\rho\exp(\alpha_0+\beta_0'h(t_i))h(t_i)K_b(t-t_i)}{\big(1+\rho\exp(\alpha_0+\beta_0'h(t_i))\big)^2}\Big)\\
&= -\frac{1}{n_0}\Big[n_0\int\frac{\rho\exp(\alpha_0+\beta_0'h(y))h(y)K_b(t-y)}{\big(1+\rho\exp(\alpha_0+\beta_0'h(y))\big)^2}dG(y)\\
&\qquad +n_1\int\frac{\rho\big(\exp(\alpha_0+\beta_0'h(y))\big)^2h(y)K_b(t-y)}{\big(1+\rho\exp(\alpha_0+\beta_0'h(y))\big)^2}dG(y)\Big]\\
&= -\int\frac{\rho\exp(\alpha_0+\beta_0'h(y))h(y)K_b(t-y)}{1+\rho\exp(\alpha_0+\beta_0'h(y))}dG(y)\\
&= -\rho B(t) \qquad (3.12)
\end{aligned}
$$

where

$$
B(t)\equiv\int\frac{\exp\big(\alpha_0+\beta_0'h(y)\big)h(y)K_b(t-y)}{1+\rho\exp\big(\alpha_0+\beta_0'h(y)\big)}dG(y).
$$

Let $H_1(t)=H_1(t;\alpha_0,\beta_0)$ and define

$$
H_2(t)=\frac{\rho}{n}\begin{pmatrix} A(t) & B(t)\end{pmatrix}\mathbf{S}^{-1}\begin{pmatrix}\partial l\{\alpha_0,\beta_0\}/\partial\alpha\\ \partial l\{\alpha_0,\beta_0\}/\partial\beta\end{pmatrix}. \qquad (3.13)
$$

From Lemma 1, we have

$$
\begin{pmatrix}\rho A(t) & \rho B(t)\end{pmatrix}\begin{pmatrix}\hat\alpha-\alpha_0\\ \hat\beta-\beta_0\end{pmatrix}=H_2(t)+o_p(n^{-1/2}).
$$

**Lemma 3.** *The function $\hat g(t)$ admits an approximation uniformly in $t$,*

$$
\hat g(t)=H_1(t)-H_2(t)+R_n(t), \qquad (3.14)
$$

*and the remainder term $R_n(t)$ satisfies $sup_t|R_n(t)|=o_p(n^{-1/2})$.*

*Proof.* We study the Taylor expansion of $\hat{g}(t)$ at $(\alpha_0, \beta_0)$. Let $\delta_n = \|(\hat{\alpha}, \hat{\beta}) - (\alpha_0, \beta_0)\|$.

By Lemma 2, we know

$$
\begin{aligned}
\hat{g}(t) &= \frac{1}{n_0} \sum_{i=1}^{n} \frac{K_b(t - t_i)}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))} \\
&= H_1(t; \alpha_0, \beta_0) + \begin{pmatrix} \frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \alpha} \\ \frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \beta} \end{pmatrix}' \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} + o_p(\delta_n) \\
&= H_1(t) + \begin{pmatrix} E\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \alpha} \\ E\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \beta} \end{pmatrix}' \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \\
&\quad + \begin{pmatrix} \frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \alpha} - E\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \alpha} \\ \frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \beta} - E\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \beta} \end{pmatrix}' \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} + o_p(\delta_n).
\end{aligned}
\tag{3.15}
$$

Let

$$
R_{n1}(t) = \begin{pmatrix} \frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \alpha} - E\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \alpha} \\ \frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \beta} - E\frac{\partial H_1(t;\alpha_0,\beta_0)}{\partial \beta} \end{pmatrix}' \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix}
$$

and

$$
R_n(t) = o_p(n^{-1/2}) + R_{n1}(t) + o_p(\delta_n).
$$

Thus

$$
\begin{aligned}
\hat{g}(t) &= H_1(t) - \begin{pmatrix} \rho A(t) & \rho B(t) \end{pmatrix} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} + R_{n1}(t) + o_p(\delta_n) \\
&= H_1(t) - H_2(t) + o_p(n^{-1/2}) + R_{n1}(t) + o_p(\delta_n) \\
&= H_1(t) - H_2(t) + R_n(t).
\end{aligned}
$$

From Lemma 1, the estimator $(\hat{\alpha}, \hat{\beta})$ is consistent in probability. So $\delta_n =$

$o_p(n^{-1/2})$. Again, Lemma 1 implies that $\sup_t |R_{n1}(t)| = o_p(n^{-1/2})$. So we have $\sup_t |R_n(t)| = o_p(n^{-1/2})$ which completes the proof. $\square$

Therefore, $H_1(t) - H_2(t)$ is an approximation of $\hat{g}(t)$ uniformly in $t$. In order to prove the weak convergence of $W_n(t) = \frac{\sqrt{nb}}{2\sqrt{g(t)}}(\hat{g}(t) - \tilde{g}(t))$, according to Lemma 3, we only need to show that $\frac{\sqrt{nb}}{2\sqrt{g(t)}}(H_1(t) - H_2(t) - \tilde{g}(t))$ converges weakly to a Gaussian process.

Next we investigate the structure of the finite dimensional distribution of $\frac{\sqrt{nb}}{2\sqrt{g(t)}}(H_1(t) - H_2(t) - \tilde{g}(t))$.

**Lemma 4.**

$$E\Big(H_1(t) - H_2(t) - \tilde{g}(t)\Big) = 0. \tag{3.16}$$

*Proof.* Obviously, $E(H_2(t)) = 0$ since $E(\partial l(\alpha_0, \beta_0)/\partial(\alpha, \beta)) = 0$

$$\begin{aligned}
E(H_1(t)) &= \frac{1}{n_0} E \sum_{i=1}^{n} \frac{K_b(t - t_i)}{1 + \rho \exp(\alpha_0 + \beta_0' h(t_i))} \\
&= \frac{1}{n_0} \Big[ n_0 \int \frac{K_b(t - y)}{1 + \rho \exp(\alpha_0 + \beta_0' h(y))} dG(y) \\
&\qquad + n_1 \int \frac{K_b(t - y) \exp(\alpha_0 + \beta_0' h(y))}{1 + \rho \exp(\alpha_0 + \beta_0' h(y))} dG(y) \Big] \\
&= \int K_b(t - y) dG(y)
\end{aligned}$$

and

$$E\big(\tilde{g}(t)\big) = \frac{1}{n_0} E \sum_{i=1}^{n_0} K_b(t - u_i) = \int K_b(t - y) dG(y).$$

So

$$E\Big(H_1(t) - H_2(t) - \tilde{g}(t)\Big) = 0.$$

$\square$

## 3.4 Variance-Covariance Structure

Next we consider the variance-covariance of the process $\frac{\sqrt{nb}}{2\sqrt{g(t)}}(H_1(t)-H_2(t)-\tilde{g}(t))$.

$$
\begin{aligned}
&\mathrm{Cov}\Big(\frac{\sqrt{nb}}{2\sqrt{g(s)}}\big(H_1(s)-H_2(s)-\tilde{g}(s)\big),\frac{\sqrt{nb}}{2\sqrt{g(t)}}\big(H_1(t)-H_2(t)-\tilde{g}(t)\big)\Big)\\
&=\frac{nb}{4\sqrt{g(s)g(t)}}\Big\{E\Big(\big(H_1(s)-\tilde{g}(s)\big)\big(H_1(t)-\tilde{g}(t)\big)\Big)\\
&\qquad\qquad -E\Big(\big(H_1(s)-\tilde{g}(s)\big)H_2(t)\Big)-E\Big(H_2(s)\big(H_1(t)-\tilde{g}(t)\big)\Big)\\
&\qquad\qquad +E\Big(H_2(s)H_2(t)\Big)\Big\}. \qquad (3.17)
\end{aligned}
$$

Consider the first part in (3.17). We have,

$$
\begin{aligned}
&n_0^2E\Big[\Big(H_1(s)-\tilde{g}(s)\Big)\Big(H_1(t)-\tilde{g}(t)\Big)\Big]\\
&=n_0^2E\Big[\Big(\frac{1}{n_0}\sum_{i=1}^{n}\frac{K_b(s-t_i)}{1+\rho\exp(\alpha_0+\beta_0'h(t_i))}-\frac{1}{n_0}\sum_{j=1}^{n_0}K_b(s-u_j)\Big)\\
&\qquad\times\Big(\frac{1}{n_0}\sum_{i=1}^{n}\frac{K_b(t-t_i)}{1+\rho\exp(\alpha_0+\beta_0'h(t_i))}-\frac{1}{n_0}\sum_{j=1}^{n_0}K_b(t-u_j)\Big)\Big]\\
&=\Big[E\Big(\sum_{i=1}^{n_1}\frac{K_b(s-z_i)}{1+\rho\exp(\alpha_0+\beta_0'h(z_i))}\Big)\cdot\Big(\sum_{i=1}^{n_1}\frac{K_b(t-z_j)}{1+\rho\exp(\alpha_0+\beta_0'h(z_j))}\Big)\\
&\quad -E\Big(\sum_{i=1}^{n_1}\frac{K_b(s-z_i)}{1+\rho\exp(\alpha_0+\beta_0'h(z_i))}\Big)E\Big(\sum_{j=1}^{n_0}\frac{\rho\exp(\alpha_0+\beta_0'h(u_j))}{1+\rho\exp(\alpha_0+\beta_0'h(u_j))}K_b(t-u_j)\Big)\\
&\quad -E\Big(\sum_{j=1}^{n_0}\frac{\rho\exp(\alpha_0+\beta_0'h(u_j))}{1+\rho\exp(\alpha_0+\beta_0'h(u_j))}K_b(s-u_j)\Big)E\Big(\sum_{i=1}^{n_1}\frac{K_b(t-z_i)}{1+\rho\exp(\alpha_0+\beta_0'h(z_i))}\Big)\\
&\quad +E\Big(\sum_{j=1}^{n_0}\frac{\rho\exp(\alpha_0+\beta_0'h(u_j))}{1+\rho\exp(\alpha_0+\beta_0'h(u_j))}K_b(s-u_j)\Big)\\
&\qquad\times\Big(\sum_{j=1}^{n_0}\frac{\rho\exp(\alpha_0+\beta_0'h(u_j))}{1+\rho\exp(\alpha_0+\beta_0'h(u_j))}K_b(t-u_j)\Big)\Big]\\
&\equiv\Big(I_1-I_2-I_3+I_4\Big). \qquad (3.18)
\end{aligned}
$$

We shall derive every term in (3.18) as follows.

$$\begin{aligned}
I_1 &= n_1 E\Big(\frac{K_b(s-z_i)K_b(t-z_i)}{\big(1+\rho\exp(\alpha_0+\beta_0'h(z_i))\big)^2}\Big) \\
&\quad +(n_1^2-n_1)E\Big(\frac{K_b(s-z_i)}{1+\rho\exp(\alpha_0+\beta_0'h(z_i))}\Big)E\Big(\frac{K_b(t-z_j)}{1+\rho\exp(\alpha_0+\beta_0'h(z_j))}\Big) \\
&= n_0\int\frac{\rho\exp(\alpha_0+\beta_0'h(x))}{\big(1+\rho\exp(\alpha_0+\beta_0'h(x))\big)^2}K_b(s-x)K_b(t-x)dG(x) \\
&\quad +(n_1^2-n_1)A(s)A(t) \\
I_2 &= n_1\int\frac{K_b(s-x)}{1+\rho\exp(\alpha_0+\beta_0'h(x))}\exp(\alpha_0+\beta_0'h(x))dG(x) \\
&\quad \times n_0\int\frac{\rho\exp(\alpha_0+\beta_0'h(x))}{1+\rho\exp(\alpha_0+\beta_0'h(x))}K_b(t-x)dG(x) \\
&= n_0n_1\rho A(s)A(t) \\
I_3 &= n_0\int\frac{\rho\exp(\alpha_0+\beta_0'h(x))}{1+\rho\exp(\alpha_0+\beta_0'h(x))}K_b(s-x)dG(x) \\
&\quad \times n_1\int\frac{K_b(t-x)}{1+\rho\exp(\alpha_0+\beta_0'h(x))}\exp(\alpha_0+\beta_0'h(x))dG(x) \\
&= n_0n_1\rho A(s)A(t) \\
I_4 &= n_0E\Big(\big(\frac{\rho\exp(\alpha_0+\beta_0'h(u_j))}{1+\rho\exp(\alpha_0+\beta_0'h(u_j))}\big)^2K_b(s-u_j)K_b(t-u_j)\Big) \\
&\quad +(n_0^2-n_0)E\Big(\frac{\rho\exp(\alpha_0+\beta_0'h(u_i))}{1+\rho\exp(\alpha_0+\beta_0'h(u_i))}K_b(s-u_i)\Big) \\
&\quad \times E\Big(\frac{\rho\exp(\alpha_0+\beta_0'h(u_j))}{1+\rho\exp(\alpha_0+\beta_0'h(u_j))}K_b(t-u_j)\Big) \\
&= n_0\int\big(\frac{\rho\exp(\alpha_0+\beta_0'h(x))}{1+\rho\exp(\alpha_0+\beta_0'h(x))}\big)^2K_b(s-x)K_b(t-x)dG(x) \\
&\quad +(n_0^2-n_0)\rho^2A(s)A(t).
\end{aligned}$$

Define

$$C(s,t)=\int\frac{\exp(\alpha_0+\beta_0'h(y))}{1+\rho\exp(\alpha_0+\beta_0'h(y))}K_b(s-y)K_b(t-y)dG(y).$$

So, it follows that

$$
\begin{aligned}
& E\Big((H_1(s)-\tilde{g}(s))(H_1(t)-\tilde{g}(t))\Big) \\
= \;& \frac{1}{n_0^2}\Big(I_1 - I_2 - I_3 + I_4\Big) \\
= \;& \frac{1}{n_0^2}\Big(n_0\int \frac{\rho\exp(\alpha_0+\beta_0'h(x))}{\big(1+\rho\exp(\alpha_0+\beta_0'h(x))\big)^2}K_b(s-x)K_b(t-x)dG(x) \\
& \quad +(n_1^2-n_1)A(s)A(t) - 2n_0n_1\rho A(s)A(t) \\
& \quad +n_0\int \big(\frac{\rho\exp(\alpha_0+\beta_0'h(x))}{1+\rho\exp(\alpha_0+\beta_0'h(x))}\big)^2 K_b(s-x)K_b(t-x)dG(x) \\
& \quad +(n_0^2-n_0)\rho^2A(s)A(t)\Big) \\
= \;& \frac{1}{n_0}\Big(\rho C(s,t) - \rho(1+\rho)A(s)A(t)\Big).
\end{aligned}
$$

Note that if all the variables are restricted to $[-L, L]$, then $bC(s,t)$ is finite. Let $y = s - bx$. Then

$$
bC(s,t) = \int_{(s-L)/b}^{(s+L)/b} \frac{\exp(\alpha_0+\beta_0'h(s-bx))}{1+\rho\exp(\alpha_0+\beta_0'h(s-bx))}K(x)K(\frac{s-t}{b}-x)g(s-bx)dx
$$

is bounded for any $b > 0$. Specifically, when $s = t$

$$
bC(s,s) = \int_{\frac{s-L}{b}}^{\frac{s+L}{b}} \frac{\exp(\alpha_0+\beta_0'h(s-bx))}{1+\rho\exp(\alpha_0+\beta_0'h(s-bx))}K(x)^2g(s-bx)dx.
$$

If the kernel function $K(\cdot)$ has compact support (e.g. Epanechnikov, Uniform or Biweight) and if $b << |s-t|$, such that $|s-t|/b$ is greater than the range of the compact support, then $K(x)\cdot K(\frac{s-t}{b}-x)$ is zero for any $x$ and thus the value of $C(s,t)$ vanishes. To avoid this situation, in this chapter we only use a kernel whose support is all of $\mathcal{R}$ (e.g. Gaussian).

Next consider the fourth term in (3.17). From Qin & Zhang (1997) and Lu

(2007)[25], we have

$$\mathbf{S}^{-1}V\mathbf{S}^{-1} = \frac{1+\rho}{\rho}A^{-1} - \begin{pmatrix} \frac{(1+\rho)^2}{\rho} & 0 \\ 0 & 0 \end{pmatrix},$$

thus for the fourth term in (3.17),

$$\begin{aligned}
& E\Big((H_2(s)H_2'(t)\Big) \\
= \;& \frac{\rho^2}{n^2}E\Bigg\{\begin{pmatrix} A(s) & B(s) \end{pmatrix}\mathbf{S}^{-1}\begin{pmatrix} \partial l\{\alpha_0,\beta_0\}/\partial\alpha \\ \partial l\{\alpha_0,\beta_0\}/\partial\beta \end{pmatrix} \\
& \quad \times\left(\begin{pmatrix} A(t) & B(t) \end{pmatrix}\mathbf{S}^{-1}\begin{pmatrix} \partial l\{\alpha_0,\beta_0\}/\partial\alpha \\ \partial l\{\alpha_0,\beta_0\}/\partial\beta \end{pmatrix}\right)'\Bigg\} \\
= \;& \frac{\rho^2}{n}\begin{pmatrix} A(s) & B(s) \end{pmatrix}\mathbf{S}^{-1}V\mathbf{S}^{-1}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} \\
= \;& \frac{\rho^2}{n}\begin{pmatrix} A(s) & B(s) \end{pmatrix}\frac{1+\rho}{\rho}A^{-1}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} \\
& -\frac{\rho^2}{n}\begin{pmatrix} A(s) & B(s) \end{pmatrix}\begin{pmatrix} \frac{(1+\rho)^2}{\rho} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} \\
= \;& \frac{\rho(1+\rho)}{n}\begin{pmatrix} A(s) & B(s) \end{pmatrix}A^{-1}\begin{pmatrix} A(t) \\ B(t) \end{pmatrix} - \frac{\rho(1+\rho)^2}{n}A(s)A(t).
\end{aligned}$$

Furthermore, from Qin & Zhang (1997) and Lu (2007)[25], we have

$$\mathrm{Cov}(\sqrt{nb}(H_1(s)-\tilde{g}(s)), \sqrt{nb}H_2(t)) = \mathrm{Cov}(\sqrt{nb}H_2(s), \sqrt{nb}H_2(t))$$

and

$$\mathrm{Cov}(\sqrt{nb}(H_2(s)-\tilde{g}(s)), \sqrt{nb}H_1(t)) = \mathrm{Cov}(\sqrt{nb}H_2(s), \sqrt{nb}H_2(t)).$$

So (3.17) becomes,

$$
\begin{aligned}
& \mathrm{Cov}\Big(\frac{\sqrt{nb}}{2\sqrt{g(s)}}\big(H_1(s)-H_2(s)-\tilde{g}(s)\big),\frac{\sqrt{nb}}{2\sqrt{g(t)}}\big(H_1(t)-H_2(t)-\tilde{g}(t)\big)\Big) \\
= \;& \frac{nb}{4\sqrt{g(s)g(t)}}\Big\{E\Big(\big(H_1(s)-\tilde{g}(s)\big)\big(H_1(t)-\tilde{g}(t)\big)\Big)-E\Big(\big(H_1(s)-\tilde{g}(s)\big)H_2(t)\Big) \\
& \qquad -E\Big(H_2(s)\big(H_1(t)-\tilde{g}(t)\big)\Big)+E\Big(H_2(s)H_2(t)\Big)\Big\} \\
= \;& \frac{nb}{4\sqrt{g(s)g(t)}}\Big\{\frac{1}{n_0}\big(\rho C(s,t)-\rho(1+\rho)A(s)A(t)\big)+\frac{\rho(1+\rho)^2}{n}A(s)A(t) \\
& \qquad -\frac{\rho(1+\rho)}{n}\begin{pmatrix} A(s) & B(s)\end{pmatrix}A^{-1}\begin{pmatrix} A(t) \\ B(t)\end{pmatrix}\Big\} \\
= \;& \frac{\rho(1+\rho)b}{4\sqrt{g(s)g(t)}}\left\{C(s,t)-\begin{pmatrix} A(s) & B(s)\end{pmatrix}A^{-1}\begin{pmatrix} A(t) \\ B(t)\end{pmatrix}\right\}.
\end{aligned}
$$

## 3.5 Weak Convergence of $W_n(t)$

Define

$$
V(s,t)=\frac{\rho(1+\rho)b}{4\sqrt{g(s)g(t)}}\left\{C(s,t)-\begin{pmatrix} A(s) & B(s)\end{pmatrix}A^{-1}\begin{pmatrix} A(t) \\ B(t)\end{pmatrix}\right\}
$$

**Theorem 1.** *Assume that the underling density $g(\cdot)$, the tilt function $h(\cdot)$ and the kernel $K(\cdot)$ all are Lipschitz continuous. The process $W_n(t)=\frac{\sqrt{nb}}{2\sqrt{g(t)}}\big(\hat{g}(t)-\tilde{g}(t)\big)$ converges weakly in [-L,L] to a Gaussian process W(t) with mean 0 and*

$$
\begin{aligned}
& Cov(W(s),W(t))=V(s,t) \\
& =\frac{\rho(1+\rho)b}{4\sqrt{g(s)g(t)}}\left\{C(s,t)-\begin{pmatrix} A(s) & B(s)\end{pmatrix}A^{-1}\begin{pmatrix} A(t) \\ B(t)\end{pmatrix}\right\}. \qquad (3.19)
\end{aligned}
$$

Note: Bondell (2007)[4] states a similar result for $\sqrt{n}(\hat{g}(t) - \tilde{g}(t))$ with fixed bandwidth $b = 1$ without proof.

