MINIMUM DISTANCE ESTIMATION ON REGRESSION MODELS WITH DEPENDENT ERRORS AND GOODNESS OF FIT TEST OF ERRORS

By

Jiwoong Kim

A DISSERTATION

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

Statistics - Doctor of Philosophy

2016

ABSTRACT

MINIMUM DISTANCE ESTIMATION ON REGRESSION MODELS WITH DEPENDENT ERRORS AND GOODNESS OF FIT TEST OF ERRORS

By

Jiwoong Kim

This dissertation consists of three chapters. The first chapter introduces the models under consideration and motivates problems of interest. A brief literature review is also provided in this chapter.

The second chapter investigates minimum distance estimators of the parameters in the linear regression model with dependent errors: autoregressive errors and panel data errors. Asymptotic distributional properties of these estimators are discussed. A simulation study that compares the performance of some of these minimum distance estimators with Gaussian maximum likelihood, the generalized least squares, and the ordinary least squares estimators is also included. This simulation shows the superiority of the minimum distance estimator over the other estimators.

The third chapter compares two asymptotic distribution free methods for fitting an error distribution in the one sample location-scale model: Khmaladze transformation and empirical likelihood methods. The comparison is made from the perspective of empirical level and power obtained from simulations. When testing for normal and logistic null distributions, we try various alternative distributions and find that Khmaladze transformation method has better power in most cases.

ACKNOWLEDGMENTS

I wish to express my deep appreciation and thanks to Professor Hira L. Koul for suggesting thesis problems and guiding my doctoral study.

I would also like to thank Professor Vidyadhar S. Mandrekar for teaching courses which widened my understanding over the probability. Also, I thank Professors Daniel Hayes, Chae Young Lim, and Marianne Huebner for serving on my dissertation committee.

I thank my wonderful parents and brother for their constant love and support which enabled me to finally complete doctoral study.

Finally I would like to thank my friends Akshita Chawla, Abishek Kaul, and Daewoo Pak for their support during my doctoral study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO SYMBOLS

In what follows, for any functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, $|f|_H^2$ and $|f - g|_H^2$ denote that $\int f^2(y)\, dH(y)$ and $\int \{f(y) - g(y)\}^2\, dH(y)$, respectively. For a real vector $u \in \mathbb{R}^q$, $\|u\|$ denotes Euclidean norm. For any random variable Y, $\|Y\|_p$ denotes $(E|Y|^p)^{1/p}$ whenever it exists.

Chapter 1

Introduction

In the past, extensive research has been done on m.d. estimation method, where the estimators are obtained by minimizing some dispersions or pseudo distances between the data and the underlying model. Many authors have demonstrated robustness properties of a variety of m.d. estimators. In the context of the one sample problem, Beran (1977) showed that the minimum Hellinger distance estimator that minimizes Hellinger distance between the modeled parametric density and its nonparametric estimator is robust against small perturbation in the underlying model. Through the Monte Carlo methods, Parr and Schucany (1979) demonstrated robustness of several m.d. estimators of the location parameter based on various L_2-type distances. In their simulation, they generated a variety of mixtures of distributions and showed several L_2-type m.d. estimators are competitive in terms of asymptotic variances and robustness. Millar (1981) proved local asymptotic minimaxity of a fairly large class of Cramér-von Mises (CvM) type m.d. estimators. Donoho and Liu (1988a) showed that a large class of these m.d. estimators is "automatically" robust in the sense that they have the smallest possible variance among estimators when there are departures from the assumed model. Donoho and Liu (1988b), however, demonstrated that the minimum Kolmogorov-Smirnov distance estimator of normal location model can be affected by outliers located beyond the $\pm 3\sigma$ points and hence have an arbitrarily large asymptotic variance. All of these papers deal with the one or two sample models only.