*Proof.* Since $C(s,t)$, $A(s)$ and $A(t)$ are continuous, $V(s,t)$ is continuous and bandwidth is fixed. So $V(s,t)$ is bounded for both $s$ and $t$ in $[-L, L]$. Thus finite-dimensional convergence follows from Lindeberg-Feller. For tightness, following Corollary 16.9 in Kallenberg 2002[19], we need to check the Kolmogorov-Chentsov criterion which is $E(W_n(s) - W_n(t))^2 \le c|s-t|^2$ for some constant $c < \infty$.

$$
\begin{aligned}
&E(W_n(s) - W_n(t))^2 \\
&= E(W_n(s))^2 + E(W_n(t))^2 - 2E(W_n(s)W_n(t)) \\
&= V(s,s) + V(t,t) - 2V(s,t) \\
&= \frac{\rho(1+\rho)b}{4}\Big\{\Big(\frac{C(s,s)}{g(s)} + \frac{C(t,t)}{g(t)} - 2\frac{C(s,t)}{\sqrt{g(s)g(t)}}\Big) - \Big(a_1\big(\frac{A(s)}{\sqrt{g(s)}} - \frac{A(t)}{\sqrt{g(t)}}\big)^2 \\
&\qquad + a_3\big(\frac{B(s)}{\sqrt{g(s)}} - \frac{B(t)}{\sqrt{g(t)}}\big)^2 + 2a_2\big(\frac{A(s)}{\sqrt{g(s)}} - \frac{A(t)}{\sqrt{g(t)}}\big)\big(\frac{B(s)}{\sqrt{g(s)}} - \frac{B(t)}{\sqrt{g(t)}}\big)\Big)\Big\} \\
&= \frac{\rho(1+\rho)b}{4}\Big\{\Big(\frac{C(s,s)}{g(s)} + \frac{C(t,t)}{g(t)} - 2\frac{C(s,t)}{\sqrt{g(s)g(t)}}\Big) - \Big(\sqrt{a_1}\big(\frac{A(s)}{\sqrt{g(s)}} - \frac{A(t)}{\sqrt{g(t)}}\big) \\
&\qquad + \frac{a_2}{\sqrt{a_1}}\big(\frac{B(s)}{\sqrt{g(s)}} - \frac{B(t)}{\sqrt{g(t)}}\big)\Big)^2 + \big(\frac{a_2^2}{a_1} - a_3\big)\Big(\frac{B(s)}{\sqrt{g(s)}} - \frac{B(t)}{\sqrt{g(t)}}\Big)^2\Big\}
\end{aligned}
$$

where

$$
A^{-1} = \begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix}.
$$

Let

$$
\begin{aligned}
a(y) &= \frac{\exp(\alpha_0 + \beta_0' h(y))g(y)}{1 + \rho\exp(\alpha_0 + \beta_0' h(y))} \\
c(y) &= \frac{\exp(\alpha_0 + \beta_0' h(y))h(y)g(y)}{1 + \rho\exp(\alpha_0 + \beta_0' h(y))}.
\end{aligned}
$$

We assume that $g(\cdot)$ is bounded away from zero in $[-L, L]$. Since $g(\cdot)$, $h(\cdot)$, and $K(\cdot)$ all are Lipschitz continuous, $\exists\ C_1 > 0$ s.t.

$$\begin{aligned}
\Big|\frac{A(s)}{\sqrt{g(s)}} - \frac{A(t)}{\sqrt{g(t)}}\Big| &= \Big|\int \frac{a(y)}{\sqrt{g(s)}} K_b(s-y)dy - \int \frac{a(y)}{\sqrt{g(t)}} K_b(t-y)dy\Big| \\
&\le \int \Big|\frac{a(s-bx)}{\sqrt{g(s)}} - \frac{a(t-bx)}{\sqrt{g(t)}}\Big| K(x)dx \\
&\le \int \Big|a(s-bx) - a(t-bx)\Big| \frac{1}{\sqrt{g(s)}} K(x)dx \\
&\quad + \int \Big|\frac{1}{\sqrt{g(s)}} - \frac{1}{\sqrt{g(t)}}\Big| \cdot \Big|a(t-bx)\Big| K(x)dx \\
&\le C_1 \Big|s-t\Big|
\end{aligned}$$

and similarly $\exists\ C_2 > 0$ s.t.

$$\begin{aligned}
\Big|\frac{B(s)}{\sqrt{g(s)}} - \frac{B(t)}{\sqrt{g(t)}}\Big| &= \Big|\int \frac{c(y)}{\sqrt{g(s)}} K_b(s-y)dy - \int \frac{c(y)}{\sqrt{g(t)}} K_b(t-y)dy\Big| \\
&\le \int \Big|\frac{c(s-bx)}{\sqrt{g(s)}} - \frac{c(t-bx)}{\sqrt{g(t)}}\Big| K(x)dx \\
&\le \int \Big|c(s-bx) - c(t-bx)\Big| \frac{1}{\sqrt{g(s)}} K(x)dx \\
&\quad + \int \Big|\frac{1}{\sqrt{g(s)}} - \frac{1}{\sqrt{g(t)}}\Big| \cdot \Big|c(t-bx)\Big| K(x)dx \\
&\le C_2 \Big|s-t\Big|.
\end{aligned}$$

Furthermore, $\exists\ C_3 > 0$ s.t.

$$\begin{aligned}
& b \cdot \Big|\frac{C(s,s)}{g(s)} + \frac{C(t,t)}{g(t)} - 2\frac{C(s,t)}{\sqrt{g(s)g(t)}}\Big| = \Big|b \cdot \int a(y)\Big(\frac{K_b(s-y)}{\sqrt{g(s)}} - \frac{K_b(t-y)}{\sqrt{g(t)}}\Big)^2 dy\Big| \\
&= \Big|\int a(s-bx)\Big(\frac{K(x)}{\sqrt{g(s)}} - \frac{K(\frac{t-s}{b}+x)}{\sqrt{g(t)}}\Big)^2 dx\Big| \\
&\le \frac{1}{2}\Big|\int a(s-bx)K^2(x)\Big(\frac{1}{\sqrt{g(s)}} - \frac{1}{\sqrt{g(t)}}\Big)^2 dx\Big| \\
&\quad + \frac{1}{2}\Big|\int \frac{a(s-bx)}{g(t)}\Big(K(x) - K(\frac{t-s}{b}+x)\Big)^2 dx\Big| \\
&\le C_3 \Big|s-t\Big|^2.
\end{aligned}$$

Thus $\exists C^* > 0$ s.t.

$$E\Big|W_n(s) - W_n(t)\Big|^2 \leq C^*\Big|s - t\Big|^2$$

And it is easy to see that the sequence $W_n(0)$ is tight by Chebyshev's inequality. Thus following the Corollary 16.9 in Kallenberg 2002[19], the tightness of $W_n(t)$ is proved. The weak convergence of $W_n(t)$ follows. $\square$

*Remark* 1. Since $\sqrt{g(t)}$ is continuous, $\sqrt{g(t)}$ is bounded in $[-L, L]$. If the bandwidth is fixed as $b = 1$, then $\sqrt{n}\{\hat{g}(t) - \tilde{g}(t)\} = 2\sqrt{g(t)}W_n(t)$ converges weakly in $[-L, L]$ to a Gaussian process $2\sqrt{g(t)}W(t)$ with mean 0 and covariance function given by

$$4\sqrt{g(s)g(t)}V(s,t) = \rho(1+\rho)\left\{C(s,t) - \begin{pmatrix} A(s) & B(s) \end{pmatrix} A^{-1} \begin{pmatrix} A(t) \\ B(t) \end{pmatrix}\right\}$$

This result was stated by Bondell(2007)[4] for $I_n^B$ for without proof.

## 3.6 Convergence in Distribution to Linear Combination of Chi-square Variables

We have shown that $W_n(t)$ converges weakly to a Gaussian process $W(t)$ with mean 0 and covariance (3.19). The Gaussian process $W(t)$ can be represented in terms of its eigenfunction expansion by the Karhunen-Loève theorem. Before stating the theorem, we introduce the Mercer kernel.

A function $V(s,t) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is called a Mercer kernel if it satisfies,

1. $V(s,t)$ is continuous

2. Symmetric: $V(s,t) = V(t,s)$

3. $V(s,t)$ is positive definite; that is, for all finite sequences of points $x_1, x_2, ..., x_n$ and all choices of real numbers $c_1, c_2, ..., c_n$, $\sum_{i=1}^{n} \sum_{j=1}^{n} V(t_i, x_j) c_i c_j \geq 0$.

**Theorem 2** (Karhunen-Loève theorem)**.** *Let $W(t)$ be a zero-mean square integrable stochastic process over a closed and bounded interval $[-L, L]$, with continuous covariance function $V(s,t)$. Then $V(s,t)$ is a Mercer kernel. Let $\{e_K(t)\}_{k=1}^{\infty}$ be an orthonormal basis of $L^2([-L, L])$ formed by the eigenfunctions of*

$$\int_{-L}^{L} V(s,t) e_k(s) ds = \lambda_k e_k(t) \tag{3.20}$$

*with respective eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$. Then W(t) admits the following representation*

$$W(t) = \sum_{k=1}^{\infty} Z_k e_k(t)$$

*where the convergence is in $L^2$, uniform in $t$ and*

$$Z_k = \int_{-L}^{L} W(t) e_k(t) dt.$$

*Furthermore, if the original process $W(t)$ is Gaussian, the random variables $Z_k$ are independent normal random variables with zero-mean and variance $\lambda_k$.*

Let

$$\zeta_k = \frac{1}{\lambda_k} Z_k^2.$$

Since $W(t)$ is Gaussian, we have $\zeta_k$ follows a chi-square distribution with one degree of freedom.

Hence, by Theorem 1 and Theorem 2,

$$
\begin{aligned}
\int_{-L}^{L} W_n(t)^2 dt &\to \int_{-L}^{L} W^2(t)dt && (\text{ Theorem 1 }) \\
&= \int_{-L}^{L} (\sum_{k=1}^{\infty} Z_k e_k(t))^2 dt && (\text{ Theorem 2 }) \\
&= \sum_{k=1}^{\infty} Z_k^2 \\
&= \sum_{k=1}^{\infty} \lambda_k \zeta_k.
\end{aligned}
$$

Thus we have derived the following theorem,

**Theorem 3.** *Assume that the underling density $g(\cdot)$, the tilt function $h(\cdot)$ and the kernel $K(\cdot)$ all are Lipschitz continuous, and $L$ is sufficiently large. The sequence $\{\lambda_k\}_{k=1}^{\infty}$ is defined as (3.20) in Theorem 2. The random variable $\int_{-L}^{L} W_n(t)^2 dt$ tends in distribution to $\sum_{k=1}^{\infty} \lambda_k \zeta_k$, where $\{\zeta_k\}$ follow chi-square distribution with one degree of freedom.*

Next we will derive a numerical approach for $I_n$. From the standard consistency result in Devroye and Györfi(1985)[7], we have that $\int |g(t) - \hat{g}(t)|dt$ and $\int |g(t) - \tilde{g}(t)|dt$ both converge to 0 with probability one if $b \to 0$ and $nb \to \infty$. And by Theorem 1, $E\big(nb(\hat{g}(t) - \tilde{g}(t))^2\big) = 4g(t)EW_n(t)^2$ is finite. Thus when the sample size $n$ is large enough, $\forall \epsilon > 0$, $\exists b > 0$ s.t. $\int_{-L}^{L} D_n(t)dt < \epsilon$ almost surely. And then for fixed $b$, we can derive Theorem 1 and Theorem 3. By (3.8), $I_n = \int_{-L}^{L} W_n(t)^2 dt + \int_{-L}^{L} D_n(t)dt$. Thus $I_n$ can be approximated numerically by $\int_{-L}^{L} W_n(t)^2 dt$ and then equals to $\sum_{k=1}^{\infty} \lambda_k \zeta_k$.

**Corollary 1.** *When sample size $n$ is large enough, we can choose and fix the bandwidth $b$ which makes $\int_{-L}^{L} D_n(t)dt$ small. Thus $I_n = nb\int_{-L}^{L}\left(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\right)^2 dt$ can be approximated by $\sum_{k=1}^{\infty}\lambda_k\zeta_k$ where the sequence $\{\lambda_k\}_{k=1}^{\infty}$ is defined as (3.20) and $\{\zeta_k\}$ follow chi-square distribution with one degree of freedom.*

However it is difficult to derive the analytic expression for the distribution of $\sum_{k=1}^{\infty}\lambda_k\zeta_k$. Therefore we apply a bootstrap procedure as in Qin and Zhang (1997)[27] in the next section.

Bondell (2007)[4] considered the test statistic $I_n^B$ as (1.5) and he assumed that the bandwidth is fixed at $b = 1$. For the sake of comparison, we modify his test statistic as

$$I_n^{Bb} = nb\int_{-L}^{L}\left(\hat{g}(t) - \tilde{g}(t)\right)^2 dt \tag{3.21}$$

where the bandwidth $b$ needs to be selected properly.

## 3.7 Numerical Study

In this section we compare the performance of our $I_n$ with the Kommogorov-Simirnov-type statistic $\Delta_n$ in (1.2) of Qin and Zhang (1997)[27], the test statistic $I_n^B$ in (1.5) of Bondell (2007)[4], and the modified test statistic $I_n^{Bb}$ as in (3.21). Thus, these four test statistics $\Delta_n$, $I_n^B$, $I_n^{Bb}$, and $I_n$ are used in goodness-of-fit testing to validate model (1.1).

In the following simulations, we use the Gaussian kernel in our density estimators, $K(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. The value of bandwidth $b$ is needed in the calculations of $I_n^{Bb}$ and $I_n$. From standard consistency results, $b \to 0$ and $nb \to \infty$

are needed. For a given sample, where the sample size $n$ is large enough, we fix the bandwidth at a proper value which numerically makes $\hat{g}(\cdot)$ and $\tilde{g}(\cdot)$ close to $g(\cdot)$ and simultaneously makes $\int_{-L}^{L} D_n(t)dt$ small. Thus we fix the bandwidth so that $4/n < b < 0.1$. Experience shows that when $n = 1000$, $b = \sqrt{\frac{4}{n} \cdot 0.1} = 0.02$ is a good choice.

### 3.7.1 Bootstrap Procedure

The bootstrap procedure which was suggested by Qin and Zhang (1997)[27] simulates the distribution of the test statistic and its quantiles. The steps are as follows:

1. Generate samples $U$ and $Z$ following model (1.1).

2. Obtain semiparametric estimates $(\hat{\alpha}, \hat{\beta})$ and $\hat{G}(x)$ from the fused sample $T$, and empirical $\tilde{G}(x)$ from the reference sample $U$ only.

3. Generate bootstrap data $U^*$ and $Z^*$ from $d\hat{G}(x)$ and $\exp(\hat{\alpha} + x\hat{\beta})d\hat{G}(x)$ respectively.

4. Obtain the estimated $(\hat{\alpha}^*, \hat{\beta}^*)$ and $\hat{G}^*(x)$ from the fused sample $T^* = (U^*, Z^*)$, and the empirical $\tilde{G}^*(x)$ from the bootstrap reference sample $U^*$.

5. Derive the bootstrap version of the test statistics $\Delta_n^* = \sup_t \sqrt{n}|\hat{G}^*(t) - \tilde{G}^*(t)|$, $I_n^{B*} = n\int \left(\hat{g}^*(t) - \tilde{g}^*(t)\right)^2 dt$, $I_n^{Bb*} = nb\int \left(\hat{g}^*(t) - \tilde{g}^*(t)\right)^2 dt$ and $I_n^* = nb\int \left(\sqrt{\hat{g}^*(t)} - \sqrt{\tilde{g}^*(t)}\right)^2 dt$. Here $\hat{g}^*(t)$ and $\tilde{g}^*(t)$ are derived from $\hat{G}^*(x)$ and $\tilde{G}^*(x)$ respectively, following Lemma 2.

6. Repeat step 3 to step 5 to generate many bootstrap replications of the test

statistics to approximate the critical values.

7. Calculate the empirical $p$-values.

We apply our proposed goodness-of-fit test statistic to simulated data following model (1.1). We generate $U \sim N(0,1)$ and $Z \sim N(0,2)$ with sample sizes $(n_0, n_1) = (500, 500)$. The fused sample is $T = \{U, Z\}$ with size $n = 1000$. The bandwidth is fixed at $b = 0.02$. According to Example 1, the correctly specified tilt function is $h(x) = x^2$. Then from the fused sample $T$ we can obtain $(\hat{\alpha}, \hat{\beta})$ and $\hat{G}(\cdot)$, the estimated semiparametric distribution. And from Lemma (2) we obtain the kernel density estimators $\hat{g}(\cdot)$ and $\tilde{g}(\cdot)$. From this the observed value of test statistic $I_n = 0.4685$. To simulate the distribution of $I_n$, we generate $U^*$ and $Z^*$ from $d\hat{G}(x)$ and $\exp(\hat{\alpha}+x\hat{\beta})d\hat{G}(x)$ respectively. From 500 bootstrap replications of $I_n^*$ from $\{U^*, Z^*\}$, the observed $p$-value is $P(I_n^* > I_n) = 0.856$ which means that null hypothesis model (1.1) should be accepted. By applying the same procedure we obtained $p$-values corresponding to $\Delta_n$, $I_n^B$ and $I_n^{Bb}$: $P(\Delta_n^* > \Delta_n) = 0.834$, $P(I_n^{B*} > I_n^B) = 0.966$ and $P(I_n^{Bb*} > I_n^{Bb}) = 0.508$. They all validate model (1.1).

Still using the same samples $U \sim N(0,1)$ and $Z \sim N(0,2)$ of size $(n_0, n_1) = (500, 500)$, replacing the correctly specified tilt function $h(x) = x^2$ by the misspecified tilt function $h(x) = x$, from the same procedure we get the observed value of $I_n = 3.022$ and corresponding $p$-value $P(I_n^* > I_n) = 0$, which indicates a strong rejection of the null hypothesis model (1.1). The $p$-values corresponding to the other three tests are all zero and so model (1.1) is rejected by all the tests.

### 3.7.2 Monte-Carlo Simulation

Next we run Monte-Carlo simulations to investigate the distributions of the $p$-values corresponding to four test statistics $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $I_n$. Recall from (1.2), (1.5), (3.21) and (3.5),

$$\Delta_n = \sup_t \sqrt{n}|\hat{G}(t) - \tilde{G}(t)| \qquad I_n^B = n \int_{-L}^{L} \left(\hat{g}(t) - \tilde{g}(t)\right)^2 dt$$

$$I_n^{Bb} = nb \int_{-L}^{L} \left(\hat{g}(t) - \tilde{g}(t)\right)^2 dt \qquad I_n = nb \int_{-L}^{L} \left(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\right)^2 dt$$

Note that bandwidth $b = 1$ in $I_n^B$ and $b = 0.02$ in both $I_n^{Bb}$ and $I_n$. We ran 100 Monte-Carlo simulation repetitions. In each repetition, we generated $U \sim N(0,1)$ and $Z \sim N(0,2)$ with sample sizes $(n_0, n_1) = (500, 500)$. After deriving $(\hat{\alpha}, \hat{\beta})$ and $\hat{G}(x)$ from fused sample $T = \{U, Z\}$, we repeat the bootstrap procedure 500 times from step 3 to step 6. Therefore we get the distributions of the observed $p$-values corresponding to each test statistic from 100 Monte-Carlo repetitions.

Figure 3.1 gives the distributions of $p$-values corresponding to four test statistics when the tilt function is correctly specified as $h(x) = x^2$. Figure 3.2 gives the distributions of $p$-values corresponding to four test statistics when the tilt function is misspecified as $h(x) = x$. Table 3.1 gives minimum, lower quartile, median, upper quartile and maximum values of the distributions of $p$-values corresponding to $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $I_n$. The 2nd column from right is the difference between the upper quartile value and the lower quartile value, and the last column is the variance of the distributions of the $p$-values for each statistic.

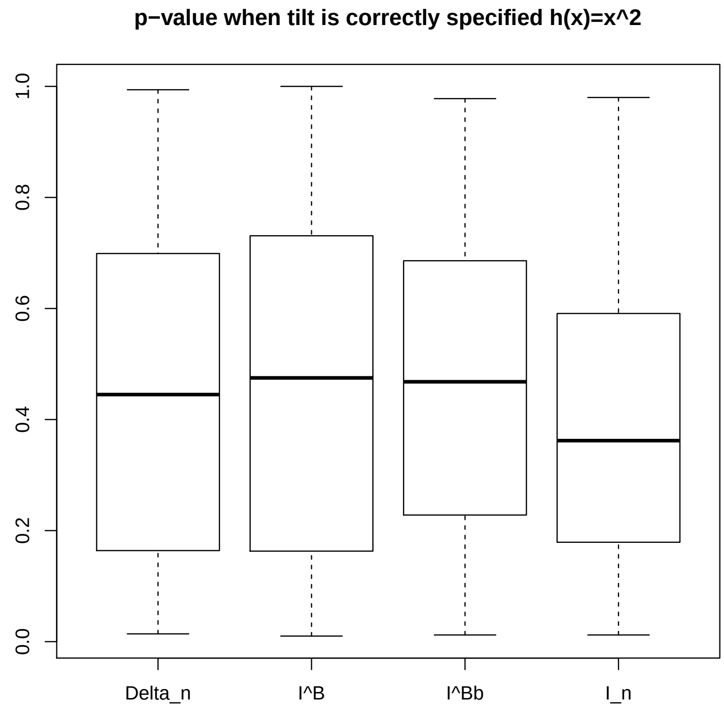


Figure 3.1: *Distributions of* $p$*-values corresponding to* $\Delta_n$, $I_n^B$, $I_n^{Bb}$ *and* $I_n$. $U \sim N(0,1)$ *and* $Z \sim N(0,2)$ *with sample sizes* $(n_0, n_1) = (500, 500)$. *Correctly specified tilt function* $h(x) = x^2$. *Bandwidth* $b = 0.02$ *for* $I_n^{Bb}$ *and* $I_n$. *Bootstrap repetitions=500, Monte-Carlo repetitions=100.*

Figure 3.2 tells us that when the tilt function is misspecified, all four test statistics lead to a strong rejection of model (1.1). Figure 3.1 tells us that when the tilt function is correctly specified, all four test statistics suggest acceptance of model (1.1) in most of the cases. Figure 3.1 also tells that the distributions of $p$-values corresponding to $I_n$ has smaller variation than those of other statistics. From Table 3.1 we see that the distribution of $p$-values corresponding to $I_n$ have the smallest variation among all four test statistics.

In the simulation above we generate samples from the normal distribution. Next we generate lognormal samples. The lognormal distribution has much heavier

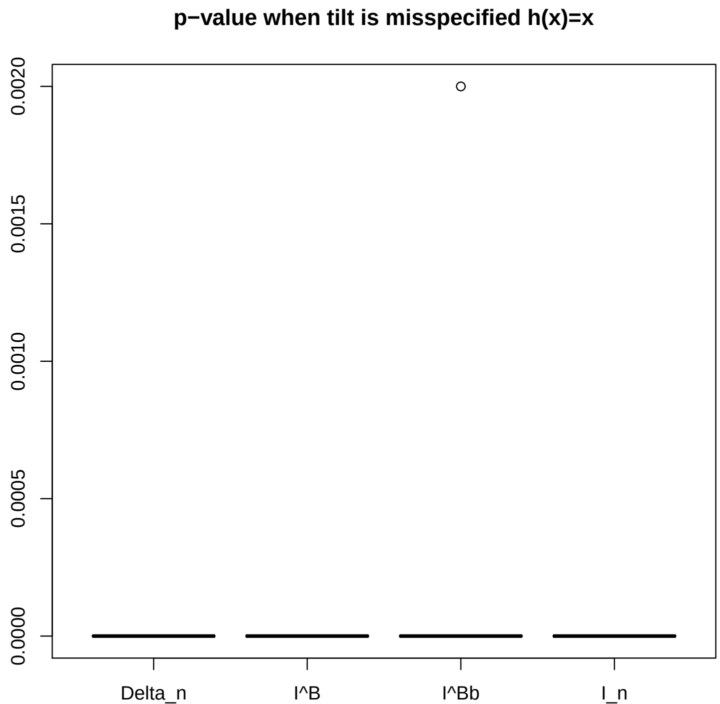


Figure 3.2: *Distributions of p-values corresponding to* $\Delta_n$, $I_n^B$, $I_n^{Bb}$ *and* $I_n$. $U \sim N(0,1)$ *and* $Z \sim N(0,2)$ *with sample sizes* $(n_0, n_1) = (500, 500)$. *The tilt function is misspecified* $h(x) = x$. *Bandwidth* $b = 0.02$ *for* $I_n^{Bb}$ *and* $I_n$. *Bootstrap repetitions=500, Monte-Carlo repetitions=100.*

tails than the normal distribution. We ran 100 Monte-Carlo simulation repetitions. In each repetition, we generated $U \sim Lognormal(0, 0.5)$ and $Z \sim Lognormal(0, 0.7)$ with sample sizes $(n_0, n_1) = (500, 500)$. According to Example 3, the correctly specified tilt function is $h(x) = (\log(x))^2$. After deriving $(\hat{\alpha}, \hat{\beta})$ and $\hat{G}(x)$ from the fused sample $T = \{U, Z\}$, we repeated the bootstrap procedure 500 times from step 3 to step 6 above. From this we obtained observed the distributions of $p$-values corresponding to each test statistic by 100 Monte-Carlo repetitions.

Figure 3.3 gives the distributions of $p$-values corresponding to four test statistics when tilt function is correctly specified as $h(x) = (\log(x))^2$. Figure 3.4 gives the distributions of $p$-values corresponding to four test statistics when tilt function is misspecified as $h(x) = \log(x)$. Table 3.2 gives the minimum, lower quartile, median, upper quartile and maximum values of the distributions of $p$-values corresponding to $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $I_n$. The 2nd column from right is the difference between upper quartile and lower quartile, and the last column is the variance of the distributions of $p$-values for each statistic.

The simulation results from the lognormal distribution are very similar to those from the normal distribution. For all four test statistics, Figure 3.4 indicates strong rejections of model (1.1) when the tilt function is misspecified, however Figure 3.3 suggests acceptances of model (1.1) when the tilt function is correctly specified. From 3.3 and Table 3.2, it is observed that the distribution of the $p$-values

Table 3.1: Distributions of $p$-values corresponding to $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $I_n$. $U \sim$ Normal (0,1), $Z \sim$ Normal (0,2) with sample size $(n_0, n_1) = (500, 500)$. Correctly specified tilt function $h(x) = x^2$. Bandwidth $b = 0.02$ for $I_n^{Bb}$ and $I_n$. Bootstrap repetitions=500, Monte-Carlo repetitions=100.

| statistic | minimum | lower quartile | median | upper quartile | maximum | upper -lower | variance |
|---|---|---|---|---|---|---|---|
| $\Delta_n$ | 0.014 | 0.164 | 0.445 | 0.699 | 0.994 | 0.535 | 0.09331524 |
| $I_n^B$ | 0.010 | 0.163 | 0.475 | 0.731 | 1.000 | 0.568 | 0.09487272 |
| $I_n^{Bb}$ | 0.012 | 0.228 | 0.468 | 0.686 | 0.978 | 0.458 | 0.08061943 |
| $I_n$ | 0.012 | 0.179 | 0.362 | 0.591 | 0.980 | 0.412 | 0.07121728 |

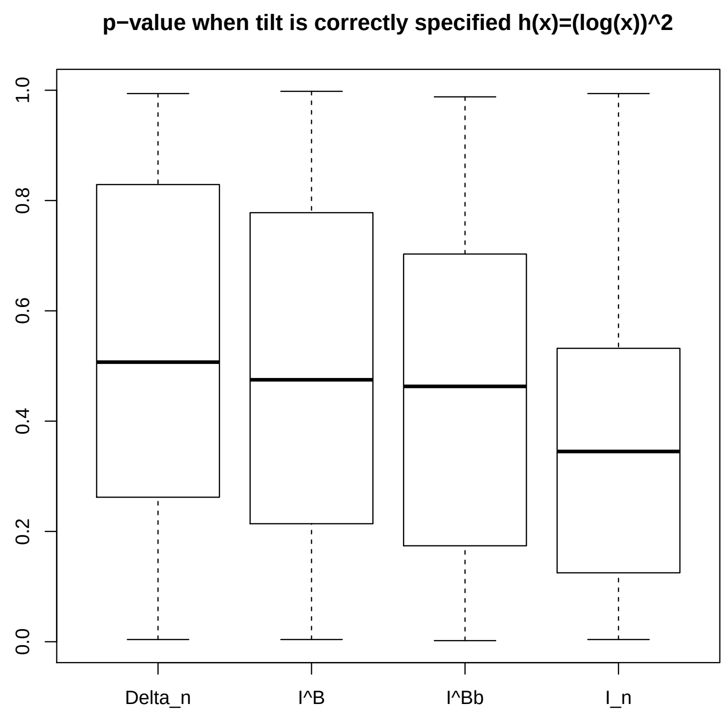


Figure 3.3: *Distributions of p-values corresponding to* $\Delta_n$*,* $I_n^B$*,* $I_n^{Bb}$ *and* $I_n$*.* $U \sim Lognormal(0, 0.5)$ *and* $Z \sim Lognormal(0, 0.7)$ *with sample size* $(n_0, n_1) = (500, 500)$*. Correctly specified tilt function* $h(x) = (\log(x))^2$*. Bandwidth* $b = 0.02$ *for* $I_n^{Bb}$ *and* $I_n$*. Bootstrap repetitions=500, Monte-Carlo repetitions=100.*

corresponding to $I_n$ has the smallest variation among all four test statistics.

Since the distribution of $\Delta_n$ is completely unknown, the goodness-of-fit test based on $\Delta_n$ requires simulations to obtain $p$-values. Since $I_n^B$ uses a fixed bandwidth $b = 1$, the density estimators $\hat{g}(\cdot)$ and $\tilde{g}(\cdot)$ are not consistent. Thus, the test statistic $I_n^B$ may not accurately measure the difference between two competing densities. Hence $I_n^B$ may not provide correct decisions. The limiting distribution of $I_n$, derived in this chapter, and $I_n^{Bb}$ have similar limiting distribution. Thus $I_n$ and $I_n^{Bb}$ can provide more reliable goodness-of-fit test decisions. In the simulations above, the distribution of $p$-values of $I_n$ has the smallest variation as compared with the other

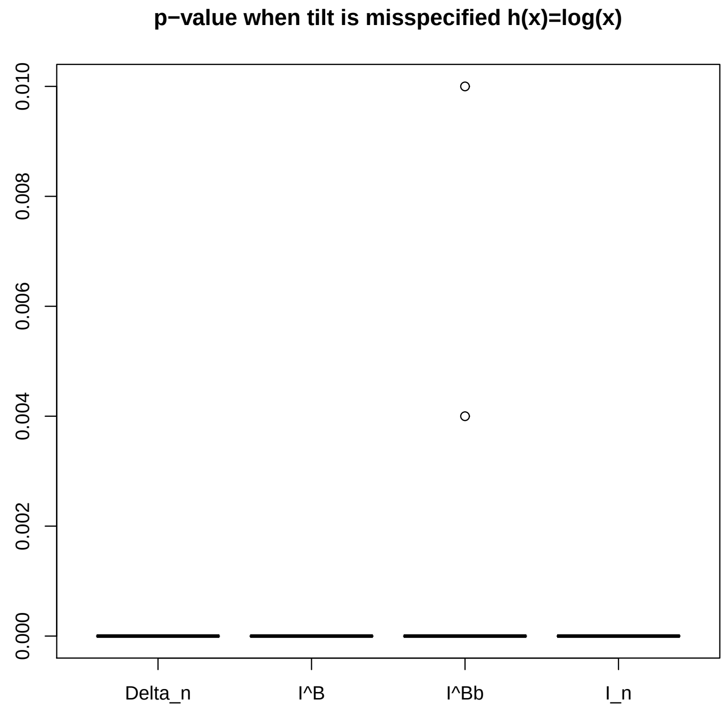


Figure 3.4: *Distributions of p-values corresponding to* $\Delta_n$, $I_n^B$, $I_n^{Bb}$ *and* $I_n$. $U \sim Lognormal(0, 0.5)$ *and* $Z \sim Lognormal(0, 0.7)$ *with sample size* $(n_0, n_1) = (500, 500)$. *The tilt function is misspecified* $h(x) = \log(x)$. *Bandwidth* $b = 0.02$ *for* $I_n^{Bb}$ *and* $I_n$. *Bootstrap repetitions=500, Monte-Carlo repetitions=100.*

three test statistics, but all four test statistics reach the same goodness-of-fit test decisions.

Table 3.2: Distributions of $p$-values corresponding to $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $I_n$. $U \sim$ Lognormal (0,0.5), $Z \sim$ Lognormal (0,0.7) with sample size $(n_0, n_1) = (500, 500)$. Correctly specified tilt function $h(x) = (\log(x))^2$. Bandwidth $b = 0.02$ for $I_n^{Bb}$ and $I_n$. Bootstrap repetitions=500, Monte-Carlo repetitions=100.

| test statistic | mini- mum | lower quartile | median | upper quartile | maxi- mum | upper -lower | variance |
|---|---|---|---|---|---|---|---|
| $\Delta_n$ | 0.004 | 0.262 | 0.507 | 0.829 | 0.994 | 0.567 | 0.0985849 |
| $I_n^B$ | 0.004 | 0.214 | 0.475 | 0.778 | 0.998 | 0.564 | 0.09622541 |
| $I_n^{Bb}$ | 0.002 | 0.174 | 0.463 | 0.703 | 0.988 | 0.529 | 0.08825601 |
| $I_n$ | 0.004 | 0.125 | 0.345 | 0.532 | 0.994 | 0.407 | 0.06814273 |

# Chapter 4

# Another Test Statistic $J_n$

In this chapter we propose a new test statistic which is a modified version of Cheng and Chu (2004)[6] by using the structure of Hellinger distance. The limiting distribution of the new test statistic is derived so that the goodness-of-fit test can be performed without the bootstrap procedure.

## 4.1 Review of Cheng and Chu (2004)

Cheng and Chu (2004)[6] proposed a goodness-of-fit test statistic defined as the integrated squared difference between two competing kernel density estimators $\hat{g}$ in (1.3) and $\tilde{g}$ in (1.4). Their test statistic is defined as (1.6),

$$J_n^C = \int \big(\hat{g}(t) - \tilde{g}(t)\big)^2 dt.$$

As in Lemma 2, $\hat{g}$ is the semiparametric kernel density estimator of the underlying reference density $g$,

$$\hat{g}(t) = \frac{1}{n_0} \sum_{i=1}^{n} \frac{1}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))} K_b(t - t_i).$$

We see that $\hat{g}(\cdot)$ depends on$(\hat{\alpha}, \hat{\beta})$ which is the maximum likelihood estimator of $(\alpha, \beta)$, derived by solving the two score equations in (2.8). From Lemma 1 we know that $(\hat{\alpha}, \hat{\beta})$ is strongly consistent. Cheng and Chu (2004)[6] applied a second-order Taylor expansion of the semiparametric kernel density estimate $\hat{g}$ at $(\alpha, \beta)$. Then

$\hat{g}$ is decomposed into several parts. Next they used the decomposition of $\hat{g}$ and followed the idea of degenerate U-statistics as in Hall (1984)[16] to decompose the test statistic $J_n^C$ into terms with centered components. One term was asymptotically normal and the other terms were negligible. Therefore the test statistic $J_n^C$ is asymptotically normal. For completeness, in the next section we will discuss their results.

## 4.2 Notation and Decomposition of $\hat{g}(\cdot)$

We shall need the decomposition of $\hat{g}$ in our development. As in (2.4), $w(t) = \exp(\alpha + \beta' h(t))$. Define

$$p(t) = \frac{1}{n_0 + n_1 w(t)} = \frac{1}{n_0} \cdot \frac{1}{1 + \rho w(t)} \tag{4.1}$$

and

$$\hat{w}(t) = \exp(\hat{\alpha} + \hat{\beta}' h(t))$$
$$\hat{p}(t) = \frac{1}{n_0 + n_1 \hat{w}(t)} = \frac{1}{n_0} \cdot \frac{1}{1 + \rho \hat{w}(t)},$$

where $(\hat{\alpha}, \hat{\beta})$ are the MLE that we derived by solving (2.8).

Then

$$\frac{\partial p}{\partial \alpha} = -\frac{1}{n_0} \frac{\rho w(t)}{(1 + \rho w(t))^2} = -p^2(t) n_1 w(t)$$
$$\frac{\partial p}{\partial \beta} = -\frac{1}{n_0} \frac{\rho w(t) h(t)}{(1 + \rho w(t))^2} = -p^2(t) n_1 w(t) h(t).$$

And as in (2.9),

$$\hat{p}(t_i) \equiv \frac{1}{n_0} \cdot \frac{1}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))} = \frac{1}{n_0} \cdot \frac{1}{1 + \rho \hat{w}(t_i)}.$$

Recalling the consistency of $(\hat{\alpha}, \hat{\beta})$, we apply a second-order Taylor expansion to $\hat{p}(t)$ at $(\alpha, \beta)$,

$$
\begin{aligned}
&\hat{p}(t) \\
&= p(t) + \begin{pmatrix} \frac{\partial p}{\partial \alpha} & \frac{\partial p}{\partial \beta'} \end{pmatrix} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\beta} - \beta \end{pmatrix} + \mathcal{O}\Big(p(t)\big((\hat{\alpha} - \alpha) + h(t)'(\hat{\beta} - \beta)\big)^2\Big) \\
&= p(t) - p^2(t) n_1 w(t)\Big((\hat{\alpha} - \alpha) + h(t)'(\hat{\beta} - \beta)\Big) + \mathcal{O}\Big(p(t)\big((\hat{\alpha} - \alpha) \\
&\quad + h(t)'(\hat{\beta} - \beta)\big)^2\Big).
\end{aligned}
$$

As in Lemma 2,

$$
\begin{aligned}
&\hat{g}(t) \\
&= \frac{1}{n_0} \sum_{i=1}^{n} \frac{1}{1 + \rho \exp(\hat{\alpha} + \hat{\beta}' h(t_i))} K_b(t - t_i) \\
&= \sum_{i=1}^{n} \hat{p}(t_i) K_b(t - t_i) \\
&= \sum_{i=1}^{n} \Big[p(t_i) - p^2(t_i) n_1 w(t_i)\Big((\hat{\alpha} - \alpha) + h(t_i)'(\hat{\beta} - \beta)\Big) + \mathcal{O}\Big(p(t_i)\big((\hat{\alpha} - \alpha) \\
&\qquad + h(t_i)'(\hat{\beta} - \beta)\big)^2\Big)\Big] \cdot K_b(t - t_i) \\
&= Q_1(t) - q_2(t)(\hat{\alpha} - \alpha) - q_3(t)'(\hat{\beta} - \beta) + Q_5(t), \qquad (4.2)
\end{aligned}
$$

where

$$
\begin{aligned}
Q_1(t) &= \sum_{i=1}^{n} p(t_i) K_b(t - t_i) \\
Q_2(t) &= \sum_{i=1}^{n} p^2(t_i) n_1 w(t_i) K_b(t - t_i) \\
q_2(t) &= E(Q_2(t)) \\
Q_3(t) &= \sum_{i=1}^{n} p^2(t_i) n_1 w(t_i) h(t_i) K_b(t - t_i) \\
q_3(t) &= E(Q_3(t)) \\
Q_4(t) &= \mathcal{O}\Big( \sum_{i=1}^{n} p(t_i) \big( (\hat{\alpha} - \alpha) + h(t_i)'(\hat{\beta} - \beta) \big)^2 K_b(t - t_i) \Big) \\
Q_5(t) &= \big( q_2(t) - Q_2(t) \big)(\hat{\alpha} - \alpha) + \big( q_3(t) - Q_3(t) \big)'(\hat{\beta} - \beta) + Q_4(t).
\end{aligned}
$$

The following assumptions are needed.

(A1). The probability density function $g$ is positive on $\mathbb{R}$ and has two Lipschitz continuous derivatives.

(A2). The kernel function $K$ is a Lipschitz continuous and symmetric probability density function with support $[-1, 1]$.

(A3). $n_1/n_0 \to \rho$ as $n \to \infty$, and the value of the bandwidth $b$ satisfies $b \to 0$ and $nb \to \infty$, as $n \to \infty$.

Let $g^{(2)}$ be the second derivative of $g$, and let

$$
\begin{aligned}
\kappa_{sq} &= \int_{-1}^{1} K(x)^2 dx \\
\kappa_2 &= \int_{-1}^{1} x^2 K(x) dx \\
M(t) &= \frac{1}{1+\rho} + \frac{\rho}{1+\rho} w(t) = \frac{1}{np(t)}.
\end{aligned}
$$

**Lemma 5** (Cheng and Chu(2004)). *If model (1.1) holds and assumption (A1)-(A3) are satisfied, then we have the following asymptotic results:*

$$E(Q_1(t)) = g(t) + \frac{1}{2}b^2\kappa_2 g^{(2)}(t) + \mathcal{O}(b^3) \tag{4.3}$$

$$Var(Q_1(t)) = \frac{1}{nb}\cdot\frac{\kappa_{sq}g(t)}{M(t)} + \mathcal{O}(\frac{1}{n}) \tag{4.4}$$

$$E(Q_2(t)) = \frac{\rho}{1+\rho}\cdot\frac{w(t)g(t)}{M(t)} + \mathcal{O}(b^2) \tag{4.5}$$

$$Var(Q_2(t)) = \frac{1}{nb}\Big(\frac{\rho}{1+\rho}w(t)\Big)^2\frac{\kappa_{sq}g(t)}{M^3(t)} + \mathcal{O}(\frac{1}{n}) \tag{4.6}$$

$$E(Q_3(t)) = \frac{\rho}{1+\rho}\cdot\frac{h(t)w(t)g(t)}{M(t)} + \mathcal{O}(b^2) \tag{4.7}$$

$$Var(Q_3(t)) = \frac{1}{nb}\Big(\frac{\rho}{1+\rho}w(t)\Big)^2\frac{\kappa_{sq}h(t)'h(t)g(t)}{M^3(t)} + \mathcal{O}(\frac{1}{n}) \tag{4.8}$$

$$E(Q_4(t)) = \mathcal{O}(\frac{1}{n}) \tag{4.9}$$

$$E(Q_4(t)^2) = \mathcal{O}(\frac{1}{n^2}) \tag{4.10}$$

$$E(\hat{\alpha}-\alpha) = o(\frac{1}{\sqrt{n}}),\ E(\hat{\beta}-\beta) = o(\frac{1}{\sqrt{n}}),\ Var(\hat{\alpha}) = \mathcal{O}(\frac{1}{n}),\ Var(\hat{\beta}) = \mathcal{O}(\frac{1}{n}). \tag{4.11}$$

The proof of Lemma 5 is provided in Appendix A.

## 4.3 A New Statistic

We define our test statistic as the Hellinger distance,

$$J_n = \int \big(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\big)^2 dt. \tag{4.12}$$

Note that in Chapter 3 our new test statistic is $I_n = \int \Big(\sqrt{nb}(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)})\Big)^2 dt$. We started from deriving weak convergence of $\sqrt{nb}(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)})$ which is inside of the integral. In this chapter we decompose $J_n$ entirely.

We shall use the Epanechnikov kernel which was first used in density estimation by Epanechnikov (1969) [8]

$$K(x) = \frac{3}{4}(1-x^2)I_{[-1,1]}(x) \tag{4.13}$$

where $I_{[-1,1]}(x)$ is an indicator function. In this chapter and in the following chapters, all the kernel functions used in statistics $J_n$ as (4.12) and $J_n^C$ as (1.6) are the Epanechnikov kernel. The interval of the integral in our statistic defined in (4.12) is actually bounded by the largest value in the samples. Let $L = \max_{1\le i\le n}|t_i| + 1$. Then all the integrals are considered in $[-L, L]$.

Consider

$$\begin{aligned}
&\big(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\big)^2 = \left(\frac{\hat{g}(t)-\tilde{g}(t)}{\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)}}\right)^2 \\
&= \frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)} + \left(\Big(\frac{\hat{g}(t)-\tilde{g}(t)}{\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)}}\Big)^2 - \frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)}\right) \\
&= \frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)} + \Big(4g(t) - (\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2\Big)\frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)(\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2} \\
&= \frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)} + \big(2\sqrt{g(t)} - \sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\big)\big(2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}\big) \\
&\quad \times \frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)(\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2} \\
&= \frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)} + \big(\sqrt{g(t)} - \sqrt{\hat{g}(t)}\big)\Big(\frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2}\Big)(\hat{g}(t)-\tilde{g}(t))^2 \\
&\quad + \big(\sqrt{g(t)} - \sqrt{\tilde{g}(t)}\big)\Big(\frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2}\Big)(\hat{g}(t)-\tilde{g}(t))^2 \\
&= \frac{(\hat{g}(t)-\tilde{g}(t))^2}{4g(t)} \\
&\quad + (g(t)-\hat{g}(t))\frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2(\sqrt{g(t)}+\sqrt{\hat{g}(t)})}(\hat{g}(t)-\tilde{g}(t))^2 \\
&\quad + (g(t)-\tilde{g}(t))\frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2(\sqrt{g(t)}+\sqrt{\tilde{g}(t)})}(\hat{g}(t)-\tilde{g}(t))^2.
\end{aligned}$$

Let

$$J_n^0 = \int \frac{(\hat{g}(t) - \tilde{g}(t))^2}{4g(t)} dt \tag{4.14}$$

$$J_n^1 = \int (g(t) - \hat{g}(t)) \frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)})^2(\sqrt{g(t)} + \sqrt{\hat{g}(t)})} (\hat{g}(t) - \tilde{g}(t))^2 dt$$

$$J_n^2 = \int (g(t) - \tilde{g}(t)) \frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)})^2(\sqrt{g(t)} + \sqrt{\tilde{g}(t)})} (\hat{g}(t) - \tilde{g}(t))^2 dt,$$

then $J_n = J_n^0 + J_n^1 + J_n^2$. In the next section we focus on $J_n^0$.

## 4.4 Decomposition of $J_n^0$

Let

$$Q^r(t) = q_2(t)(\hat{\alpha} - \alpha) + q_3(t)'(\hat{\beta} - \beta) - Q_5(t)$$

then

$$\hat{g}(t) = Q_1(t) - Q^r(t).$$

Let

$$q_0(t) = E(\tilde{g}(t)), \qquad q_1(t) = E(Q_1(t)).$$

Then we have

$$\begin{aligned}
J_n^0 &= \int \frac{(\hat{g}(t) - \tilde{g}(t))^2}{4g(t)} dt \\
&= \int \frac{1}{4g(t)} (Q_1(t) - Q^r(t) - \tilde{g}(t))^2 dt \\
&= \int \frac{1}{4g(t)} \Big\{ \Big[ E(Q_1(t)) - E(\tilde{g}(t)) \Big] + \Big[ \big(Q_1(t) - E(Q_1(t))\big) \\
&\qquad\qquad - \big(\tilde{g}(t) - E(\tilde{g}(t))\big) \Big] - Q^r(t) \Big\}^2 dt \\
&= \int \left\{ \frac{\Big(q_1(t) - q_0(t)\Big) + \Big(\big(Q_1(t) - q_1(t)\big) - \big(\tilde{g}(t) - q_0(t)\big)\Big) - Q^r(t)}{2\sqrt{g(t)}} \right\}^2 dt.
\end{aligned}$$

Define

$$J_1 = \int \Big(\frac{q_1(t) - q_0(t)}{2\sqrt{g(t)}}\Big)^2 dt$$

$$J_2 = \int \Big(\frac{\big(Q_1(t) - q_1(t)\big) - \big(\tilde{g}(t) - q_0(t)\big)}{2\sqrt{g(t)}}\Big)^2 dt$$

$$J_3 = \int \Big(\frac{Q^r(t)}{2\sqrt{g(t)}}\Big)^2 dt$$

$$J_4 = \frac{1}{2}\int \frac{q_1(t) - q_0(t)}{g(t)}\Big[\big(Q_1(t) - q_1(t)\big) - \big(\tilde{g}(t) - q_0(t)\big)\Big] dt$$

$$J_5 = -\frac{1}{2}\int \frac{q_1(t) - q_0(t)}{g(t)} Q^r(t) dt$$

$$J_6 = -\frac{1}{2}\int \frac{\big(Q_1(t) - q_1(t)\big) - \big(\tilde{g}(t) - q_0(t)\big)}{g(t)} Q^r(t) dt.$$

Then, $J_n^0 = J_1 + J_2 + J_3 + J_4 + J_5 + J_6$.

## 4.5 Asymptotic Result for $J_n^0$

In this section, we will prove that $J_2$ is asymptotically normal and that the other terms $J_1$,$J_3$,$J_4$,$J_5$ and $J_6$ all are negligible. Since $J_n^0 = J_1 + J_2 + J_3 + J_4 + J_5 + J_6$, we have that $J_n^0$ is asymptotically normal with the same distribution as $J_2$.

**Lemma 6.**

$$n\sqrt{b}\Big(J_2 - \frac{1}{4nb}\rho\kappa_{sq}\int \frac{w(t)}{M(t)} dt\Big) \xrightarrow{d} \mathbf{N}\Big(0,\ \frac{1}{8}\rho^2\kappa_c \int \Big(\frac{w(t)}{M(t)}\Big)^2 dt\Big)$$

*where*

$$\kappa_c = \int \Big(\int K(x)K(y-x)dx\Big)^2 dy$$

*is a constant number for fixed $K(\cdot)$ with compact support.*

In this chapter $K(\cdot)$ is the Epanechnikov kernel. We provide the value of $\kappa_c$ in Appendix B.

*Proof.* Notice that

$$\frac{1}{n_0} - p(t) = \rho p(t) w(t).$$

So

$$\begin{aligned}
&\Big(Q_1(t) - q_1(t)\Big) - \Big(\tilde{g}(t) - q_0(t)\Big) \\
&= \Big(\sum_{i=1}^{n} p(t_i) K_b(t - t_i) - E(\sum_{i=1}^{n} p(t_i) K_b(t - t_i))\Big) \\
&\quad - \Big(\frac{1}{n_0} \sum_{i=1}^{n_0} K_b(t - u_i) - E(\frac{1}{n_0} \sum_{i=1}^{n_0} K_b(t - u_i))\Big) \\
&= \sum_{j=1}^{n_1} \Big(p(z_j) K_b(t - z_j) - E(p(z_j) K_b(t - z_j))\Big) \\
&\quad - \sum_{i=1}^{n_0} \Big(\rho p(u_i) w(u_i) K_b(t - u_i) - E(\rho p(u_i) w(u_i) K_b(t - u_i))\Big).
\end{aligned}$$

Define

$$X_i(t) = -\frac{1}{2\sqrt{g(t)}} \Big(\rho p(u_i) w(u_i) K_b(t - u_i) - E(\rho p(u_i) w(u_i) K_b(t - u_i))\Big)$$

$$i = 1, 2, ..., n_0$$

$$Y_j(t) = \frac{1}{2\sqrt{g(t)}} \Big(p(z_j) K_b(t - z_j) - E(p(z_j) K_b(t - z_j))\Big) \quad j = 1, 2, ..., n_1$$

$$\Big\{T_k(t)\Big\} = \Big\{X_1(t), X_2(t), ..., X_{n_0}(t), Y_1(t), Y_2(t), ..., Y_{n_1}(t)\Big\}$$

$$k = 1, 2, ..., n_0, ..., n.$$

Then

$$\begin{aligned}
J_2 &= \int \Big(\frac{\big(Q_1(t) - q_1(t)\big) - \big(\tilde{g}(t) - q_0(t)\big)}{2\sqrt{g(t)}}\Big)^2 dt \\
&= \int \Big(\sum_{k=1}^{n} T_k(t)\Big)^2 dt \\
&= \int \sum_{k=1}^{n} T_k^2(t) dt + \int 2 \sum_{1 \le i < j \le n} T_i(t) T_j(t) \; dt. \qquad (4.15)
\end{aligned}$$

Consider the first part in (4.15). From the calculation of $n_0 EX_i^2(t) + n_1 EY_j^2(t)$ below, we can see that $EX_i^2(t)$ and $EY_j^2(t)$ both are continuous functions of $t$. Since the integral is considered in a bounded interval $[-L, L]$, we have $E|\int X_i^2(t)dt| < \infty$ and $E|\int Y_j^2(t)dt| < \infty$. Thus

$$
\begin{aligned}
&\int \sum_{k=1}^{n} T_k^2(t)\ dt = \int \sum_{i=1}^{n_0} X_i^2(t) + \sum_{j=1}^{n_1} Y_j^2(t)\ dt \\
&= n_0 \cdot \frac{\sum_{i=1}^{n_0} \int X_i^2(t)dt}{n_0} + n_1 \cdot \frac{\sum_{j=1}^{n_1} \int Y_j^2(t)dt}{n_1} \\
&\xrightarrow[a.s.]{LLN} n_0 E\Big(\int X_i^2(t)dt\Big) + n_1 E\Big(\int Y_j^2(t)dt\Big) = \int \Big(n_0 EX_i^2(t) + n_1 EY_j^2(t)\Big)dt.
\end{aligned}
\tag{4.16}
$$

Next we calculate $n_0 EX_i^2(t) + n_1 EY_j^2(t)$.