Consider the linear regression model

$$Y_i = x_i'\beta + \varepsilon_i, \tag{1.1}$$

where $E\varepsilon_i \equiv 0$, $x_i = (1, x_{i1}, ..., x_{ip})' \in \mathbb{R}^{p+1}$ with $x_{ij}, j = 1, \cdots, p, i = 1, \cdots, n$ being non random design variables, and where $\beta = (\beta_0, \beta_1, ..., \beta_p)' \in \mathbb{R}^{p+1}$ is the parameter vector of interest. When ε_i's are independent identically distributed (i.i.d.) random variables (r.v.'s) with a known distribution function (d.f.) F, Koul and de Wet (1983) proposed a class of m.d. estimators by minimizing a class of L_2-distances between a weighted residual empirical processes and the error d.f. F. Koul (1985, 1986) extended this methodology to the case where the error distribution is unknown but symmetric around zero, and to autoregressive models with symmetric innovation, respectively. Furthermore, it was shown in these papers that in terms of the asymptotic variance some of these m.d. estimators remain competitive with other estimators including Wilcoxon rank, the least absolute deviation, the least square and normal scores estimators of β. Also robustness of these m.d. estimators against the extremes in the errors such as certain gross errors was established. For more detail, see chapter 5 in the monograph by Koul (2002). Given the above mentioned efficiency properties of the m.d. estimators, it is expected that the m.d. estimation method will still remain competitive when we extend the domain of its applications to the linear regression model with dependent errors, which is the aim of this dissertation. Chapter 2 demonstrates empirical efficiency - in terms of bias and asymptotic variance - of the m.d. estimation method through simulation studies when it is applied to the linear regression model with autoregressive (AR) errors and one with panel data.

Section 2.1 considers the linear regression model with AR errors. In both social sciences (e.g. economics) and natural sciences (e.g. climatology, hydrology) linear regression models with AR errors are found to be highly useful. For example, time series of total Gross National Product (GNP) in the United States during the postwar era has shown the linear time trend - often referred to as long term trend. In the series, the cyclical fluctuation around the linear time trend - often referred to as business cycles or short term shock - is commonly observed. The true underlying structure of these business cycles has been of interest to the macroeconomists. A class of macroeconomists, known as real business-cycle school, believes these business cycles respond to the changes in the long term trend; see, e.g., Prescott (1987). It is well-known that the cyclical fluctuation of many macroeconomic variables - including GNP, total consumption, etc. - is generally autocorrelated, and AR model can explain it well. The global temperature and its long term trend in the climatology is another example which shows that the linear model with AR errors is of good use. Bloomfield (1992) decomposed the time series of the past 100 year global temperature into the linear long term trend and the short term errors and fitted various models including AR model to the errors. He showed that AR model is fitted better to the errors.

Section 2.2 considers the linear regression model with panel data. The repeated observations on the same cross section unit, e.g., household, individual, firm, are called panel data; some authors prefer to call them longitudinal data. The most famous example of panel data is the Panel Study of Income Dynamics (PSID), collected by the Institute of Social Research at the University of Michigan. PSID has collected information about the annual economic status such as employment status, income changes, and marital status from more than 5000 households since 1968. Panel data analysis has been popular in the literature of econometrics due to several merits. These merits brought attention of econometricians to the panel data analysis, and extensive research has been done on the panel data; see Baltagi (2001, p.5) and references therein. Among all merits, the exemplary merit of the panel data analysis is that it considers heterogeneity of cross section units through time-invariant individual effect. Moulton (1986) showed that inappropriate usage of ordinary least squares estimation without considering the individual effect results in extremely huge downward bias of standard errors of the estimators. Another merit is attributed to the fundamental structure of the panel data; repeated observations on the same cross section unit is appropriate to study the dynamics of some social phenomena, and hence, the panel data analysis provides policy makers with good understanding of them. For example, high unemployment has been plaguing most countries during the post war era, and policy makers struggle against unemployment since it has the devastating affection on overall society such as poverty and surges of crime rate. Increasing the unemployment benefits during the time of rampant unemployment in the society has always been controversial. One criticism of unemployment benefits is it causes moral hazard, that is, it induces less effort of job-search. Cross-sectional data can provide unemployment rate at a fixed time while panel data are able to show the proportion of individuals whose employment status changed over the time periods. Upon the time of determining whether unemployment is chronic or transitory, panel data are very useful and provide guidance for policy makers on how to make a government initiative to increase unemployment benefits. For example, Ashenfelter (1978) investigated how federal training program affects future income of program participants through comparing groups of participants and nonparticipants. For each group, he analyzed the income records before and after the program and drew a conclusion that participating in the program positively affects an increase of future income.