$$
\begin{aligned}
&n_0 EX_i^2(t) + n_1 EY_j^2(t) \\
&= \frac{n_0}{4g(t)} E\Big(\rho p(u_i)w(u_i)K_b(t-u_i) - E(\rho p(u_i)w(u_i)K_b(t-u_i))\Big)^2 \\
&\quad + \frac{n_1}{4g(t)} E\Big(p(z_j)K_b(t-z_j) - E(p(z_j)K_b(t-z_j))\Big)^2 \\
&= \frac{n_0}{4g(t)} E\Big(\rho p(u_i)w(u_i)K_b(t-u_i)\Big)^2 - \frac{n_0}{4g(t)}\Big(E(\rho p(u_i)w(u_i)K_b(t-u_i))\Big)^2 \\
&\quad + \frac{n_1}{4g(t)} E\Big(p(z_j)K_b(t-z_j)\Big)^2 - \frac{n_1}{4g(t)}\Big(E(p(z_j)K_b(t-z_j))\Big)^2 \\
&= \frac{n_0}{4g(t)} \int \rho^2 p^2(x) w^2(x) K_b^2(t-x) g(x)dx - \frac{n_0}{4g(t)}\Big(\int \rho p(x)w(x)K_b(t-x)g(x)dx\Big)^2 \\
&\quad + \frac{n_1}{4g(t)} \int p^2(x) K_b^2(t-x) w(x) g(x)dx - \frac{n_1}{4g(t)}\Big(\int p(x)K_b(t-x)w(x)g(x)dx\Big)^2 \\
&= \frac{1}{4g(t)} \int \Big(n_0\rho^2 w(x) + n_1\Big) p^2(x) K_b^2(t-x) w(x) g(x)dx \\
&\quad - \frac{n_0\rho^2 + n_1}{4g(t)}\Big(\int p(x)K_b(t-x)w(x)g(x)dx\Big)^2 \\
&= \frac{\rho}{4g(t)} \int p(x) K_b^2(t-x) w(x) g(x)dx - \frac{n\rho}{4g(t)}\Big(\int p(x)K_b(t-x)w(x)g(x)dx\Big)^2.
\end{aligned}
$$

Let $\frac{t-x}{b} = y$, then $x = t - by$. Since $K(\cdot)$ is symmetric, $\int yK(y)dy = 0$ and $\int yK^2(y)dy = 0$. Apply the Taylor expansion to $p(\cdot), w(\cdot)$ and $g(\cdot)$,

$$\begin{aligned}
&\int p(x)K_b(t-x)w(x)g(x)dx \\
&= \int p(t-by)\frac{1}{b}K(y)w(t-by)g(t-by)bdy \\
&= \int \big(p(t) - byp'(t)\big)K(y)\big(w(t) - byw'(t)\big)\big(g(t) - byg'(t)\big)dy + \mathcal{O}(\frac{b^2}{n}) \\
&= p(t)w(t)g(t)\int k(y)dy + \mathcal{O}(\frac{b^2}{n}) \\
&= \frac{1}{n}\cdot\frac{w(t)g(t)}{M(t)} + \mathcal{O}(\frac{b^2}{n}) \qquad (4.17)
\end{aligned}$$

and

$$\begin{aligned}
&\int p(x)K_b^2(t-x)w(x)g(x)dx \\
&= \int p(t-by)\frac{1}{b^2}K^2(y)w(t-by)g(t-by)bdy \\
&= \frac{1}{b}\int \big(p(t) - byp'(t)\big)K^2(y)\big(w(t) - byw'(t)\big)\big(g(t) - byg'(t)\big)dy + \mathcal{O}(\frac{b}{n}) \\
&= \frac{1}{b}p(t)w(t)g(t)\int K^2(y)dy + \mathcal{O}(\frac{b}{n}) \\
&= \frac{1}{nb}\cdot\frac{w(t)g(t)}{M(t)}\kappa_{sq} + \mathcal{O}(\frac{b}{n}), \qquad (4.18)
\end{aligned}$$

so

$$\begin{aligned}
n_0EX_i^2(t) + n_1EY_j^2(t) &= \frac{\rho}{4g(t)}\cdot\frac{1}{nb}\cdot\frac{w(t)g(t)}{M(t)}\kappa_{sq} - \frac{n\rho}{4g(t)}\Big(\frac{1}{n}\frac{w(t)g(t)}{M(t)}\Big)^2 \\
&= \frac{\rho}{4nb}\cdot\frac{w(t)}{M(t)}\kappa_{sq} - \frac{\rho}{4ng(t)}(\frac{w(t)}{M(t)})^2g(t) \\
&= \frac{\rho}{4nb}\cdot\frac{w(t)}{M(t)}\kappa_{sq} - \mathcal{O}(\frac{1}{n}).
\end{aligned}$$

Then for the first part in (4.15), we have

$$\int\sum_{k=1}^{n}T_k^2(t)dt \xrightarrow[a.s.]{LLN} \int\frac{\rho}{4nb}\cdot\frac{w(t)}{M(t)}\kappa_{sq}dt.$$

Next we investigate the 2nd part of (4.15). Note that $\{T_k(t)\}$ are independent and $E(T_k(t)) = 0$. Furthermore,

$$E(\int T_i(t)T_j(t)dt) = \int E\Big\{E\Big(T_i(t)T_j(t)\Big|t_i\Big)\Big\}dt = 0.$$

Let

$$w_{ij} = \begin{cases} 2\int T_i(t)T_j(t)dt & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

then

$$\int 2 \sum_{1\le i<j\le n} T_i(t)T_j(t)dt = \sum_{1\le i<j\le n} w_{ij} \tag{4.19}$$

and $E(w_{ij}|t_i) = 0$. Thus $\sum_{1\le i<j\le n} w_{ij}$ satisfies the 'clean' condition as in Jong (1987)[17]. Next we calculate the variance of $\sum_{1\le i<j\le n} w_{ij}$.

Define

$$\begin{aligned} W_a &= \sum_{1\le i<j\le n} w_{ij}^2 \\ W_b &= \sum_{1\le i<j<k\le n} \big(w_{ij}w_{ik} + w_{ji}w_{jk} + w_{ki}w_{kj}\big) \\ W_c &= \sum_{1\le i<j<k<l\le n} \big(w_{ij}w_{kl} + w_{ik}w_{jl} + w_{il}w_{jk}\big). \end{aligned}$$

Then

$$\big(\sum_{1\le i<j\le n} w_{ij}\big)^2 = W_a + 2W_b + 2W_c.$$

Note that for any general term in $W_b$, the product $w_{ij}w_{ik}$ has zero expectation since

$$E(w_{ij}w_{ik}) = E(E(w_{ij}w_{ik}|t_i,t_k)) = Ew_{ik}E(w_{ij}|t_i) = 0.$$

So we have $EW_b = 0$, and obviously, $EW_c = 0$. Let $\sigma^2(n)$ be the variance of $\sum_{1\le i<j\le n} w_{ij}$. Then

$$\sigma^2(n) = Var(\sum_{1\le i<j\le n} w_{ij}) = E(\sum_{1\le i<j\le n} w_{ij})^2 = E(W_a + W_b + W_c) = EW_a.$$

Thus

$$
\begin{aligned}
\sigma^2(n) &= E(\sum_{1\le i<j\le n} w_{ij}^2) = 4\sum_{1\le i<j\le n} E\Big(\int T_i(t)T_j(t)dt\Big)^2 \\
&= 4\sum_{1\le i<j\le n_0} E\Big(\int X_i(t)X_j(t)dt\Big)^2 + 4\sum_{1\le i<j\le n_1} E\Big(\int Y_i(t)Y_j(t)dt\Big)^2 \\
&\quad + 4\sum_{\substack{1\le i\le n_0\\ 1\le j\le n_1}} E\Big(\int X_i(t)Y_j(t)dt\Big)^2 \\
&= 4\sum_{1\le i<j\le n_0} E\Big(\int X_i(s)X_j(s)ds\int X_i(t)X_j(t)dt\Big) \\
&\quad + 4\sum_{1\le i<j\le n_1} E\Big(\int Y_i(s)Y_j(s)ds\int Y_i(t)Y_j(t)dt\Big) \\
&\quad + 4\sum_{\substack{1\le i\le n_0\\ 1\le j\le n_1}} E\Big(\int X_i(s)Y_j(s)ds\int X_i(t)Y_j(t)dt\Big) \\
&= 4\sum_{1\le i<j\le n_0} \int\int E\Big(X_i(s)X_i(t)\Big)E\Big(X_j(s)X_j(t)\Big)dsdt \\
&\quad + 4\sum_{1\le i<j\le n_1} \int\int E\Big(Y_i(s)Y_i(t)\Big)E\Big(Y_j(s)Y_j(t)\Big)dsdt \\
&\quad + 4\sum_{\substack{1\le i\le n_0\\ 1\le j\le n_1}} \int\int E\Big(X_i(s)X_i(t)\Big)E\Big(Y_j(s)Y_j(t)\Big)dsdt \\
&= 4\int\int \frac{n_0(n_0-1)}{2}\Big(E(X_i(s)X_i(t))\Big)^2 + \frac{n_1(n_1-1)}{2}\Big(E(Y_j(s)Y_j(t))\Big)^2 \\
&\quad + n_0n_1E\Big(X_i(s)X_i(t)\Big)E\Big(Y_j(s)Y_j(t)\Big)dsdt \\
&= 2\int\int \Big(n_0EX_i(s)X_i(t) + n_1EY_j(s)Y_j(t)\Big)^2 - n_0\Big(E(X_i(s)X_i(t))\Big)^2 \\
&\quad - n_1\Big(E(Y_j(s)Y_j(t))\Big)^2 dsdt.
\end{aligned}
$$

In order to obtain the value of $\sigma(n)$, we need to calculate $E(X_i(s)X_i(t))$ and

$E(Y_j(s)Y_j(t))$. We have,

$$
\begin{aligned}
& E\big(X_i(s)X_i(t)\big) \\
&= E\Big\{\frac{1}{4\sqrt{g(s)g(t)}}\Big(\rho p(u_i)w(u_i)K_b(s-u_i) - E(\rho p(u_i)w(u_i)K_b(s-u_i))\Big) \\
&\qquad \times \Big(\rho p(u_i)w(u_i)K_b(t-u_i) - E(\rho p(u_i)w(u_i)K_b(t-u_i))\Big)\Big\} \\
&= \frac{\rho^2}{4\sqrt{g(s)g(t)}}\Big\{E\Big(p^2(u_i)w^2(u_i)K_b(s-u_i)K_b(t-u_i)\Big) \\
&\qquad\qquad - E\Big(p(u_i)w(u_i)K_b(s-u_i)\Big)E\Big(p(u_i)w(u_i)K_b(t-u_i)\Big)\Big\} \\
&= \frac{\rho^2}{4\sqrt{g(s)g(t)}}\Big\{\int p^2(x)w^2(x)K_b(s-x)K_b(t-x)g(x)dx \\
&\qquad\qquad - \int p(x)w(x)K_b(s-x)g(x)dx \cdot \int p(x)w(x)K_b(t-x)g(x)dx\Big\}.
\end{aligned}
$$

Let $\frac{t-x}{b} = y$, then $x = t - by$ and $x - s = (t-s) - by$. We have

$$
\begin{aligned}
& \int p^2(x)w^2(x)K_b(s-x)K_b(t-x)g(x)dx \\
&= \int p^2(t-by)w^2(t-by)\frac{1}{b^2}K(y)K(\frac{t-s}{b}-y)g(t-by)bdy \\
&= \frac{1}{b}p^2(t)w^2(t)g(t)\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2}) \\
&= \frac{1}{n^2 b}(\frac{w(t)}{M(t)})^2 g(t)\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2}).
\end{aligned} \tag{4.20}
$$

Combining this with (4.17), we have

$$
\begin{aligned}
& E(X_i(s)X_i(t)) \\
&= \frac{\rho^2}{4\sqrt{g(s)g(t)}}\Big\{\frac{1}{n^2 b}(\frac{w(t)}{M(t)})^2 g(t)\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2}) \\
&\qquad\qquad - \Big(\frac{1}{n}\cdot\frac{w(s)g(s)}{M(s)} + \mathcal{O}(\frac{b^2}{n})\Big)\Big(\frac{1}{n}\cdot\frac{w(t)g(t)}{M(t)} + \mathcal{O}(\frac{b^2}{n})\Big)\Big\} \\
&= \frac{\rho^2}{4n^2 b}(\frac{w(t)}{M(t)})^2\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2}).
\end{aligned} \tag{4.21}
$$

Similarly,

$$
\begin{aligned}
&E(Y_j(s)Y_j(t)) \\
&= E\Big\{\frac{1}{4\sqrt{g(s)g(t)}}\Big(p(z_j)K_b(s-z_j) - E(p(z_j)K_b(s-z_j))\Big) \\
&\qquad \times \Big(p(z_j)K_b(t-z_j) - E(p(z_j)K_b(t-z_j))\Big)\Big\} \\
&= \frac{1}{4\sqrt{g(s)g(t)}}\Big\{E\Big(p^2(z_j)K_b(s-z_j)K_b(t-z_j)\Big) \\
&\qquad\qquad\qquad - E\Big(p(z_j)K_b(s-z_j)\Big)E\Big(p(z_j)K_b(t-z_j)\Big)\Big\} \\
&= \frac{1}{4\sqrt{g(s)g(t)}}\Big\{\int p^2(x)K_b(s-x)K_b(t-x)w(x)g(x)dx \\
&\qquad\qquad\qquad - \int p(x)K_b(s-x)w(x)g(x)dx \cdot \int p(x)K_b(t-x)w(x)g(x)dx\Big\} \\
&= \frac{1}{4\sqrt{g(s)g(t)}}\Big\{p^2(t)w(t)g(t)\frac{1}{b}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2}) \\
&\qquad\qquad\qquad - \frac{1}{n^2}(\frac{w(s)g(s)}{M(s)})(\frac{w(t)g(t)}{M(t)}) - \mathcal{O}(\frac{b^2}{n^2})\Big\} \\
&= \frac{1}{4n^2b}\cdot\frac{w(t)}{M^2(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2}). \qquad (4.22)
\end{aligned}
$$

So

$$
\begin{aligned}
&n_0EX_i(s)X_i(t) + n_1EY_j(s)Y_j(t) \\
&= n_0\frac{\rho^2}{4n^2b}(\frac{w(t)}{M(t)})^2\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy \\
&\quad + n_1\frac{1}{4n^2b}\cdot\frac{w(t)}{M^2(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy \\
&\quad + \mathcal{O}(\frac{1}{n}) \\
&= \frac{\rho}{4nb}\cdot\frac{w(t)}{M(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n}).
\end{aligned}
$$

Let $(t-s)/b = r$, then $s = t - br$. Therefore,

$$
\begin{aligned}
\sigma^2(n) = 2\int\int \Big(n_0 EX_i(s)X_i(t) + n_1 EY_j(s)Y_j(t)\Big)^2 - n_0\Big(E(X_i(s)X_i(t))\Big)^2 \\
- n_1\Big(E(Y_j(s)Y_j(t))\Big)^2 dsdt \\
= 2\int\int \Big(\frac{\rho}{4nb}\cdot\frac{w(t)}{M(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy\Big)^2 dsdt \\
- 2n_0\int\int \Big(\frac{\rho^2}{4n^2b}(\frac{w(t)}{M(t)})^2\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy\Big)^2 dsdt \\
- 2n_1\int\int \Big(\frac{1}{4n^2b}\cdot\frac{w(t)}{M^2(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy\Big)^2 dsdt \\
= \int \Big(\frac{\rho^2}{8n^2b^2}\cdot\frac{w^2(t)}{M^2(t)} - \frac{n_0\rho^4}{8n^4b^2}\cdot\frac{w^4(t)}{M^4(t)} - \frac{n_1}{8n^4b^2}\cdot\frac{w^2(t)}{M^4(t)}\Big)g(t)\int\frac{1}{g(t-br)} \\
\times (\int K(y)K(r-y)dy\Big)^2 bdrdt \\
= \frac{\rho^2}{8n^2b}\kappa_c\int\left(\frac{w(t)}{M(t)}\right)^2 dt + \mathcal{O}(\frac{1}{n^3b}).
\end{aligned}
\tag{4.23}
$$

In order to Apply Theorem (2.1) in Jong (1987)[17], we need to check two conditions in Theorem (2.1) in Jong (1987)[17]:

$$
\sigma(n)^{-2}\max_{1\le i\le n}\sum_{1\le j\le n} Ew_{ij}^2 \to 0, \qquad n\to\infty. \tag{4.24}
$$

$$
\sigma(n)^{-4}E(\sum_{1\le i<j\le n} w_{ij})^4 \to 3, \qquad n\to\infty. \tag{4.25}
$$

For condition (4.24),

$$
\begin{aligned}
\sum_{1\le j\le n} Ew_{ij}^2 &= 4\sum_{1\le j\le n} E(\int T_i(t)T_j(t)dt)^2 \\
&= 4E\Big(\sum_{1\le j\le n_0}(\int T_i(t)X_j(t)dt)^2 + \sum_{n_0<j\le n}(\int T_i(t)Y_{j-n_0}(t)dt)^2\Big).
\end{aligned}
$$

For $1 \leq i \leq n_0$,

$$
\begin{aligned}
& \sum_{1\leq j\leq n} Ew_{ij}^2 \\
= \ & 4E\Big(\sum_{1\leq j\leq n_0}(\int X_i(t)X_j(t)dt)^2 + \sum_{n_0<j\leq n}(\int X_i(t)Y_{j-n_0}(t)dt)^2\Big) \\
= \ & 4E\Big(\sum_{1\leq j\leq n_0}\int\int X_i(t)X_j(t)X_i(s)X_j(s)dsdt \\
& \quad + \sum_{n_0<j\leq n}\int\int X_i(t)Y_{j-n_0}(t)X_i(s)Y_{j-n_0}(s)dsdt\Big) \\
= \ & 4\sum_{\substack{1\leq j\leq n_0\\ i\neq j}}\int\int E\Big(X_i(t)X_i(s)\Big)E\Big(X_j(t)X_j(s)\Big)dsdt \\
& +4\sum_{n_0<j\leq n}\int\int E\Big(X_i(t)X_i(s)\Big)E\Big(Y_{j-n_0}(t)Y_{j-n_0}(s)\Big)dsdt \\
= \ & 4(n_0-1)\int\int\Big(\frac{\rho^2}{4n^2b}(\frac{w(t)}{M(t)})^2\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2})\Big)^2 dsdt \\
& +4n_1\int\int\Big(\frac{\rho^2}{4n^2b}(\frac{w(t)}{M(t)})^2\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2})\Big) \\
& \qquad\qquad \times\Big(\frac{1}{4n^2b}\frac{w(t)}{M^2(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2})\Big)dsdt \\
= \ & \frac{\rho^4(n_0-1)}{4n^4b^2}\int\int(\frac{w(t)}{M(t)})^4\cdot\frac{g(t)}{g(t-br)}\Big(\int K(y)K(r-y)dy\Big)^2 bdrdt \\
& +\frac{\rho^2 n_1}{4n^4b^2}\int\int\frac{w^3(t)}{M^4(t)}\cdot\frac{g(t)}{g(t-br)}\Big(\int K(y)K(r-y)dy\Big)^2 bdrdt + \mathcal{O}(\frac{1}{n^3}) \\
= \ & \mathcal{O}(\frac{1}{n^3b}).
\end{aligned}
$$

And for $n_0 < i \le n$, let $k = i - n_0$,

$$
\begin{aligned}
& \sum_{1\le j\le n} E w_{ij}^2 \\
= \;& 4E\Big(\sum_{1\le j\le n_0} (\int Y_k(t)X_j(t)dt)^2 + \sum_{n_0<j\le n} (\int Y_k(t)Y_{j-n_0}(t)dt)^2\Big) \\
= \;& 4E\Big(\sum_{1\le j\le n_0} \int\int Y_k(t)X_j(t)Y_k(s)X_j(s)dsdt \\
& \quad + \sum_{1\le j\le n_1} \int\int Y_k(t)Y_{j-n_0}(t)Y_k(s)Y_{j-n_0}(s)dsdt\Big) \\
= \;& 4\sum_{1\le j\le n_0} \int\int E\Big(Y_k(t)Y_k(s)\Big)E\Big(X_j(t)X_j(s)\Big)dsdt \\
& +4\sum_{\substack{n_0<j\le n\\ k\ne j}} \int\int E\Big(Y_k(t)Y_k(s)\Big)E\Big(Y_{j-n_0}(t)Y_{j-n_0}(s)\Big)dsdt \\
= \;& 4n_0\int\int \Big(\frac{1}{4n^2b}\cdot\frac{w(t)}{M^2(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2})\Big) \\
& \qquad \times\Big(\frac{\rho^2}{4n^2b}(\frac{w(t)}{M(t)})^2\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy + \mathcal{O}(\frac{1}{n^2})\Big)dsdt \\
& +4(n_1-1)\int\int \Big(\frac{1}{4n^2b}\cdot\frac{w(t)}{M^2(t)}\sqrt{\frac{g(t)}{g(s)}}\int K(y)K(\frac{t-s}{b}-y)dy \\
& \qquad\qquad +\mathcal{O}(\frac{1}{n^2})\Big)^2 dsdt \\
= \;& \mathcal{O}(\frac{1}{n^3 b}).
\end{aligned}
\tag{4.26}
$$

Also from (4.23) we know that

$$\sigma^2(n) = \mathcal{O}(\frac{1}{n^2 b}).$$

Thus

$$\sigma(n)^{-2}\max_{1\le i\le n}\sum_{1\le j\le n} E w_{ij}^2 = \mathcal{O}(\frac{1}{n}) \to 0,$$

so condition (4.24) is satisfied.

For condition (4.25), we define

$$
\begin{aligned}
W_I &= \sum_{1\le i<j\le n} Ew_{ij}^4 \\
W_{II} &= \sum_{1\le i<j<k\le n} (Ew_{ij}^2 w_{ik}^2 + Ew_{ji}^2 w_{jk}^2 + Ew_{ki}^2 w_{kj}^2) \\
W_{III} &= \sum_{1\le i<j<k\le n} (Ew_{ij}^2 w_{ki} w_{kj} + Ew_{ik}^2 w_{ji} w_{jk} + Ew_{kj}^2 w_{ij} w_{ik}) \\
W_{IV} &= \sum_{1\le i<j<k<l\le n} (Ew_{ij} w_{ik} w_{lj} w_{lk} + Ew_{ij} w_{il} w_{kj} w_{kl} + Ew_{ik} w_{il} w_{jk} w_{jl}) \\
W_V &= \sum_{1\le i<j<k<l\le n} (Ew_{ij}^2 w_{kl}^2 + Ew_{ik}^2 w_{jl}^2 + Ew_{il}^2 w_{jk}^2),
\end{aligned}
$$

then

$$
\begin{aligned}
\Big(\sum_{1\le i<j\le n} w_{ij}\Big)^4 &= \Big[\big(\sum_{1\le i<j\le n} w_{ij}\big)^2\Big]^2 \\
&= (W_a + 2W_b + 2W_c)^2 \\
&= W_a^2 + 4W_b^2 + 4W_c^2 + 4W_aW_b + 4W_aW_c + 8W_bW_c.
\end{aligned}
$$

By direct calculation, we have

$$
\begin{aligned}
EW_a^2 &= W_I + 2W_{II} + 2W_V \\
EW_b^2 &= W_{II} + 2W_{III} + 4W_{IV} \\
EW_c^2 &= 2W_{IV} + W_V \\
EW_aW_b &= W_{III} \\
EW_aW_c &= 0 \\
EW_bW_c &= 0,
\end{aligned}
$$

so that,

$$
E\Big(\sum_{1\le i<j\le n} w_{ij}\Big)^4 = W_I + 6W_{II} + 12W_{III} + 24W_{IV} + 6W_V. \tag{4.27}
$$

Consider the general term in $W_I$,

$$\begin{aligned} Ew_{ij}^4 &= 16E(\int T_i(t)T_j(t)dt)^4 \\ &= 16\int\int\int\int E(\prod_{k=1}^{4} T_i(t_k)) \cdot E(\prod_{k=1}^{4} T_j(t_k))dt_1dt_2dt_3dt_4. \end{aligned}$$

Next we need to calculate $E(\prod_{k=1}^{4} T_i(t_k))$. Note here $\{t_k\}$ $(k=1,2,3,4)$ are integration variables, not random. For $1 \le i \le n_0$, $T_i(t_k) = X_i(t_k)$ $(k=1,2,3,4.)$

$$\begin{aligned} &E(\prod_{k=1}^{4} T_i(t_k)) = E(\prod_{k=1}^{4} X_i(t_k)) \\ &= \frac{1}{16\sqrt{\prod_{k=1}^{4} g(t_k)}} E\Big\{\prod_{k=1}^{4}\Big[\rho p(u_i)w(u_i)K_b(t_k-u_i) - E\big(\rho p(u_i)w(u_i) \\ &\qquad\qquad \times K_b(t_k-u_i)\big)\Big]\Big\} \\ &= \frac{\rho^4}{16\sqrt{\prod_{k=1}^{4} g(t_k)}}\big(L_1 - L_2 + L_3 - L_4\big) \end{aligned} \tag{4.28}$$

where

$$\begin{aligned} L_1 &= E\Big[p^4(u_i)w^4(u_i)\prod_{k=1}^{4} K_b(t_k-u_i)\Big] \\ L_2 &= \sum_{k=1}^{4}\Big\{E\Big[p^3(u_i)w^3(u_i)\frac{\prod_{m=1}^{4} K_b(t_m-u_i)}{K_b(t_k-u_i)}\Big] \cdot E\Big[p(u_i)w(u_i)K_b(t_k-u_i)\Big]\Big\} \\ L_3 &= \sum_{\substack{\{k_1,k_2,k_3,k_4\}=\{1,2,3,4\},\\ \{1,3,2,4\},\{1,4,2,3\},\{2,3,1,4\},\\ \{2,4,1,3\},\{3,4,1,2\}}} \Big\{E\Big[p^2(u_i)w^2(u_i)K_b(t_{k_1}-u_i)K_b(t_{k_2}-u_i)\Big] \\ &\qquad \times E\Big[p(u_i)w(u_i)K_b(t_{k_3}-u_i)\Big] \cdot E\Big[p(u_i)w(u_i)K_b(t_{k_4}-u_i)\Big]\Big\} \\ L_4 &= 3\prod_{k=1}^{4} E\Big[p(u_i)w(u_i)K_b(t_k-u_i)\Big]. \end{aligned}$$

We will get the order of $L_1, L_2, L_3$ and $L_4$ by deriving general term of them. Let $\frac{t_1-x}{b} = y$, $\frac{t_1-t_2}{b} = s_2$, $\frac{t_1-t_3}{b} = s_3$, $\frac{t_1-t_4}{b} = s_4$, then $t_2 = t_1 - bs_2$, $t_3 = t_1 - bs_3$, and $t_4 = t_1 - bs_4$.

For the general term in $L_4$,

$$\begin{aligned}
Ep(u_i)w(u_i)K_b(t_k-u_i) &= \int p(x)w(x)K_b(t_k-x)g(x)dx \\
&= \int p(t_k-by)w(t_k-by)\frac{1}{b}K(y)g(t_k-by)bdy \\
&= p(t_k)w(t_k)g(t_k)-\mathcal{O}(\frac{b^2}{n}) \\
&= \frac{1}{n}\cdot\frac{w(t_k)g(t_k)}{M(t_k)}+\mathcal{O}(\frac{b^2}{n}) \\
&= \mathcal{O}(\frac{1}{n}). \qquad (4.29)
\end{aligned}$$

Thus $L_4$ in (4.28) is of order $\mathcal{O}(\frac{1}{n^4})$. For the general term in $L_3$,

$$\begin{aligned}
&Ep^2(u_i)w^2(u_i)K_b(t_1-u_i)K_b(t_2-u_i) \\
&= \int p^2(x)w^2(x)K_b(t_1-x)K_b(t_2-x)g(x)dx \\
&= \int p^2(t_1-by)w^2(t_1-by)\frac{1}{b^2}K(y)K(\frac{t_1-t_2}{b}-y)g(t_1-by)bdy \\
&= \frac{1}{b}p^2(t_1)w^2(t_1)g(t_1)\int K(y)K(s_2-y)dy+\mathcal{O}(\frac{1}{n^2}) \\
&= \mathcal{O}(\frac{1}{n^2b}). \qquad (4.30)
\end{aligned}$$

Combining (4.29) and (4.30), we see that $L_3$ in (4.28) is of order $\mathcal{O}(\frac{1}{n^4b})$. Consider the general term in $L_2$,

$$\begin{aligned}
&Ep^3(u_i)w^3(u_i)K_b(t_1-u_i)K_b(t_2-u_i)K_b(t_3-u_i) \\
&= \int p^3(x)w^3(x)K_b(t_1-x)K_b(t_2-x)K_b(t_3-x)g(x)dx \\
&= \int p^3(t_1-by)w^3(t_1-by)\frac{1}{b^3}K(y)K(\frac{t_1-t_2}{b}-y)K(\frac{t_1-t_3}{b}-y)g(t_1-by)bdy \\
&= \frac{1}{n^3b^2}\cdot\frac{w^3(t_1)g(t_1)}{M^3(t_1)}\int K(y)K(s_2-y)K(s_3-y)dy+\mathcal{O}(\frac{1}{n^3b}) \\
&= \mathcal{O}(\frac{1}{n^3b^2}). \qquad (4.31)
\end{aligned}$$

Thus combining (4.31) and (4.29), we see that $L_2$ in (4.28) is of order $\mathcal{O}(\frac{1}{n^4b^2})$. For $L_1$,

$$\begin{aligned}
&Ep^4(u_i)w^4(u_i)\prod_{k=1}^{4}K_b(t_k-u_i)\\
&=\int p^4(x)w^4(x)\prod_{k=1}^{4}K_b(t_k-x)g(x)dx\\
&=\int p^4(t_1-by)w^4(t_1-by)\frac{1}{b^4}K(y)\prod_{k=2}^{4}K(\frac{t_1-t_k}{b}-y)g(t_1-by)bdy\\
&=\frac{1}{n^4b^3}\cdot\frac{w^4(t_1)g(t_1)}{M^4(t_1)}\int K(y)K(s_2-y)K(s_3-y)K(s_4-y)dy+\mathcal{O}(\frac{1}{n^4b^2})\\
&=\mathcal{O}(\frac{1}{n^4b^3}). \qquad (4.32)
\end{aligned}$$

Thus $L_1$ in (4.28) is of order $\mathcal{O}(\frac{1}{n^4b^3})$. Then $L_2$, $L_3$ and $L_4$ are of lower order than $L_1$. So we have

$$E(\prod_{k=1}^{4}T_i(t_k))=\mathcal{O}(\frac{1}{n^4b^3}) \qquad for \qquad 1\le i\le n_0. \qquad (4.33)$$

For $n_0<i\le n$, let $l=i-n_0$, then $T_i(t_k)=Y_l(t_k)$ $(k=1,2,3,4)$. We have

$$\begin{aligned}
&E(\prod_{k=1}^{4}T_i(t_k))\\
&=E(\prod_{k=1}^{4}Y_l(t_k))\\
&=\frac{1}{16\sqrt{\prod_{k=1}^{4}g(t_k)}}E\Big\{\prod_{k=1}^{4}\Big[p(z_l)K_b(t_k-z_l)-E\big(p(z_l)K_b(t_k-z_l)\big)\Big]\Big\}\\
&=\frac{1}{16\sqrt{\prod_{k=1}^{4}g(t_k)}}E\big(L_1'-L_2'+L_3'-L_4'\big). \qquad (4.34)
\end{aligned}$$

where

$$
\begin{aligned}
L_1' &= E\Big[p^4(z_l)\prod_{k=1}^{4} K_b(t_k - z_l)\Big] \\
L_2' &= \sum_{k=1}^{4}\Big\{E\Big[p^3(z_l)\frac{\prod_{m=1}^{4} K_b(t_m - z_l)}{K_b(t_k - z_l)}\Big]\cdot E\Big[p(z_l)K_b(t_k - z_l)\Big]\Big\} \\
L_3' &= \sum_{\substack{\{k_1,k_2,k_3,k_4\}=\{1,2,3,4\},\\ \{1,3,2,4\},\{1,4,2,3\},\{2,3,1,4\},\\ \{2,4,1,3\},\{3,4,1,2\}}} \Big\{E\Big[p^2(z_l)K_b(t_{k_1} - z_l)K_b(t_{k_2} - z_l)\Big]\cdot E\Big[p(z_l) \\
&\qquad \times K_b(t_{k_3} - z_l)\Big]\cdot E\Big[p(z_l)K_b(t_{k_4} - z_l)\Big]\Big\} \\
L_4' &= \prod_{k=1}^{4} E\Big[p(z_l)K_b(t_k - z_l)\Big].
\end{aligned}
$$

We can obtain the order of $L_1'$, $L_2'$, $L_3'$ and $L_4'$ by deriving general term of them as before. The result is similar. $L_1'$ is of order $\mathcal{O}(\frac{1}{n^4b^3})$, $L_2'$ is of order $\mathcal{O}(\frac{1}{n^4b^2})$, $L_3'$ is of order $\mathcal{O}(\frac{1}{n^4b})$ and $L_4'$ is of order $\mathcal{O}(\frac{1}{n^4})$. Therefore, we have

$$
E(\prod_{k=1}^{4} T_i(t_k)) = \mathcal{O}(\frac{1}{n^4b^3}) \qquad for \qquad n_0 < i \leq n. \tag{4.35}
$$

Combining (4.33) and (4.35), we have for any $i \in \{1, 2, \cdots, n\}$,

$$
E(\prod_{k=1}^{4} T_i(t_k)) = \mathcal{O}(\frac{1}{n^4b^3}). \tag{4.36}
$$

So

$$
\begin{aligned}
Ew_{ij}^4 &= \frac{1}{16}\rho^{4I_{[1,n_0]}(i)+4I_{[1,n_0]}(j)}\int\int\int\int \mathcal{O}(\frac{1}{n^4b^3})\mathcal{O}(\frac{1}{n^4b^3})b^3 ds_4 ds_3 ds_2 dt_1 \\
&= \mathcal{O}(\frac{1}{n^8b^3}).
\end{aligned}
$$

Thus

$$
W_I = \sum_{1\leq i<j\leq n} Ew_{ij}^4 = \binom{n}{2}\mathcal{O}(\frac{1}{n^8b^3}) = \mathcal{O}(\frac{1}{n^6b^3}). \tag{4.37}
$$

Consider the general term in $W_{II}$, we need the following result from combining (4.21) and (4.22),

$$E\Big(T_i(t_1)T_i(t_2)\Big) = \mathcal{O}(\frac{1}{n^2 b}), \tag{4.38}$$

and by using the result in (4.36) and (4.38),

$$\begin{aligned}
&Ew_{ij}^2 w_{ik}^2 \\
&= E\big[(2\int T_i(t)T_j(t)dt)^2(2\int T_i(t)T_k(t)dt)^2\big] \\
&= 16E\int\int\int\int T_i(t_1)T_j(t_1)T_i(t_2)T_j(t_2)T_i(t_3)T_k(t_3)T_i(t_4)T_k(t_4)dt_4dt_3dt_2dt_1 \\
&= 16\int\int\int\int E(\prod_{l=1}^{4} T_i(t_l))E(T_j(t_1)T_j(t_2))E(T_k(t_3)T_k(t_4))dt_4dt_3dt_2dt_1 \\
&= \int\int\int\int \mathcal{O}(\frac{1}{n^4 b^3})\mathcal{O}(\frac{1}{n^2 b})\mathcal{O}(\frac{1}{n^2 b}) \cdot b^3 ds_4ds_3ds_2dt_1 \\
&= \mathcal{O}(\frac{1}{n^8 b^2}).
\end{aligned}$$

Then

$$\begin{aligned}
W_{II} &= \sum_{1\le i<j<k\le n} (Ew_{ij}^2 w_{ik}^2 + Ew_{ji}^2 w_{jk}^2 + Ew_{ki}^2 w_{kj}^2) \\
&= 3\binom{n}{3}\mathcal{O}(\frac{1}{n^8 b^2}) \\
&= \mathcal{O}(\frac{1}{n^5 b^2}).
\end{aligned} \tag{4.39}$$

Consider $W_{III}$, since

$$|2w_{ik}w_{kj}| \le w_{ik}^2 + w_{kj}^2,$$

compare general term of $W_{III}$ with the general term of $W_{II}$,

$$Ew_{ij}^2 w_{ki}w_{kj} + Ew_{ik}^2 w_{ji}w_{jk} + Ew_{kj}^2 w_{ij}w_{ik} \le Ew_{ij}^2 w_{ik}^2 + Ew_{ji}^2 w_{jk}^2 + Ew_{ki}^2 w_{kj}^2.$$

Thus we have $|W_{III}| \le W_{II}$.

Consider the general term in $W_{IV}$, we will use the result in (4.38),

$$\begin{aligned}
& Ew_{ij}w_{ik}w_{lj}w_{lk} \\
&= 16E\Big(\int T_i(t)T_j(t)dt \int T_i(t)T_k(t)dt \int T_l(t)T_j(t)dt \int T_l(t)T_k(t)dt\Big) \\
&= 16 \int\int\int\int ET_i(t_1)T_i(t_2) \cdot ET_j(t_1)T_j(t_3) \cdot ET_k(t_2)T_k(t_4) \\
&\qquad \cdot ET_l(t_3)T_l(t_4)dt_4dt_3dt_2dt_1 \\
&= \int\int\int\int \mathcal{O}(\frac{1}{n^2b}) \cdot \mathcal{O}(\frac{1}{n^2b}) \cdot \mathcal{O}(\frac{1}{n^2b}) \cdot \mathcal{O}(\frac{1}{n^2b}) \cdot b^3 ds_4ds_3ds_2dt_1 \\
&= \mathcal{O}(\frac{1}{n^8b}).
\end{aligned}$$

Thus we derive $W_{IV}$ as

$$\begin{aligned}
W_{IV} &= \sum_{1\le i<j<k<l\le n} (Ew_{ij}w_{ik}w_{lj}w_{lk} + Ew_{ij}w_{il}w_{kj}w_{kl} + Ew_{ik}w_{il}w_{jk}w_{jl}) \\
&= 3\binom{n}{4}\mathcal{O}(\frac{1}{n^8b}) \\
&= \mathcal{O}(\frac{1}{n^4b}). \qquad (4.40)
\end{aligned}$$

Consider the general term in $W_V$, we will use the result in (4.38),

$$\begin{aligned}
Ew_{ij}^2w_{kl}^2 &= E[w_{ij}^2] \cdot E[w_{kl}^2] \\
&= \big(Ew_{ij}^2\big)^2 \\
&= \Big(4E\int\int T_i(t)T_j(t)T_i(s)T_j(s)dsdt\Big)^2 \\
&= \Big(4\int\int ET_i(t)T_i(s) \cdot ET_j(t)T_j(s)dsdt\Big)^2 \\
&= \Big(\int\int \mathcal{O}(\frac{1}{n^2b})\mathcal{O}(\frac{1}{n^2b})bdrdt\Big)^2 \\
&= \mathcal{O}(\frac{1}{n^8b^2}).
\end{aligned}$$

So

$$\begin{aligned} W_V &= \sum_{1\le i<j<k<l\le n} (Ew_{ij}^2 w_{kl}^2 + Ew_{ik}^2 w_{jl}^2 + Ew_{il}^2 w_{jk}^2) \\ &= 3\binom{n}{4}\mathcal{O}(\frac{1}{n^8 b^2}) \\ &= \mathcal{O}(\frac{1}{n^4 b^2}). \end{aligned} \tag{4.41}$$

Combining (4.37),(4.39),(4.40) and (4.41), and the fact that $|W_{III}| \le W_{II}$, we have that $W_I$,$W_{II}$, $W_{III}$ and $W_{IV}$ all are of lower order than $W_V$. So

$$E\Big(\sum_{1\le i<j\le n} w_{ij}\Big)^4 = 6W_V + o(W_V). \tag{4.42}$$

Furthermore,

$$\begin{aligned} \sigma^4(n) &= (EW_a)^2 = (E\sum_{1\le i<j\le n} w_{ij}^2)^2 \\ &= 2W_V + \sum_{1\le i<j\le n} (Ew_{ij}^2)^2 + 2\sum_{1\le i<j\le n}\sum_{\substack{1\le k\le n \\ k\ne i,j}} Ew_{ki}^2 \cdot Ew_{kj}^2. \end{aligned} \tag{4.43}$$

Since

$$\sum_{1\le i<j\le n} (Ew_{ij}^2)^2 = \binom{n}{2}\cdot\mathcal{O}(\frac{1}{n^4 b})\cdot\mathcal{O}(\frac{1}{n^4 b}) = \mathcal{O}(\frac{1}{n^6 b^2}),$$

and

$$\sum_{1\le i<j\le n}\sum_{\substack{1\le k\le n \\ k\ne i,j}} Ew_{ki}^2 \cdot Ew_{kj}^2 = \binom{n}{2}(n-2)\mathcal{O}(\frac{1}{n^4 b})\mathcal{O}(\frac{1}{n^4 b}) = \mathcal{O}(\frac{1}{n^5 b^2}),$$

then in (4.43), the 2nd and 3rd terms are of lower order than $W_V$. So

$$\sigma^4(n) = 2W_V + o(W_V). \tag{4.44}$$

Combining (4.42) and (4.44), we have

$$\sigma(n)^{-4} E(\sum_{1\le i<j\le n} w_{ij})^4 \to 3, \qquad as \qquad n\to\infty.$$

we see that condition (4.25) is satisfied.

Applying Theorem 2.1 in Jong (1987)[17], we have

$$\sigma(n)^{-1} \sum_{1 \le i < j \le n} w_{ij} \xrightarrow{d} \mathbf{N}(0,1)$$

so that

$$\sigma^{-1}(n) \int 2 \sum_{1 \le i < j \le n} T_i(t) T_j(t) dt \xrightarrow{d} \mathbf{N}(0,1)$$

and we have got

$$\int \sum_{k=1}^{n} T_k^2(t) dt \xrightarrow[a.s.]{LLN} \int \frac{\rho}{4nb} \cdot \frac{w(t)}{M(t)} \kappa_{sq} dt.$$

As in (4.15),

$$J_2 = \int \sum_{k=1}^{n} T_k^2(t) dt + \int 2 \sum_{1 \le i < j \le n} T_i(t) T_j(t) dt.$$

By Slutsky's theorem, $J_2$ converges in distribution to normal distribution with mean $\int \frac{\rho}{4nb} \cdot \frac{w(t)}{M(t)} \kappa_{sq} dt$ and variance $\sigma(n)^2 = \frac{\rho^2}{8n^2 b} \kappa_c \int (\frac{w(t)}{M(t)})^2 dt$. Thus Lemma 6 is proved.

$\square$

Next we will prove $J_1, J_3, J_4, J_5, J_6$ are negligible.

**Lemma 7.**

$$E(Q_5(t)) = \mathcal{O}(\frac{1}{n\sqrt{b}}) \qquad E(Q_5(t)^2) = \mathcal{O}(\frac{1}{n^2 b}) \tag{4.45}$$

$$J_1 = \mathcal{O}(b^6) \tag{4.46}$$

$$E(J_3) = \mathcal{O}(\frac{1}{n}) \qquad E(J_3^2) = \mathcal{O}(\frac{1}{n^2}) \tag{4.47}$$

$$E(J_4) = 0 \qquad E(J_4^2) = \mathcal{O}(\frac{b^5}{n}) \tag{4.48}$$

$$E(J_5) = o(\frac{b^3}{\sqrt{n}}) \qquad E(J_5^2) = \mathcal{O}(\frac{b^6}{n}) \tag{4.49}$$

$$E(J_6) = \mathcal{O}(\frac{1}{n}) \qquad E(J_6^2) = \mathcal{O}(\frac{1}{n^2}). \tag{4.50}$$

*Proof.* Combining (4.6), (4.11) in Lemma 5 and Cauchy-Schwarz inequality, we have

$$E\{(q_2(t)-Q_2(t))(\hat{\alpha}-\alpha)\} \leq \sqrt{Var(Q_2(t))}\cdot\sqrt{E(\hat{\alpha}-\alpha)^2} = \mathcal{O}(\frac{1}{n\sqrt{b}}).$$

Similarly we can get $E\{(q_3(t)-Q_3(t))(\hat{\beta}-\beta)\} \leq \mathcal{O}(\frac{1}{n\sqrt{b}})$. Thus $E(Q_5(t)) = E\{(q_2(t)-Q_2(t))(\hat{\alpha}-\alpha)\} + E\{(q_3(t)-Q_3(t))(\hat{\beta}-\beta)\} + E(Q_4(t)) = \mathcal{O}(\frac{1}{n\sqrt{b}})$. Furthermore, $E(Q_5(t)^2) = \mathcal{O}(\frac{1}{n^2 b})$. Hence (4.45) is proved.

From Silverman (1986)[29], we know

$$q_0(t) = E(\tilde{g}(t)) = g(t) + \frac{1}{2}b^2 g^{(2)}(t)\kappa_2 + \mathcal{O}(b^3).$$

Combined with (4.3) in Lemma 5,

$$q_1(t) = E(Q_1(t)) = g(t) + \frac{1}{2}b^2\kappa_2 g^{(2)}(t) + \mathcal{O}(b^3).$$

Note that the integral is in the bounded interval by the sample,

$$J_1 = \int (\frac{q_1(t)-q_0(t)}{2\sqrt{g(t)}})^2 dt = \mathcal{O}(b^6),$$

thus (4.46) is proved.

To prove (4.47), combining (4.6), (4.11) in Lemma 5 and (4.45) and by using Cauchy-Schwarz inequality, we have

$$\begin{aligned}
E(J3) &= E\int (\frac{Q^r(t)}{2\sqrt{g(t)}})^2 dt = \int E(\frac{q_2(t)(\hat{\alpha}-\alpha)+q_3(t)(\hat{\beta}-\beta)-Q_5(t)}{2\sqrt{g(t)}})^2 dt \\
&= \int \frac{q_2^2(t)}{4g(t)}E(\hat{\alpha}-\alpha)^2 dt + \int \frac{q_3^2(t)}{4g(t)}E(\hat{\beta}-\beta)^2 dt + \int \frac{EQ_5^2(t)}{4g(t)} dt \\
&\quad + \int \frac{q_2(t)q_3(t)}{2g(t)}E\{(\hat{\alpha}-\alpha)(\hat{\beta}-\beta)\}dt - \int \frac{q_2(t)}{2g(t)}E\{Q_5(t)(\hat{\alpha}-\alpha)\}dt \\
&\quad - \int \frac{q_3(t)}{2g(t)}E\{Q_5(t)(\hat{\beta}-\beta)\}dt \\
&= \mathcal{O}(\frac{1}{n}).
\end{aligned}$$

Thus $EJ_3^2 = \mathcal{O}(\frac{1}{n^2})$ and (4.47) is proved.

Since $E\{(Q_1(t) - q_1(t)) - (\tilde{g}(t) - q_0(t))\} = 0$, $EJ_4 = 0$. By Cauchy-Schwarz inequality, (4.46) and Lemma 6,

$$E(J_4^2) \le 4\Big(\int (\frac{q_1(t) - q_0(t)}{2\sqrt{g(t)}})^2 dt\Big)\Big(E\int (\frac{(Q_1(t) - q_1(t)) - (\tilde{g}(t) - q_0(t))}{2\sqrt{g(t)}})^2 dt\Big)$$

$$= 4J_1 \cdot EJ_2$$

$$= \mathcal{O}(\frac{b^5}{n}).$$

Thus (4.48) follows.

For (4.49), $E(J_5) = 2\int \frac{q_1(t)-q_0(t)}{g(t)} E(Q^r(t))dt = o(\frac{b^3}{\sqrt{n}})$. Thus $E(J_5^2) = \mathcal{O}(\frac{b^6}{n})$.

For (4.50), since $(Q_1(t) - q_1(t)) - (\tilde{g}(t) - q_0(t)) = \mathcal{O}_p(1)$ and $E(Q^r(t)) = o(\frac{1}{n})$, thus $E(J_6) = \mathcal{O}(\frac{1}{n})$ and $E(J_6^2) = \mathcal{O}(\frac{1}{n^2})$.

□

When $nb^6 \to 0$, from Lemma (6) and Lemma (7) we know that compared with $J_2$, the other component $J_1$, $J_3$, $J_4$, $J_5$ and $J_6$ are negligible. And $J_n^0 = J_1 + J_2 + J_3 + J_4 + J_5 + J_6$, therefore, $J_n^0$ converges to the same distribution as $J_2$.

**Lemma 8.** *When $nb^6 \to 0$,*

$$n\sqrt{b}\Big(J_n^0 - \frac{1}{4nb}\rho\kappa_{sq}\int \frac{w(t)}{M(t)}dt\Big) \xrightarrow{d} \mathbf{N}\Big(0,\ \frac{1}{8}\rho^2\kappa_c \int \Big(\frac{w(t)}{M(t)}\Big)^2 dt\Big).$$

## 4.6 Asymptotic Result for $J_n$

We are using the Epanechnikov kernel, and all the integrals are considered in a bounded interval $[-L, L]$. So

$$\frac{2\sqrt{g(t)} + \sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)} + \sqrt{\tilde{g}(t)})^2(\sqrt{g(t)} + \sqrt{\hat{g}(t)})}(\hat{g}(t) - \tilde{g}(t))^2$$

and

$$\frac{2\sqrt{g(t)}+\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)}}{4g(t)(\sqrt{\hat{g}(t)}+\sqrt{\tilde{g}(t)})^2(\sqrt{g(t)}+\sqrt{\tilde{g}(t)})}(\hat{g}(t)-\tilde{g}(t))^2$$

are both bounded. As in the consistency result in Devroye and Györfi(1985)[7], we have that $\int |g(t)-\hat{g}(t)|dt$ and $\int |g(t)-\tilde{g}(t)|dt$ both converge to 0 with probability one if $b \to 0$ and $nb \to \infty$. Therefore $J_n^1 \to 0$ and $J_n^2 \to 0$ in probability. Since $J_n = J_n^0 + J_n^1 + J_n^2$, by Lemma 8 and Slutsky's theorem, we now finally have our main result.