Motivated by the above mentioned merits of panel data, Section 2.2 investigates applications of the m.d. estimation method to the regression model with panel data. In the sequel we refer to the regression model with panel data as panel regression model. Consider panel regression model

$$
\begin{aligned}
Y_{it} &= \boldsymbol{x}'_{it}\beta + \varepsilon_{it}, \qquad\qquad (1.2)\\
\varepsilon_{it} &= \gamma_i + \nu_{it}, \qquad\qquad i = 1, 2, ..., n;\ t = 1, 2, ..., T,
\end{aligned}
$$

where i and t denote cross section and time, respectively, $\boldsymbol{x}_{it} = (x_{it}^1, ..., x_{it}^p)' \in \mathbb{R}^p$ are non random design variables, and $\beta = (\beta_1, ..., \beta_p)' \in \mathbb{R}^p$ is the parameter vector of interest. We distinguish the vector $\boldsymbol{x}_{it}$ above from the scalar x_{it} in the model (1.1) by writing it in a bold font. The notable difference between the panel regression model and the model (1.1) is the fact that the variables of the first vary with cross section unit and time. Another difference is the fact that the error ε_{it} is decomposed into two unobservable components: time-invariant individual effect (γ_i) and the remainder disturbance (ν_{it}) which varies with cross section and time.

When we implement the CvM type m.d. estimation method, one common problem we encounter is the choice of the integrating measure. Koul (2002) showed that choosing optimal measure in the case of the i.i.d. errors in the model (1.1) yields the most efficient estimators of regression parameters. For example, the Lebesgue and degenerate measures are recommended for the efficient estimators of the regression parameters when the error follows the logistic and Laplace distributions, respectively; see Koul (2002, p.207) for the detail. Therefore, specifying the exact distribution of the error plays an important role in obtaining the efficient estimators of regression parameters.

Chapter 3 discusses a goodness-of-fit testing problem which sheds some light on specifying the error distribution. A classical goodness-of-fit testing problem, i.e., the problem of testing whether a random sample comes from a specific distribution or from a given parametric family of distributions, has been of interest to many fields for a long time. For example, the normality of sample has been commonly assumed in the vast literature of social and physical sciences. Since the final result, then, heavily depends on normality assumption, goodness-of-fit test for normality has been a critical issue.

For goodness-of-fit test for general distributions, the various parametric and nonparametric tests have been proposed in the literature. The best known exemplary parametric and nonparametric tests are χ^2 test and Kolmogorov-Smirnov (K-S) test. The most attractive advantage of the K-S test is that asymptotic distribution of its test statistic under the null hypothesis does not depend on the null distribution, when fitting a known continuous distribution. However, it loses this property when fitting a parametric family of distributions; see, e.g., Durbin (1973).

Seeking tests which preserve the desirable feature of being distribution free, we come up with the two celebrated methods: Khmaladze martingale transformation (KMT) and empirical likelihood (EL). The KMT and EL methods are implemented to test for a parametric location-scale family of distributions; then a comparison of these two methods is made in order to show which method is superior. To that end, we report empirical levels and powers of KMT and EL methods in Chapter 3. In order to compare the two methods in terms of the empirical levels, we generate a random sample of observations from the chosen null distributions and compute those of the two methods. Similarly, we generate a random sample of observations from various alternative distributions and compute the corresponding empirical powers of the two methods.