**Theorem 4.** *If model (1.1) holds and assumptions (A1)-(A3) are satisfied, and if the bandwidth $b$ satisfies $nb^6 \to 0$ as $n \to \infty$, then $J_n = \int (\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\,)^2\, dt$ has limiting distribution given by*

$$n\sqrt{b}\Big(J_n - \frac{1}{4nb}\rho\kappa_{sq}\int \frac{w(t)}{M(t)}dt\Big) \xrightarrow{d} \mathbf{N}\Big(0,\ \frac{1}{8}\rho^2\kappa_c \int \Big(\frac{w(t)}{M(t)}\Big)^2 dt\Big)$$

*where*

$$\kappa_{sq} = \int K(x)^2 dx, \qquad\qquad \kappa_c = \int \Big(\int K(x)K(y-x)dx\Big)^2 dy.$$

*Remark* 2. Since the Epanechnikov kernel is our kernel of choice, $\kappa_{sq}$ and $\kappa_c$ are constant. We will derive their values in Appendix B.

*Remark* 3. Let

$$m = \frac{1}{4nb}\rho\kappa_{sq}\int \frac{w(t)}{M(t)}dt$$

and

$$v = \frac{1}{8n^2b}\rho^2\kappa_c\int (\frac{w(t)}{M(t)})^2 dt.$$

Then from Theorem 4, we know that under model (1.1), the limiting distribution of $J_n$ is normal with mean $m$ and variance $v$. The mean $m$ can be estimated by

$$\hat{m} = \frac{1}{4b}\rho\kappa_{sq}\sum_{k=1}^{n}\frac{\hat{w}(t_k)}{\hat{g}(t_k)}\hat{p}(t_k)^2 \tag{4.51}$$

and the variance $v$ can be estimated by

$$\hat{v} = \frac{1}{8b}\rho^2\kappa_c\sum_{k=1}^{n}\frac{\hat{w}(t_k)^2}{\hat{g}(t_k)}\hat{p}(t_k)^3. \tag{4.52}$$

*Remark* 4. Combining Theorem 4 and Remark 3 we have

$$\frac{1}{\sqrt{\hat{v}}}(J_n - \hat{m}) \to \mathbf{N}(0,1) \qquad as\ n \to \infty \tag{4.53}$$

Let the significance level be $\alpha_S$ and let $z_{\alpha_S}$ denote the point having probability $\alpha_S$ to the right of it in the standard normal distribution. Then we can use the test which rejects model (1.1) if

$$J_n \geq \hat{m} + \sqrt{\hat{v}}z_{\alpha_S}. \tag{4.54}$$

*Remark* 5. We can derive the $p$-value of the goodness-of-fit test by using (4.53). Let $J_n^{ob}$ be the observed statistic which is

$$J_n^{ob} = \int_{-L}^{L}\left(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\right)^2 dt$$

Since the Epanechnikov kernel is our kernel of choice, $\hat{g}(t)$ and $\tilde{g}(t)$ both are 0 when $|t| > L$. Thus the integral range for $J_n^{ob}$ is defined in $[-L, L]$. Then from (4.53), the $p$-value can be approximated by

$$1 - \Phi\big(\frac{1}{\sqrt{\hat{v}}}(J_n^{ob} - \hat{m})\big). \tag{4.55}$$

*Remark* 6 (Bandwidth Selection). When we deal with nonparametric and semiparametric problems, the choice of an appropriate bandwidth for the kernel estimate of the underlying density is always crucial and important. In our Theorem 4 the bandwidth $b$ is assumed to satisfy

$$\begin{aligned} & b \to 0 \\ & nb \to \infty \\ & nb^6 \to 0 \qquad \qquad as \quad n \to \infty \end{aligned}$$

Thus we have $\frac{1}{n} << b << \frac{1}{n^{1/6}}$. That is, we need the bandwidth $b$ to be between $\frac{1}{n}$ and $\frac{1}{n^{1/6}}$ and not close to either limit. For example the range of the value of the bandwidth $b$ could be $\frac{5}{n} \leq b \leq (\frac{0.2}{n})^{1/6}$.

# Chapter 5

# Simulation Studies for $J_n$

## 5.1 Overview

In this chapter, we will present Monte Carlo simulations to evaluate the performance of our goodness-of-fit test for the semiparametric density ratio model. The test statistic $J_n$ is defined in (4.12) and its asymptotic properties are analyzed in Chapter 4. The reference and distortion samples will be computer-generated randomly. We restrict attention to normal, gamma, and lognormal samples.

First, we will derive our asymptotic approximation of the distribution of the test statistic as in Theorem 4, and get the estimated distribution of the test statistic from generated samples. The theoretical distribution and real-data distribution are investigated under various bandwidth selections. According to Remark 6, the bandwidth $b$ should be selected in the range $\frac{5}{n} \le b \le (\frac{0.2}{n})^{1/6}$. We apply an equally spaced grid search to derive the optimal bandwidth.

Furthermore, using the method in Remark 5, we calculate the $p$-value as (4.55) under two cases. One is the correct selection of the tilt function. In this case model (1.1) is correct and the null hypothesis should be accepted (i.e., the model is correctly specified). In the second case, we intentionally misspecified the tilt function as in Fokianos and Kaimi (2006)[11], where the null hypothesis should be rejected (i.e., the model is misspecified). As a comparison, we obtain the $p$-values of the test corresponding to the statistic $J_n^c$ which was defined in Cheng and Chu (2004)[6].

## 5.2 Behavior of $J_n$

In this section, we generate the reference and distortion samples from normal, gamma and lognormal distributions. The asymptotic mean and variance of $J_n$ are calculated following Theorem 4, and also empirically from the samples. In all the simulations, the total sample size $n = n_0 + n_1$ is 1000, where $n_0$ is the size of the reference sample and $n_1$ is the size of the distortion sample. According to Remark 6, the range of the bandwidth will be $[\frac{5}{n}, (\frac{0.2}{n})^{1/6}] = [0.005, 0.2418]$.

Consider the kth Monte-Carlo repetition, let $\hat{m}_k$ be the theoretical mean of $J_n$ as in Theorem 4. Let $\tilde{m}_k$ be the value of the test statistic $J_n$ calculated from the kth computer-generated sample. Since the Epanechnikov kernel is our kernel of choice, the integral range of $J_n$ is only considered in $[-L, L]$, i.e. $J_n = \int_{-L}^{L} \left(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\right)^2 dt$. Thus by many Monte-Carlo repetitions, the optimal bandwidth can be obtained by minimizing $|\sum \hat{m}_k - \sum \tilde{m}_k|$ with respect to the bandwidth, over equally spaced grid points in the range $[0.005, 0.2418]$.

### 5.2.1 Normal (0,1) and Normal (0,2)

We generate the reference sample $U$ from $N(0, 1)$, and the distortion sample $Z$ from $N(0, 2)$. According to model (1.1) and Example 1, $h(x) = x^2$. The sample sizes $\{n_0, n_1\}$ are first taken as $\{400, 600\}$ and then $\{600, 400\}$. For each pair combination, there are 200 Monte-Carlo runs. Table 5.1 gives the simulation results. Note that in each Monte-Carlo repetition, $\hat{m}_k$ is the theoretical mean of the $J_n$. Following Theorem 4, $\hat{m}_k$ is obtained as in (4.51), and $\tilde{m}_k$ is the value of $J_n$, calculated from the definition $J_n = \int_{-L}^{L} \left(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\right)^2 dt$ by using the simulated samples. We compare the mean of $\{\hat{m}_k\}$ with the mean of $\{\tilde{m}_k\}$ under multiple bandwidth selections and get the relative percentage of the difference of the means of $\hat{m}_k$ and $\tilde{m}_k$ in the last column. A large difference means the approximated distribution of $J_n$ as in Theorem 4 is not accurate because the bandwidth is not chosen properly. In the table 5.1 optimal bandwidth is labeled by '$\star$'. Figure 5.1 shows the box plot for comparison of $\hat{m}_k$ and $\tilde{m}_k$ when bandwidth is optimal and when it is not optimal.

### 5.2.2 Gamma Distribution

In this section we generate the reference sample $U$ from $Gamma(3, 1)$, and the distortion sample $Z$ from $Gamma(1, 1)$. According to model (1.1) and Example 2, $h(x) = \log(x)$. The sample sizes $\{n_0, n_1\}$ are $\{400, 600\}$. The number of Monte-Carlo repetitions is 200. Table 5.2 gives the simulation results.

Furthermore, we generate the reference sample $U$ from $Gamma(1, 1)$, and the distortion sample $Z$ from $Gamma(1, 0.2)$. According to model (1.1) and Example 2, the correct tilt is $h(x) = x$. The sample sizes $\{n_0, n_1\}$ are $\{500, 500\}$. The number of Monte-Carlo repetitions is 200. Table 5.3 gives the simulation result.

Table 5.1: $N(0,1)$ and $N(0,2)$. $U \sim N(0,1)$, $Z \sim N(0,2)$, $\star$ optimal bandwidth for the statistics $J_n$.

| $U$ | $Z$ | band- | $\{\hat{m}_k\}$ mean of | $\{\tilde{m}_k\}$ mean of | relative % of |
|---|---|---|---|---|---|
| sample size | | -width | theoretical | empirical | the difference |
| 400 | 600 | 0.005 | 0.1884818 | 0.2875685 | 34.46% |
| | | 0.0178 | 0.1006727 | 0.1030558 | 2.31% |
| | | 0.02$\star$ | 0.09371565 | 0.09341618 | 0.32% |
| | | 0.1 | 0.02942333 | 0.02125169 | 38.45% |
| | | 0.25 | 0.01337576 | 0.00866427 | 54.38% |
| 600 | 400 | 0.005 | 0.08372413 | 0.1460077 | 42.66% |
| | | 0.0178 | 0.04687305 | 0.05431368 | 13.70% |
| | | 0.03$\star$ | 0.0346543 | 0.03458529 | 0.20% |
| | | 0.1 | 0.0153845 | 0.01171947 | 31.27% |
| | | 0.25 | 0.0073784 | 0.004849092 | 52.16% |

Table 5.2: Gamma (3,1) and Gamma (1,1). $U \sim$ Gamma (3,1), $Z \sim$ Gamma (1,1), $\star$ optimal bandwidth for the statistics $J_n$.

| $U$ | $Z$ | band- | $\{\hat{m}_k\}$ mean of | $\{\tilde{m}_k\}$ mean of | relative % of |
|---|---|---|---|---|---|
| sample size | | -width | theoretical | empirical | the difference |
| 400 | 600 | 0.005 | 0.1184832 | 0.2083053 | 43.12% |
| | | 0.0178 | 0.04331359 | 0.06594018 | 34.31% |
| | | 0.1 | 0.008454803 | 0.009272672 | 8.82% |
| | | 0.23$\star$ | 0.003683005 | 0.003684826 | 0.05% |
| | | 0.25 | 0.003385677 | 0.003349908 | 1.07% |

Table 5.3: Gamma (1,1) and Gamma (1,0.2). $U \sim$ Gamma (1,1), $Z \sim$ Gamma (1,0.2), $\star$ optimal bandwidth for the statistics $J_n$.

| $U$ | $Z$ | band- | $\{\hat{m}_k\}$ mean of | $\{\tilde{m}_k\}$ mean of | relative % of |
|---|---|---|---|---|---|
| sample size | | -width | theoretical | empirical | the difference |
| 500 | 500 | 0.005$\star$ | 0.1434049 | 0.1471724 | 2.56% |
| | | 0.01 | 0.1185871 | 0.09683343 | 22.47% |
| | | 0.0178 | 0.09668758 | 0.06669246 | 44.98% |
| | | 0.1 | 0.03964973 | 0.017706 | 123.93% |
| | | 0.25 | 0.02122221 | 0.007987027 | 165.71% |

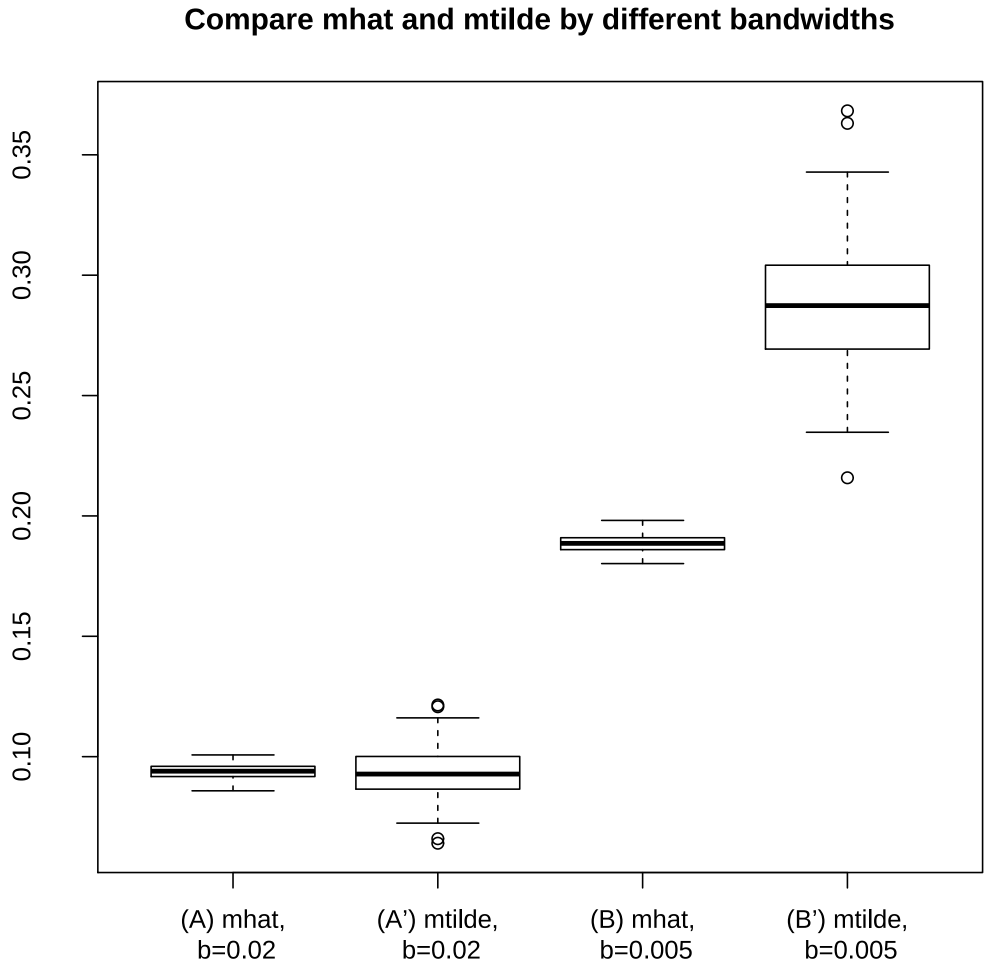


Figure 5.1: *Boxplot for* $\hat{m}_k$ *and* $\tilde{m}_k$. *The left mhat (A) and mtilde (A') are under optimal bandwidth=0.02. The bandwidth for the right mhat (B) and mtilde (B') is 0.005.*

### 5.2.3 Lognormal Distribution

In this section we generate the reference sample $U$ from $Lognormal(0, 0.5)$, and the distortion sample $Z$ from $Lognormal(0, 0.7)$. According to model (1.1) and Example 3, $h(x) = (\log(x))^2$. The sample sizes $\{n_0, n_1\}$ are $\{500, 500\}$. The number of Monte-Carlo repetitions is 200. Table 5.4 gives the simulation results.

## 5.3 Goodness-of-Fit Test

In section (5.2), we studied the behavior of $J_n$. Here we apply it. Following the method in Remark 5, we can use (4.55) to approximate the $p$-value corresponding to the test statistic $J_n$, and we reject $H_0$ if the $p$-value is small. We still generate the samples from normal, gamma and lognormal distributions. We calculate $p$-values not only when the tilt function is correctly specified, but also misspecified, as in Fokianos and Kaimi (2006)[11]. The $p$-values corresponding to the test statistic $J_n^C = \int_{-L}^{L} \left(\hat{g}(t) - \tilde{g}(t)\right)^2 dt$, which was proposed by Cheng and Chu (2004)[6], are calculated for comparison. As before, we generate reference and distortion samples from normal, gamma and lognormal distributions.

Table 5.4: Lognormal (0,0.5) and Lognormal (0,0.7). $U \sim$ Lognormal (0,0.5), $Z \sim$ Lognormal (0,0.7), $\star$ optimal bandwidth for the statistics $J_n$.

| $U$ | $Z$ | band- | $\{\hat{m}_k\}$ mean of | $\{\tilde{m}_k\}$ mean of | relative % of |
|---|---|---|---|---|---|
| sample size | | -width | theoretical | empirical | the difference |
| 500 | 500 | 0.005 | 0.07321884 | 0.1280294 | 42.81% |
| | | 0.0178 | 0.03339353 | 0.04640805 | 28.04% |
| | | 0.1 | 0.01001211 | 0.01136205 | 11.88% |
| | | 0.235$\star$ | 0.005325392 | 0.005319663 | 0.11% |
| | | 0.25 | 0.005084966 | 0.00502595 | 1.17% |

### 5.3.1 Normal (0,1) and Normal (0,2)

We generate the reference sample $U$ from $N(0,1)$, and the distortion sample $Z$ from $N(0,2)$. According to model (1.1) and Example 1, correctly specified tilt function is $h(x) = x^2$. We intentionally misspecify the tilt function by $h(x) = x$. The sample sizes $\{n_0, n_1\}$ are first taken as $\{400, 600\}$ and then $\{600, 400\}$. For each pair combination, there are 200 Monte-Carlo runs. $p$-values for both statistics $J_n^C$ and $J_n$ are calculated under various bandwidths. We pick the median of the $p$-value in 200 Monte-Carlo Repetitions. The maximum of the $p$-value in 200 Monte-Carlo repetitions is recorded when the tilt function is misspecified. Table 5.5 gives the simulation results.

### 5.3.2 Gamma Distribution

We generate the reference sample $U$ from $Gamma(3,1)$, and the distortion sample $Z$ from $Gamma(1,1)$. According to model (1.1) and Example 1, correctly specified till function is $h(x) = \log x$. We intentionally misspecify the tilt function by $h(x) = x$. The sample sizes $\{n_0, n_1\}$ are $\{400, 600\}$. $p$-values for both statistics $J_n^C$ and $J_n$ are calculated under various bandwidths. We pick the median of the $p$-value in 200 Monte-Carlo Repetitions. The maximum of the $p$-value in 200 Monte-Carlo repetitions is recorded when the tilt function is misspecified. Table 5.6 gives the simulation results. Furthermore, we generate the reference sample $U$ from $Gamma(1,1)$, and the distortion sample $Z$ from $Gamma(1,0.2)$. According to model (1.1) and Example 1, correctly specified till function is $h(x) = x$. We

Table 5.5: $N(0,1)$ and $N(0,2)$. $U \sim N(0,1)$, $Z \sim N(0,2)$, $\star$ optimal bandwidth for the statistics $J_n$, † optimal bandwidth for the statistics $J_n^C$

| $U$ | $Z$ | band- | $p$-value, column $J_n^C$ and $J_n$ are median from simulations | | | | | |
|---|---|---|---|---|---|---|---|---|
| sample | | -width | specified $h(x)=x^2$ | | misspecified $h(x)=x$ | | | |
| size | | | $J_n^C$ | $J_n$ | $J_n^C$ | max $J_n^C$ | $J_n$ | max $J_n$ |
| 400 | 600 | 0.005 | 0.5008695 | 0 | 0.010536 | 0.999999 | 0 | 0 |
| | | 0.0115† | 0.5405066 | 0.1401269 | 9.68E-06 | 0.242335 | 0 | 0 |
| | | 0.0178 | 0.5267068 | 0.3488811 | 1.96E-06 | 0.549676 | 0 | 0 |
| | | 0.02$\star$ | 0.5164572 | 0.5069916 | 2.91E-07 | 0.381980 | 0 | 0 |
| | | 0.1 | 0.6088644 | 0.9618217 | 0 | 1.43E-10 | 0 | 0 |
| | | 0.25 | 0.6810998 | 0.9316603 | 0 | 0 | 0 | 0 |
| 600 | 400 | 0.005 | 0.6725442 | 0 | 0.081080 | 0.997779 | 0 | 0 |
| | | 0.0178 | 0.561414 | 0.0510693 | 7.80E-06 | 0.208925 | 0 | 0 |
| | | 0.03$\star$ | 0.6323972 | 0.52481 | 3.75E-09 | 0.007031 | 0 | 0 |
| | | 0.1 | 0.6600704 | 0.9294418 | 0 | 3.55E-07 | 0 | 0 |
| | | 0.25† | 0.698934 | 0.9302384 | 0 | 0 | 0 | 0 |

intentionally misspecify the tilt function by $h(x) = \log x$. The sample sizes $\{n_0, n_1\}$ are $\{500, 500\}$. $p$-values for both statistics $J_n^C$ and $J_n$ are calculated under various bandwidths. We pick the median of the $p$-value in 200 Monte-Carlo Repetitions. The maximum of the $p$-value in 200 Monte-Carlo repetitions is recorded when the tilt function is misspecified. Table 5.7 gives the simulation results.

Table 5.6: Gamma (3,1) and Gamma (1,1). $U \sim$ Gamma (3,1), $Z \sim$ Gamma (1,1), $\star$ optimal bandwidth for the statistics $J_n$, $\dagger$ optimal bandwidth for the statistics $J_n^C$.

| $U$ | $Z$ | band- | $p$-value, column $J_n^C$ and $J_n$ are median from simulations | | | | | |
|---|---|---|---|---|---|---|---|---|
| sample | | -width | specified $h(x) = \log(x)$ | | misspecified $h(x) = x$ | | | |
| size | | | $J_n^C$ | $J_n$ | $J_n^C$ | max $J_n^C$ | $J_n$ | max $J_n$ |
| 400 | 600 | 0.005 | 0.6442556 | 0 | 0.329217 | 0.99172 | 0 | 0 |
| | | 0.01† | 0.5793755 | 1.33E-15 | 0.121219 | 0.98821 | 0 | 3.10E-10 |
| | | 0.0178 | 0.5693371 | 2.70E-06 | 0.033288 | 0.85205 | 0 | 0 |
| | | 0.1 | 0.6416833 | 0.404165 | 1.10E-06 | 0.36665 | 0 | 0.000252 |
| | | 0.23⋆ | 0.7039756 | 0.565106 | 2.11E-11 | 0.04825 | 0 | 6.37E-09 |
| | | 0.25 | 0.7058387 | 0.589399 | 9.13E-12 | 0.04042 | 0 | 2.71E-09 |

### 5.3.3 Lognormal Distribution

We generate the reference sample $U$ from $Lognormal(0, 0.5)$, and the distortion sample $Z$ from $Lognormal(0, 0.7)$. According to model (1.1) and Example 1, correctly specified till function is $h(x) = (\log x)^2$. We intentionally misspecify the

Table 5.7: Gamma (1,1) and Gamma (1,0.2). $U \sim$ Gamma (1,1), $Z \sim$ Gamma (1,0.2), $\star$ optimal bandwidth for the statistics $J_n$, $\dagger$ optimal bandwidth for the statistics $J_n^C$.

| $U$ | $Z$ | | $p$-value, column $J_n^C$ and $J_n$ are median from simulations | | | | | |
|---|---|---|---|---|---|---|---|---|
| sample size | | band- -width | specified $h(x) = \log(x)$ | | misspecified $h(x) = x$ | | | |
| | | | $J_n^C$ | $J_n$ | $J_n^C$ | max $J_n^C$ | $J_n$ | max $J_n$ |
| 500 | 500 | 0.005$\star$ | 0.81777 | 0.37423 | 0.7000 | 1 | 2.27E-09 | 0.99998 |
| | | 0.01 | 0.69705 | 0.99831 | 0.21356 | 1 | 0.00038 | 0.99971 |
| | | 0.013$\dagger$ | 0.70421 | 0.99948 | 0.09164 | 1 | 0.00083 | 0.99410 |
| | | 0.0178 | 0.71502 | 0.99996 | 0.02159 | 0.99992 | 0.00251 | 0.83783 |
| | | 0.1 | 0.62303 | 0.99999 | 1.27E-10 | 0.23726 | 3.50E-07 | 0.74024 |
| | | 0.25 | 0.68406 | 0.99984 | 0 | 0.00946 | 3.59E-14 | 0.05056 |

tilt function by $h(x) = \log x$. The sample sizes $\{n_0, n_1\}$ are $\{500, 500\}$. $p$-values for both statistics $J_n^C$ and $J_n$ are calculated under various bandwidths. We pick the median of the $p$-value in 200 Monte-Carlo Repetitions. The maximum of the $p$-value in 200 Monte-Carlo repetitions is recorded when the tilt function is misspecified. Table 5.8 gives the simulation results.

Table 5.8: Lognormal (0,0.5) and Lognormal (0,0.7). $U \sim$ Lognormal(0,0.5), $Z \sim$ Lognormal(0,0.7), $\star$ optimal bandwidth for the statistics $J_n$, $\dagger$ optimal bandwidth for the statistics $J_n^C$.

| $U$ | $Z$ | band- | $p$-value, column $J_n^C$ and $J_n$ are median from simulations | | | | | |
|---|---|---|---|---|---|---|---|---|
| sample | | -width | specified $h(x) = \log(x)$ | | misspecified $h(x) = x$ | | | |
| size | | | $J_n^C$ | $J_n$ | $J_n^C$ | max $J_n^C$ | $J_n$ | max $J_n$ |
| 500 | 500 | 0.005 | 0.4875377 | 0 | 0.1198135 | 0.975032 | 0 | 0 |
| | | 0.011† | 0.500015 | 3.64E-07 | 0.0456381 | 0.75716 | 0 | 2.89E-15 |
| | | 0.0178 | 0.4991361 | 0.0008704 | 0.0116544 | 0.724075 | 0 | 3.08E-13 |
| | | 0.1 | 0.6598331 | 0.3024718 | 2.07E-07 | 0.228446 | 0 | 2.22E-14 |
| | | 0.235⋆ | 0.7710084 | 0.5480933 | 7.25E-11 | 0.077927 | 0 | 2.78E-15 |
| | | 0.25 | 0.7725608 | 0.5572037 | 6.59E-11 | 0.068809 | 0 | 8.22E-15 |

## 5.4 Comparison with $J_n^C$

Since we can derive the limiting distribution of $J_n$ and $J_n^C$, bootstrap is not needed in the goodness-of-fit test using these test statistics. When the tilt function is misspecified, $J_n$ rejects $H_0$ strongly in most of the cases in our simulations, but on the other hand $J_n^C$ may give relatively large $p$-values resulting in acceptance of the null hypothesis in many misspecified situations. When the tilt function is correctly specified, the $p$-values of the test corresponding to $J_n$ are large when the bandwidth is chosen properly whereas $J_n^C$ accepts model (1.1) all the time as it should in specified cases. Thus $J_n$ performs as well as $J_n^C$ when the tilt function is correctly specified and improves upon $J_n^C$ when the tilt function is misspecified.

# Chapter 6

# Application to Radar Data

In this chapter, we apply our test statistics $I_n$ proposed in Chapter 3 and $J_n$ proposed in Chapter 4 in a two-sample radar problem, and compare them with the test statistics $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $J_n^C$.

## 6.1 Description of the Radar Data

During NASA's Tropical Rainfall Measuring Mission (TRMM) Kwajalein Experiment (KWAJEX), held during Jul.15-Sep.12,1999 in the Republic of the Marshall Islands, a C-band radar was deployed aboard NOAA ship Ronald H.Brown (RHB) and an S-band KPOL radar was deployed on Kwajalein Island at the southern end of the Kwajalein Atoll. Experimental radar reflectivity data were obtained from these two radars. Kedem *et al.* 2004[24] gives more details about the data. The data were collected in pairs referring to the radar as 'Brown' and 'Kwajalein'.

We randomly sample the data collected from the 'Brown' radar to produce random reference samples $U$, and randomly sample the data collected from the 'Kwajalein' radar to produce random distortion samples $Z$. If these two radars or their algorithms produce equidistributed reflectivity data, the null hypothesis model (1.1) should be accepted in the goodness-of-fit test by fusing $U$ and $Z$. For comparison, random reference samples $U$ and distortion samples $Z$ are sampled from the same radar to apply the test statistics. For the reason that radar data are always assumed to be distributed as lognormal or gamma, following the results from Example 2 and Example 3, we choose $h(x) = x$ and $h(x) = \log(x)$ as tilt functions.

## 6.2 Dataset From Two Different Radars

$U$ is randomly sampled from 'Brown' as the reference sample and $Z$ is randomly sampled from 'Kwajalein' as the distortion sample. The sample sizes $(n_0, n_1)$=$(1500, 1500)$. The combined sample $T = \{U, Z\}$ has size $n = 3000$. The test statistics

$$\begin{aligned}
\Delta_n &= \sup_t \sqrt{n}|\hat{G}(t) - \tilde{G}(t)| \\
I_n^B &= n \int_{-L}^{L} \big(\hat{g}(t) - \tilde{g}(t)\big)^2 dt \\
I_n^{Bb} &= nb \int_{-L}^{L} \Big(\hat{g}(t) - \tilde{g}(t)\Big)^2 dt \\
I_n &= nb \int_{-L}^{L} \Big(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\Big)^2 dt \\
J_n^C &= \int_{-L}^{L} \Big(\hat{g}(t) - \tilde{g}(t)\Big)^2 dt \\
J_n &= \int_{-L}^{L} \Big(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\Big)^2 dt
\end{aligned}$$

are used in the goodness-of-fit test of the null model (1.1), with bandwidth $b$. Following the analysis in Chapter 3 and Chapter 4, a Gaussian kernel is used for $I_n^B$, $I_n^{Bb}$ and $I_n$, and the Epanechnikov kernel is used for $J_n^C$ and $J_n$. As noted in Remark 6, the bandwidth needs to satisfy $\frac{5}{n} \leq b \leq (\frac{0.2}{n})^{1/6}$. So, the bandwidth is in the range $[0.0017, 0.201]$. We derive $p$-values of the tests corresponding to each statistic under various values of bandwidth which are $0.02, 0.05, 0.1, 0.2, 0.25$.

Table 6.1 gives the $p$-values of the goodness-of-tests corresponding to all six test statistics with various values of bandwidth. The tilt functions $h(x) = x$ and $h(x) = \log(x)$ are applied separately. The $p$-values for $J_n$ and $J_n^C$ are derived by (4.55) as in Remark 5. The $p$-values for $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $I_n$ are derived by bootstrap procedures discussed in Chapter 3. All six statistics suggest strong rejection of model (1.1) which means that these two radars operate very differently.

## 6.3 From the Same Radar

In section (6.2), all test statistics indicate rejection of model (1.1) when the reference sample is from 'Brown' and the distortion sample is from 'Kwajalein'. In this section, the reference and distortion samples are from the same radar. The sample sizes are as before, $(n_0, n_1)$=$(1500, 1500)$. The reference $U$ and distortion $Z$ are both randomly sampled from 'Brown' first and then from 'Kwajalein'. As before, a Gaussian kernel is used for $I_n^B$, $I_n^{Bb}$ and $I_n$, and Epanechnikov kernel is used for $J_n^C$ and $J_n$. The bandwidth used is in the range $[0.0017, 0.201]$. The optimal bandwidth for $J_n$ is 0.19. Since $\Delta_n$ does not depend on a bandwidth, and the bandwidth for $I_n^B$ is fixed at $b = 1$, we see that the $p$-values of the goodness-of-fit tests corresponding to $\Delta_n$ and $I_n^B$ do not change in Table 6.2 and Table 6.3. The tilt functions $h(x) = x$ and $h(x) = \log(x)$ are applied separately both in Tables 6.2 and 6.3. Very similar results are seen from these two tables. The $p$-values for $\Delta_n$, $I_n^B$, $I_n^{Bb}$, $I_n$ and $J_n^C$ lead to acceptance of model (1.1), and so does $J_n$ when the bandwidth is chosen properly. The results tell us the fact that samples $U$ and $Z$ are generated by the same algorithm.

Table 6.1: $U$ is from 'Brown' as reference and $Z$ is from 'Kwajalein' as distortion. Sample size $(n_0, n_1)$=(1500, 1500). Tile functions are $h(x) = x$ and $h(x) = \log(x)$.

| tilt function | band-width | $p$-value | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\Delta_n$ | $I_n^B$ | $I_n^{Bb}$ | $I_n$ | $J_n^C$ | $J_n$ |
| $x$ | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.1 | 0 | 0 | 0 | 0 | 1.35E-10 | 0 |
| | 0.05 | 0 | 0 | 0 | 0 | 2.73E-7 | 0 |
| | 0.02 | 0 | 0 | 0 | 0 | 0.00112 | 0 |
| $\log(x)$ | 0.25 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.2 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 0.1 | 0 | 0 | 0 | 0 | 2.48E-12 | 0 |
| | 0.05 | 0 | 0 | 0 | 0 | 2.62E-8 | 0 |
| | 0.02 | 0 | 0 | 0 | 0.002 | 0.00041 | 0 |

Table 6.2: $U$ is from 'Brown' as reference and $Z$ is from 'Brown' as distortion. Sample size $(n_0, n_1)$=$(1500, 1500)$. Tile functions are $h(x) = x$ and $h(x) = \log(x)$.

| tilt function | band-width | $p$-value | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\Delta_n$ | $I_n^B$ | $I_n^{Bb}$ | $I_n$ | $J_n^C$ | $J_n$ |
| $x$ | 0.25 | 0.57 | 0.7 | 1 | 0.994 | 0.9981411 | 0.9703743 |
| | 0.2 | 0.57 | 0.7 | 1 | 0.998 | 0.999185 | 0.9565418 |
| | 0.1 | 0.57 | 0.7 | 1 | 1 | 0.9999877 | 0.6973797 |
| | 0.05 | 0.57 | 0.7 | 1 | 1 | 1 | 0.010993 |
| | 0.02 | 0.57 | 0.7 | 1 | 1 | 1 | 0 |
| $\log(x)$ | 0.25 | 0.462 | 0.76 | 0.998 | 0.998 | 0.9983481 | 0.9853257 |
| | 0.2 | 0.462 | 0.76 | 0.998 | 0.998 | 0.9992846 | 0.9779273 |
| | 0.1 | 0.462 | 0.76 | 1 | 1 | 0.999989 | 0.8064212 |
| | 0.05 | 0.462 | 0.76 | 1 | 1 | 1 | 0.02748843 |
| | 0.02 | 0.462 | 0.76 | 1 | 1 | 1 | 0 |

Table 6.3: $U$ is from 'Kwajalein' as reference and $Z$ is from 'Kwajalein' as distortion. Sample size $(n_0, n_1)$=$(1500, 1500)$. Tile functions are $h(x) = x$ and $h(x) = \log(x)$.

| tilt function | band-width | $p$-value | | | | | |
|---|---|---|---|---|---|---|---|
| | | $\Delta_n$ | $I_n^B$ | $I_n^{Bb}$ | $I_n$ | $J_n^C$ | $J_n$ |
| $x$ | 0.25 | 0.318 | 0.768 | 1 | 0.992 | 0.9988772 | 0.9926679 |
| | 0.2 | 0.318 | 0.768 | 1 | 0.998 | 0.9995837 | 0.9904045 |
| | 0.1 | 0.318 | 0.768 | 1 | 1 | 0.9999903 | 0.6773479 |
| | 0.05 | 0.318 | 0.768 | 1 | 1 | 1 | 0.00017857 |
| | 0.02 | 0.318 | 0.768 | 1 | 1 | 1 | 0 |
| $\log(x)$ | 0.25 | 0.254 | 0.87 | 1 | 0.998 | 0.9990609 | 0.9972642 |
| | 0.2 | 0.254 | 0.87 | 1 | 1 | 0.999647 | 0.9963206 |
| | 0.1 | 0.254 | 0.87 | 1 | 1 | 0.9999915 | 0.7975165 |
| | 0.05 | 0.254 | 0.87 | 1 | 1 | 1 | 0.000779084 |
| | 0.02 | 0.254 | 0.87 | 1 | 1 | 1 | 0 |

# Chapter 7

# Conclusion

## 7.1 Overview

In this dissertation, we proposed two new goodness-of-fit test statistics $I_n$ as in (3.5) and $J_n$ as in (4.12) for DRM. Model (1.1) is the null model. Goodness-of-fit tests are needed to justify or reject the assumed model. $I_n$ is a modification of $I_n^B$ which was proposed by Bondell (2007)[4], and $J_n$ is a modification of $J_n^B$ which was proposed by Cheng and Chu (2004)[6]. In this chapter we summarize the advantages of our test statistics, $I_n$ and $J_n$, our contribution of this dissertation. The distributions of the test statistics $\Delta_n$, $I_n^B$, $I_n^{Bb}$ and $J_n^C$ are investigated to illustrate the advantage of $I_n$ and $J_n$, namely, our test statistics appear to be more symmetric, and distinguished well between correctly specified and misspecified cases. Recall from (1.2), (1.5), (3.21), (3.5), (1.6) and (4.12),

$$\Delta_n = \sup_t \sqrt{n}|\hat{G}(t) - \tilde{G}(t)|$$
$$I_n^B = n\int_{-L}^{L}\big(\hat{g}(t) - \tilde{g}(t)\big)^2 dt$$
$$I_n^{Bb} = nb\int_{-L}^{L}\Big(\hat{g}(t) - \tilde{g}(t)\Big)^2 dt$$
$$I_n = nb\int_{-L}^{L}\Big(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\Big)^2 dt$$
$$J_n^C = \int_{-L}^{L}\big(\hat{g}(t) - \tilde{g}(t)\big)^2 dt$$

and

$$J_n = \int_{-L}^{L} \left(\sqrt{\hat{g}(t)} - \sqrt{\tilde{g}(t)}\right)^2 dt.$$

## 7.2 Simulation

In this section, we derive the distributions of $\Delta_n$, $I_n^B$, $I_n^{Bb}$, $I_n$, $J_n^C$ and $J_n$ by samples from computer-generated data when the tilt functions are both correctly specified and misspecified.

### 7.2.1 Correctly Specified

Let $X \sim N(0, 1)$ with size 5000 and $Y \sim N(0, 2)$ with size 5000 be our populations. The reference $U$ and distortion $Z$ are sampled from $X$ and $Y$ respectively with sample sizes $(n_0, n_1) = (500, 500)$. As Example 1, the correctly specified tilt function is $h(x) = x^2$. Following the DRM in Chapter 2, we can obtain $\hat{\alpha}, \hat{\beta}$ and $\hat{G}$ from the fused sample $T = \{U, Z\}$, and $\tilde{G}$ from the reference sample $U$ only. Thereafter, $\hat{g}$ and $\tilde{g}$ are derived by following Lemma 2. Note that $I_n^B$, $I_n^{Bb}$ and $I_n$ use the Gaussian kernel $K(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$, and $J_n^C$ and $J_n$ use the Epanechnikov kernel as in (4.13). Bandwidth $b = 0.02$ is used for $I_n^{Bb}$, $I_n$, $J_n^C$ and $J_n$, but $I_n^B$ uses a fixed bandwidth $b = 1$. Therefore we can obtain the numerical values of $\Delta_n$, $I_n^B$, $I_n^{Bb}$, $I_n$, $J_n^C$ and $J_n$ from the reference samples $U$ and the distortion samples $Z$. We repeat this procedure 1000 times to approximate the distributions of $\Delta_n$, $I_n^B$, $I_n^{Bb}$, $I_n$, $J_n^C$ and $J_n$.

Figure 7.1 gives the histograms of the simulated distributions of $\Delta_n$, $I_n^B$, $I_n^{Bb}$,

$I_n$, $J_n^C$ and $J_n$. The histograms of $\Delta_n$ and $I_n^B$ have very long right tails. Actually we do not know their distribution functions. The right tails corresponding to $I_n^{Bb}$ and $J_n^C$ are shorter as compared with $\Delta_n$ and $I_n^B$. $I_n$ improves $I_n^{Bb}$ at both left and right tails and its histogram shape appears more symmetric. Similarly $J_n$ improves $J_n^C$ at both left and right tails, and the histogram of $J_n$ is more symmetric than that of $J_n^C$. Figure 7.2 gives the Q-Q plots of the distributions of all six test statistics. Clearly $J_n$ is the closest to normal which is its limiting distribution according to Theorem 4. Hence $J_n$ performs more accurately in goodness-of-fit tests.

### 7.2.2 Correctly Specified and Misspecified

In this section we intentionally misspecify the tilt function by $h(x) = x$. This time we only focus on $I_n^{Bb}$, $I_n$, $J_n^C$ and $J_n$. We repeat the same procedure as we have done in Section (7.2.1) above to derive the simulated distributions of $I_n^{Bb}$, $I_n$, $J_n^C$ and $J_n$. Together with the corresponding results in Section (7.2.1) , we get Figure 7.3 which gives the comparisons of the histograms of these four test statistics under both correctly specified and misspecified tilts. The histograms of the test statistics when the tilt function is correctly specified are given in red color in Figure 7.3 and those are without color when the tilt function is misspecified. We consider the change from correctly specified to misspecified. We can see that when tilt function is misspecified, all the values of the test statistics increase and therefore lead to rejection of null hypothesis $H_0$. Our proposed test statistics $I_n$ and $J_n$ increase greatly so that they make a strong rejection very conclusively under the misspecified case. However, $I_n^{Bb}$

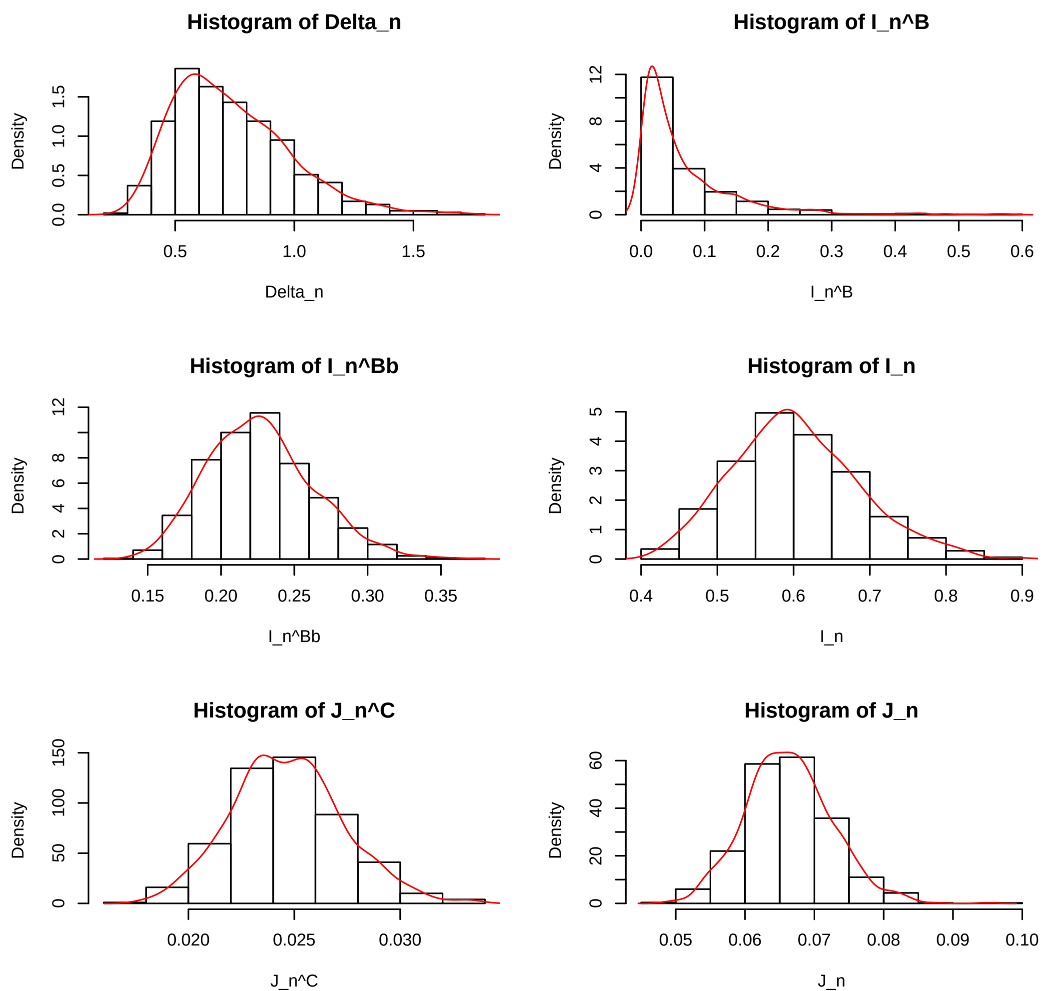


Figure 7.1: *Histograms of* $\Delta_n$, $I_n^B$, $I_n^{Bb}$, $I_n$, $J_n^C$ *and* $J_n$. *The reference* $U$ *is sampled from* $N(0,1)$. *The distortion* $Z$ *is sampled from* $N(0,2)$. *Sample sizes* $(n_0, n_1) = (500, 500)$. $I_n^B$, $I_n^{Bb}$ *and* $I_n$ *use the Gaussian kernel.* $J_n^C$ *and* $J_n$ *use the Epanechnikov kernel.* $b = 0.02$ *is the bandwidth for* $I_n^{Bb}$, $I_n$, $J_n^C$ *and* $J_n$. $b = 1$ *is the bandwidth for* $I_n^B$. *Simulation repetitions=1000.*

and $J_n^C$ do not increase much. Thus the two distributions of the same test statistic, which are under the correctly specified case and under the misspecified case, may have overlaps! This could lead to a wrong decision of the goodness-of-fit test like we have discussed in Chapter 5. On the other hand, the distributions of $I_n$ and $J_n$ do not suffer an overlap!

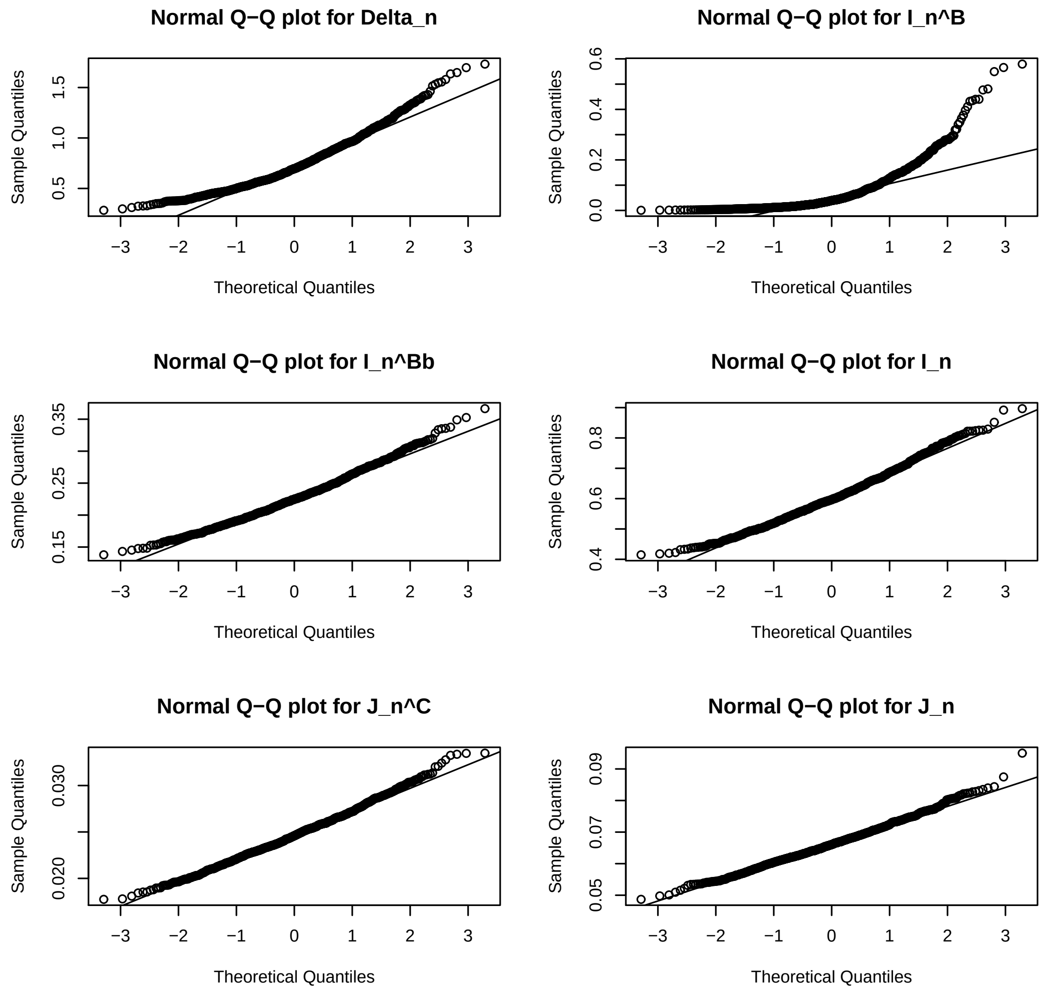


Figure 7.2: *Normal Q-Q plot for* $\Delta_n$, $I_n^B$, $I_n^{Bb}$, $I_n$, $J_n^C$ *and* $J_n$.

## 7.3 Conclusion

In this dissertation we have discussed six goodness-of-fit test statistics. $\Delta_n$ performs well in data simulations. However, the unknown distribution of $\Delta_n$ is always a problem. The fixed bandwidth used by $I_n^B$ may lead to inaccurate estimates of the underlying density. Thus the test statistic $I_n^B$ is not sufficiently reliable. $I_n$ improves $I_n^{Bb}$ by better distribution shapes on both the left and right tails when the tilt function is correctly specified, and increases more pronouncedly when the tilt

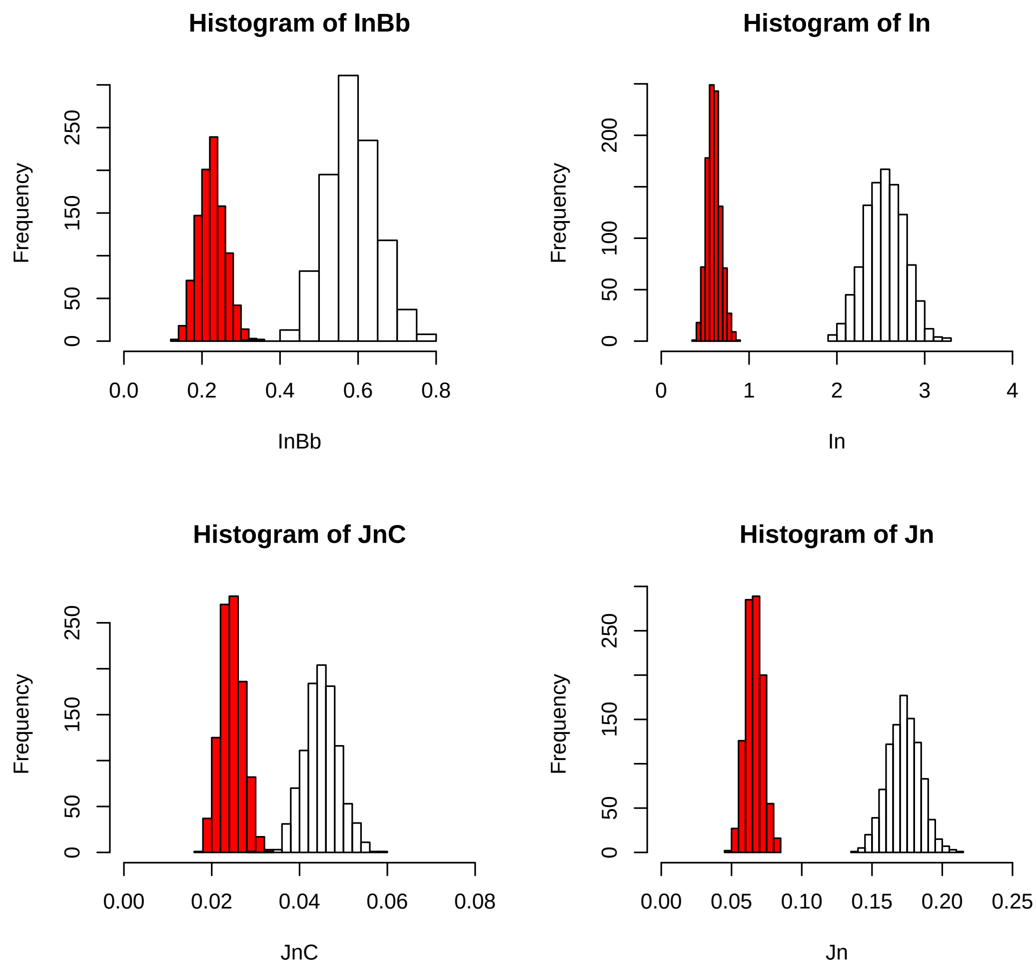


Figure 7.3: *Histograms of $I_n^{Bb}$, $I_n$, $J_n^C$ and $J_n$ when tilt functions are both correctly specified $h(x) = x^2$ (in red) and misspecified $h(x) = x$ (in white).*

function is misspecified, which helps to reject model (1.1) when it should be rejected. Although we know the distributions of $I_n$ and $I_n^{Bb}$, the bootstrap procedure is still needed to simulate the distributions of these test statistics. On the other hand, $J_n$ and $J_n^C$ can perform goodness-of-fit tests without the bootstrap procedure. It seems that the normal approximation of $J_n$ is more apparent as compared with that of $J_n^C$. $J_n$ can perform as well as $J_n^C$ in correctly specified cases and improves $J_n^C$ in misspecified cases. The distributions of our proposed statistics $I_n$ and $J_n$ appear more symmetric than those of $I_n^{Bb}$ and $J_n^C$ in our simulation study.

# Appendix A

## Proof of Lemma 5

In this section, we will provide the proof for the Lemma 5 which is not given by Cheng and Chu (2004)[6].

Let $(t-x)/b = y$, then $x = t - by$. Note that since $K(y)$ is symmetric, $\int yK(y)dy = 0$ and $\int yK^2(y)dy = 0$. These facts will be used in this section many times, including in the following proof.

To prove (4.3),

$$\begin{aligned}
E(Q_1(t)) &= E\big[\sum_{k=1}^{n} p(t_i)K_b(t-t_i)\big] \\
&= E\big[\sum_{i=1}^{n_0} p(u_i)K_b(t-u_i)\big] + E\big[\sum_{j=1}^{n_1} p(z_j)K_b(t-z_j)\big] \\
&= n_0 \int p(x)K_b(t-x)g(x)dx + n_1 \int p(x)K_b(t-x)w(x)g(x)dx \\
&= \int K_b(t-x)g(x)dx \\
&= \int K(y)g(t-by)dy \\
&= \int K(y)\cdot\big[g(t) - byg'(t) + \frac{1}{2}b^2y^2g^{(2)}(t) + \mathcal{O}(b^3)\big]dy \\
&= g(t) + \frac{1}{2}b^2g^{(2)}\int y^2K(y)dy + \mathcal{O}(b^3) \\
&= g(t) + \frac{1}{2}b^2\kappa_2 g^{(2)}(t) + \mathcal{O}(b^3).
\end{aligned}$$

(4.4) is the variance of $Q_1(t)$, we have

$$
\begin{aligned}
Var(Q_1(t)) &= Var\big[\sum_{k=1}^{n} p(t_i)K_b(t-t_i)\big] \\
&= Var\big[\sum_{i=1}^{n_0} p(u_i)K_b(t-u_i)\big] + Var\big[\sum_{j=1}^{n_1} p(z_j)K_b(t-z_j)\big] \\
&= n_0 \int p^2(x)K_b^2(t-x)g(x)dx + n_1 \int p^2(x)K_b^2(t-x)w(x)g(x)dx \\
&\quad -n_0\big(\int p(x)K_b(t-x)g(x)dx\big)^2 \\
&\quad -n_1\big(\int p(x)K_b(t-x)w(x)g(x)dx\big)^2 \\
&= \int p(x)K_b^2(t-x)g(x)dx - n_0\big(\int p(x)K_b(t-x)g(x)dx\big)^2 \\
&\quad -n_1\big(\int p(x)K_b(t-x)w(x)g(x)dx\big)^2.
\end{aligned}
$$

Since

$$
\begin{aligned}
\int p(x)K_b^2(t-x)g(x)dx &= \int p(t-by)\frac{1}{b^2}K^2(y)g(t-by)bdy \\
&= \frac{1}{b}\int \Big(p(t) - byp'(t) + \mathcal{O}(\frac{b^2}{n})\Big)K^2(y)\Big(g(t) - byg'(t) \\
&\qquad +\mathcal{O}(b^2)\Big)dy \\
&= \frac{1}{b}p(t)g(t)\int K^2(y)dy + \mathcal{O}(\frac{b}{n}) \\
&= \frac{1}{nb}\cdot\frac{\kappa_{sq}g(t)}{M(t)} + \mathcal{O}(\frac{b}{n})
\end{aligned}
$$

and

$$
\begin{aligned}
\big(\int p(x)K_b(t-x)g(x)dx\big)^2 &= \big(\int p(t-by)\frac{1}{b}K(y)g(t-by)bdy\big)^2 \\
&= \big(p(t)g(t) + \mathcal{O}(\frac{b^2}{n})\big)^2 \\
&= \frac{g^2(t)}{n^2M^2(t)} + \mathcal{O}(\frac{b^2}{n^2})
\end{aligned}
$$

and

$$\begin{aligned}\Big(\int p(x)K_b(t-x)w(x)g(x)dx\Big)^2 &= \Big(\int p(t-by)\frac{1}{b}K(y)w(t-by)g(t-by)bdy\Big)^2 \\ &= \big(p(t)w(t)g(t)+\mathcal{O}(\frac{b^2}{n})\big)^2 \\ &= \frac{w^2(t)g^2(t)}{n^2M^2(t)}+\mathcal{O}(\frac{b^2}{n^2})\end{aligned}$$

then

$$\begin{aligned}Var(Q_1(t)) &= \frac{1}{nb}\cdot\frac{\kappa_{sq}g(t)}{M(t)}+\mathcal{O}(\frac{b}{n})-n_0\cdot\frac{g^2(t)}{n^2M^2(t)}-n_1\cdot\frac{w^2(t)g^2(t)}{n^2M^2(t)}-\mathcal{O}(\frac{b^2}{n}) \\ &= \frac{1}{nb}\cdot\frac{\kappa_{sq}g(t)}{M(t)}+\mathcal{O}(\frac{1}{n}).\end{aligned}$$

To prove (4.5),

$$\begin{aligned}E(Q_2(t)) &= E\big[\sum_{k=1}^{n}p^2(t_i)n_1w(t_i)K_b(t-t_i)\big] \\ &= E\big[\sum_{i=1}^{n_0}p^2(u_i)n_1w(u_i)K_b(t-u_i)\big]+E\big[\sum_{j=1}^{n_1}p^2(z_j)n_1w(z_j)K_b(t-z_j)\big] \\ &= n_0\int p^2(x)n_1w(x)K_b(t-x)g(x)dx \\ &\quad +n_1\int p^2(x)n_1w(x)K_b(t-x)w(x)g(x)dx \\ &= \int p(x)n_1w(x)K_b(t-x)g(x)dx \\ &= \int p(t-by)n_1w(t-by)\frac{1}{b}K(y)g(t-by)bdy \\ &= n_1p(t)w(t)g(t)+\mathcal{O}(b^2) \\ &= \frac{\rho}{1+\rho}\cdot\frac{w(t)g(t)}{M(t)}+\mathcal{O}(b^2).\end{aligned}$$

To prove (4.6),

$$
\begin{aligned}
&Var\big(Q_2(t)\big)\\
&= Var\big[\sum_{k=1}^{n} p^2(t_i) n_1 w(t_i) K_b(t-t_i)\big]\\
&= Var\big[\sum_{i=1}^{n_0} p^2(u_i) n_1 w(u_i) K_b(t-u_i)\big] + Var\big[\sum_{j=1}^{n_1} p^2(z_j) n_1 w(z_j) K_b(t-z_j)\big]\\
&= n_0 \int p^4(x) n_1^2 w^2(x) K_b^2(t-x) g(x) dx\\
&\quad + n_1 \int p^4(x) n_1^2 w^2(x) K_b^2(t-x) w(x) g(x) dx\\
&\quad - n_0 \big( \int p^2(x) n_1 w(x) K_b(t-x) g(x) dx\big)^2\\
&\quad - n_1 \big( \int p^2(x) n_1 w(x) K_b(t-x) w(x) g(x) dx\big)^2\\
&= \int p^3(x) n_1^2 w^2(x) K_b^2(t-x) g(x) dx - n_0 \big( \int p^2(x) n_1 w(x) K_b(t-x) g(x) dx\big)^2\\
&\quad - n_1 \big( \int p^2(x) n_1 w(x) K_b(t-x) w(x) g(x) dx\big)^2.
\end{aligned}
$$

Like the calculations above for (4.4), we derive every term for $Var(Q_2(t))$,

$$
\begin{aligned}
\int p^3(x) n_1^2 w^2(x) K_b^2(t-x) g(x) dx &= n_1^2 \int p^3(t-by) w^2(t-by) \frac{1}{b^2} K^2(y)\\
&\qquad \times g(t-by) b dy\\
&= \frac{1}{b} n_1^2 p^3(t) w^2(t) g(t) \int K^2(y) dy + \mathcal{O}(\frac{b}{n})\\
&= \frac{1}{nb} \Big( \frac{\rho}{1+\rho} w(t) \Big)^2 \frac{\kappa_{sq} g(t)}{M^3(t)} + \mathcal{O}(\frac{b}{n})
\end{aligned}
$$

and

$$
\begin{aligned}
\big( \int p^2(x) n_1 w(x) K_b(t-x) g(x) dx\big)^2 &= \big( \int p^2(t-by) n_1 w(t-by) \frac{1}{b} K(y)\\
&\qquad \times g(t-by) b dy\big)^2\\
&= \big( n_1 p^2(t) w(t) g(t) + \mathcal{O}(\frac{b^2}{n})\big)^2\\
&= \frac{1}{n^2} (\frac{\rho}{1+\rho})^2 \frac{w^2(t) g^2(t)}{M^4(t)} + \mathcal{O}(\frac{b^2}{n^2})
\end{aligned}
$$

and

$$
\begin{aligned}
& \Big(\int p^2(x)n_1 w(x)K_b(t-x)w(x)g(x)dx\Big)^2 \\
= & \Big(\int p^2(t-by)n_1 w^2(t-by)\frac{1}{b}K(y)g(t-by)bdy\Big)^2 \\
= & \big(n_1 p^2(t)w^2(t)g(t)+\mathcal{O}(\frac{b^2}{n})\big)^2 \\
= & \frac{1}{n^2}(\frac{\rho}{1+\rho})^2\frac{w^4(t)g^2(t)}{M^4(t)}+\mathcal{O}(\frac{b^2}{n^2}).
\end{aligned}
$$

Thus

$$
\begin{aligned}
Var\big(Q_2(t)\big) &= \frac{1}{nb}\Big(\frac{\rho}{1+\rho}w(t)\Big)^2\frac{\kappa_{sq}g(t)}{M^3(t)}+\mathcal{O}(\frac{b}{n})+\frac{n_0}{n^2}(\frac{\rho}{1+\rho})^2\frac{w^2(t)g^2(t)}{M^4(t)} \\
&\quad +\mathcal{O}(\frac{b^2}{n^2})+\frac{n_1}{n^2}(\frac{\rho}{1+\rho})^2\frac{w^4(t)g^2(t)}{M^4(t)}+\mathcal{O}(\frac{b^2}{n^2}) \\
&= \frac{1}{nb}\Big(\frac{\rho}{1+\rho}w(t)\Big)^2\frac{\kappa_{sq}g(t)}{M^3(t)}+\mathcal{O}(\frac{1}{n}).
\end{aligned}
$$

Similarly we can get the mean and variance of $Q_3(t)$,

$$
\begin{aligned}
E\big(Q_3(t)\big) &= \frac{\rho}{1+\rho}\cdot\frac{h(t)w(t)g(t)}{M(t)}+\mathcal{O}(b^2) \\
Var\big(Q_3(t)\big) &= \frac{1}{nb}\Big(\frac{\rho}{1+\rho}w(t)\Big)^2\frac{\kappa_{sq}h(t)'h(t)g(t)}{M^3(t)}+\mathcal{O}(\frac{1}{n}).
\end{aligned}
$$

Next we are considering the expectation and variance of $\hat{\alpha}$ and $\hat{\beta}$. Like in section (4.6), we assume that we only consider all the variables and integrals in the bounded interval $[-L, L]$. From (2.19) in Lu (2007)[25], and combined with the assumption of $g(\cdot)$ and $K(\cdot)$, we know that $\frac{1}{n}Var\big(\partial\ell(\alpha_0,\beta_0)/\partial\alpha\big)$, $\frac{1}{n}Var\big(\partial\ell(\alpha_0,\beta_0)/\partial\beta\big)$ and $\frac{1}{n}Cov\big(\partial\ell(\alpha_0,\beta_0)/\partial\alpha,\partial\ell(\alpha_0,\beta_0)/\partial\beta\big)$ all are bounded uniformly for $n$. So we have $\frac{1}{\sqrt{n}}\big(\partial\ell(\alpha_0,\beta_0)/\partial\alpha,\partial\ell(\alpha_0,\beta_0)/\partial\beta\big)'$ is uniformly integrable. Combining Lemma 1, we have

$$
\begin{pmatrix} \hat{\alpha}-\alpha_0 \\ \hat{\beta}-\beta_0 \end{pmatrix} = \frac{1}{n}\mathbf{S}^{-1}\begin{pmatrix} \frac{\partial l(\alpha_0,\beta_0)}{\partial\alpha} \\ \frac{\partial l(\alpha_0,\beta_0)}{\partial\beta} \end{pmatrix} + \frac{1}{\sqrt{n}}\varsigma_n
$$

where $\varsigma_n$ is uniformly integrable and $\varsigma_n \to 0$ in distribution. Following Theorem 25.12 in Billingsley (1995)[2], we have

$$E(\hat{\alpha} - \alpha) = o(\frac{1}{\sqrt{n}}),\ E(\hat{\beta} - \beta) = o(\frac{1}{\sqrt{n}}),\ Var(\hat{\alpha}) = \mathcal{O}(\frac{1}{n}),\ Var(\hat{\beta}) = \mathcal{O}(\frac{1}{n})$$

Therefore, $E(Q_4(t)) = \mathcal{O}(\frac{1}{n})$ and $E(Q_4^2(t)) = \mathcal{O}(\frac{1}{n^2})$. The Lemma 5 is proved.

## Appendix B

## Values of $\kappa_{sq}$ and $\kappa_c$

We are using the Epanechnikov kernel which is

$$K(x) = \frac{3}{4}(1 - x^2)I_{[-1,1]}(x)$$

then

$$\kappa_{sq} = \int_{-1}^{1} K(x)^2 dx = \frac{9}{16}\int_{-1}^{1}(1 - x^2)^2 dx = \frac{3}{5}.$$

Let's consider $\kappa_c$,

$$\begin{aligned}
\kappa_c &= \int \Big(\int K(x)K(y-x)dx\Big)^2 dy \\
&= \int \Big(\frac{9}{16}\int (1-x^2)(1-(y-x)^2)I_{[-1,1]}(x)I_{[-1,1]}(y-x)dx\Big)^2 dy
\end{aligned}$$

Then the integral is in the area bounded by $x = -1$, $x = 1$, $y = x-1$ and $y = x+1$.

$$\begin{aligned}
\kappa_c &= \int_{-2}^{0}\Big(\frac{9}{16}\int_{-1}^{y+1}(1-x^2)(1-(y-x)^2)dx\Big)^2 dy \\
&\quad + \int_{0}^{2}\Big(\frac{9}{16}\int_{y-1}^{1}(1-x^2)(1-(y-x)^2)dx\Big)^2 dy \\
&= (\frac{9}{16})^2\Big(\int_{-2}^{0}(\frac{1}{30}y^5 - \frac{2}{3}y^3 - \frac{4}{3}y^2 + \frac{16}{15})^2 dy \\
&\quad + \int_{0}^{2}(-\frac{1}{30}y^5 + \frac{2}{3}y^3 - \frac{4}{3}y^2 + \frac{16}{15})^2 dy\Big) \\
&= 2\cdot(\frac{9}{16})^2 \int_{0}^{2}(-\frac{1}{30}y^5 + \frac{2}{3}y^3 - \frac{4}{3}y^2 + \frac{16}{15})^2 dy \\
&= 2\cdot(\frac{3}{160})^2 \int_{0}^{2}(y^{10} - 40y^8 + 80y^7 + 400y^6 - 1664y^5 + 1600y^4 + 1280y^3 \\
&\quad -2560y^2 + 1024)dy \\
&= 0.4337662
\end{aligned}$$

# Bibliography


[1] Anderson, N.H., Hall, P. and Titterington, D.M. (1994). "Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates", *Journal of Multivariate Analysis* **50** 41-54.

[2] Billingsley, P. (1995). "*Probability and measure, 3rd Edition*", Wiley, New York.

[3] Bondell, H.D. (2005). "Minimum distance estimation for the logistic regression model", *Biometrika* **92** 724-731.

[4] Bondell, H.D. (2007). "Testing goodness-of-fit in logistic case-control studies", *Biometrika* **94** 487-495.

[5] Cheng, K.F. and Chen, L.C. (2004). "Testing goodness-of-fit of a logistic regression model with caseCcontrol data ", *Journal of Statistical Planning and Inference* **124** 409-422.

[6] Cheng, K.F. and Chu, C.K. (2004). "Semiparametric Density Estimation under a Two-Sample Density Ratio Model ", *Bernoulli* **10** 583-604.

[7] Devroye, L. and Györfi, L. (1985). "Nonparametric Density Estimation: The $L_1$ View", New York: Wiley.

[8] Epanechnikov, V.A. (1969). "Nonparametric estimation of a multidimensional probability density", *Theor. Probab. Appl.*, **14** 153-158.

[9] Fokianos, K., Kedem, B., Qin, J. and Short, D.A. (2001). "A Semiparametric Approach to the One-Way Layout", *Technometrics* **43** 56-65.

[10] Fokianos, K. (2004). "Merging information for semiparametric density estimation", *J.R.Stat.Soc.Ser.B Stat. Methodol.* **66** 941-958.

[11] Fokianos, K. and Kaimi, I. (2006). "On the Effect of Misspecifying the Density Ratio Model", *Annals of the Institute of Statistical Mathematics* **58** 475-497.

[12] Gilbert, P.B. (2000). "Large sample theory of maximum likelihood estimates in semiparametric biased sampling models", *The Annals of Statistics* **28** 151-194.

[13] Gilbert, P.B. (2004). "Goodness-of-fit tests for semiparametric biased sampling models", *Journal of Statistical Planning and Inference* **118** 51-81.

[14] Gilbert, P.B., Lele, S.R. and Vardi, Y. (1999). "Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials", *Biometrika* **86** 27-43.

[15] Gill, R.D., Vardi, Y. and Wellner, J.A. (1988). "Large sample theory of empirical distributions in biased sampling models", *Annals of Statistics* **16** 1069-1112.

[16] Hall, P. (1984). "Central limit theorem for integrated square error of multivariate nonparametric density estimators", *Journal of multivariate analysis* **14** 1-16.

[17] Jong, P.D. (1987). "A central limit theorem for generalized quadratic forms", *Probability Theory and Related Fields* **75** 261-277.

[18] Katzoff, M., Zhou, w., Khan, D., Lu, G., Kedem, B. (2014). "Out-of-Sample Fusion in Risk Prediction", *Journal of Statistical Theory and Practice* **8** 444-459.

[19] Kallenberg, O. (2002). "*Foundations of Modern Probability*", Springer Science & Business Media.

[20] Kedem, B., De Oliveira, V. and Sverchkov, M. (2017). "*Statistical Data Fusion*", World Scientific, Singapore.

[21] Kedem, B., Lu, G., Wei, R. and Williams, P.D. (2008). "Forecasting mortality rates via density ratio modeling", *The Canadian Journal of Statistics* **36** 193-206.

[22] Kedem, B., Kim, E., Voulgaraki, A. and Graubard, B.I. (2009). "Two-dimensional semiparametric density ratio modeling of testicular germ cell data", *Statistics in Medicine* **28** 2147-2159.

[23] Kedem, B., Pan, L., Zhou, W. and Coelho, CA. (2016). "Interval estimation of small tail probabilitiesCapplications in food safety", *Statistics in Medicine* **35** 3229-3240.

[24] Kedem, B., Wolff, D.B. and Fokianos, K. (2004). "Statistical comparison of algorithms", *IEEE Transactions on Instrumentation and Measurement* **53** 770-776.

[25] Lu, G. (2007). "Asymptotic theory for multiple-sample semiparametric density ratio models and its application to mortality forecasting", *Ph.D. dissertation* Univ. of Maryland, College Park, MD.

[26] Prentice, R.L. and Pyke, R. (1979). "Logistic disease incidence models and case-control studies", *Biometrika* **66** 403-411.

[27] Qin, J. and Zhang, B. (1997). "A goodness of fit test for logistic regression models based on case-control data", *Biometrika* **84** 609-618.

[28] Qin, J. and Zhang, B. (2005). "Density estimation under a two-sample semi-parametric model", *Journal of Nonparametric Statistics* **17** 665-683.

[29] Silverman, B.W. (1986). "Density Estimation for Statistics and Data Analysis", *Chapman and Hall, London.*

[30] Vardi, Y. (1982). "Nonparametric Estimation in the Presence of Length Bias", *The Annals of Statistics* **10** 616-620.

[31] Voulgaraki, A., Kedem, B., and Graubard, B.I. (2012). "Semiparametric regression in testicular germ cell data", *The Annals of Applied Statistics* **6** 1185-1208.

[32] Wen, S. and Kedem, B. (2009). "A semiparametric cluster detection method a comprehensive power comparison with Kulldorff's method", *International Journal of Health Geographics* **8** Online journal without page numbers.

[33] White, H. (1982). "Maximum likelihood estimation of misspecified models", *Econometrica* **50** 1-25.

[34] Wu, J., Karunamuni, R. and Zhang, B. (2010). "Minimum Hellinger distance estimation in a two-sample semiparametric model", *Journal of Multivariate Analysis* **101** 1102-1122.

[35] Wu, J. and Karunamuni, R. (2012). "Efficient Hellinger distance estimates for semiparametric models ", *Journal of Multivariate Analysis* **107** 1-23.

[36] Xu, G. and Wang, S. (2011). "A goodness-of-fit test of logistic regression models for case-control data with measurement error", *Biometrika* **98** 877-886.

[37] Zhang, B. (1999). "A chi-squared goodness-of-fit test for logistic regression models based on case-control data", *Biometrika* **86** 531-539.

[38] Zhang, B. (2000). "A goodness of fit test for multiplicative-intercept risk models based on case-control data", *Statistica Sinica* **10** 839-865.

[39] Zhang, B. (2001). "An information matrix test for logistic regression models based on case-control data", *Biometrika* **88** 921-932.

[40] Zhang, B. (2002). "Assessing goodness-of-fit of generalized logit models based on case-control data", *Journal of Multivariate Analysis* **82** 17-38.

[41] Zhou, W. (2013). "Out of sample fusion", *Ph.D. dissertation*, University of Maryland, College Park.

[42] Zhu, Y., Wu, J. and Lu, X. (2013). "Minimum Hellinger distance estimation for a two-sample semiparametric cure rate model with censored survival data", *Computational Statistics* **28** 2495-2518.