Chapter 2

M.d. Estimation Method on Regression Model with Dependent Errors

2.1 Regression model with AR errors

Recall the linear regression model (1.1). The errors ε_i's are assumed to obey the following autoregressive model of a known order q. For some $\rho = (\rho_1, ..., \rho_q)' \in \mathbb{R}^q$,

$$\varepsilon_i = Z_i' \rho + \xi_i, \tag{2.1}$$

where the innovations $\{\xi_i,\ i = 0, \pm 1, \pm 2, ...\}$ are i.i.d. r.v.'s, ε_{i-1} is independent of ξ_i, $i \in \{0, \pm 1, \cdots\}$, and $Z_i = (\varepsilon_{i-1}, ..., \varepsilon_{i-q})' \in \mathbb{R}^q$. To ensure the stationarity of the error process we assume that all roots of the equation $z^q - \rho_1 z^{q-1} - \cdots - \rho_{q-1} z - \rho_q = 0$ are less than unity in modulus. In addition, the distribution of ξ_0 is assumed to be symmetric around zero, unknown otherwise.

Next, we shall introduce the basic processes and estimators of interest here. To begin with we assume the following about $n \times (p+1)$ design matrix X.

$$(X'X)^{-1}\, exists,\, \forall\, n \geq p+1; \quad \max_i x_i'(X'X)^{-1}x_i = o(1). \tag{2.2}$$

Now let $A = (X'X)^{-1/2}$ and a_j denote jth column of A. Let $D = ((d_{ik}))$, $1 \leq i \leq n$, $1 \leq k \leq p+1$, be an $n \times (p+1)$ matrix of real numbers and d_j denote jth column of D. As stated in Koul (2002, p.60), if $D = XA$ (i.e., $d_{ik} = x_i' a_k$), then under (2.2),

$$\sum_i^n d_{ik}^2 = 1, \qquad \max_{1\leq i\leq n} d_{ik}^2 = o(1), \qquad \text{for all } 1 \leq k \leq p+1. \tag{2.3}$$

Note that the symmetry of ξ_0 around zero implies that of the regression errors ε_i's. This motivates one to introduce, as in Koul (2002; Definition 5.3.1),

$$\begin{aligned} U_k(y,t) &:= \sum_{i=1}^n a_k' x_i \Big\{ I\big(Y_i - x_i't \leq y\big) - I\big(-Y_i + x_i't < y\big) \Big\}, \\ U(y,t) &:= (U_1(y,t), ..., U_{p+1}(y,t))', \quad y \in \mathbb{R}, \\ T(t) &:= \int \|U(y,t)\|^2\, dH(y), \quad t \in \mathbb{R}^{p+1}, \end{aligned}$$

where H is a $\sigma-$finite measure on $\mathbb{R}$ and symmetric around 0, i.e., $dH(-x) = -dH(x)$, $x \in \mathbb{R}$. Subsequently, define $\widehat{\beta}$ as

$$T(\widehat{\beta}) = \inf_t T(t). \tag{2.4}$$

Next, we define analogous estimators of ρ. Accordingly, let

$$\begin{aligned} S(y,r) &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i \Big\{ I\big(\varepsilon_i - Z_i'r \leq y\big) - I\big(-\varepsilon_i + Z_i'r < y\big) \Big\}, \quad y \in \mathbb{R}, \qquad (2.5) \\ M(r) &:= \int \|S(y,r)\|^2\, dH(y), \quad r \in \mathbb{R}^q. \end{aligned}$$

Since Z_i's are unobservable, we replace them in (2.5) by $\widehat{Z}_i = (\widehat{\varepsilon}_{i-1}, ..., \widehat{\varepsilon}_{i-q})$, where $\widehat{\varepsilon}_i = Y_i - X_i'\widehat{\beta}$. Finally, we are motivated to introduce

$$\widehat{S}(y,r) := n^{-1/2}\sum_{i=1}^{n} \widehat{Z}_i\Big\{ I\big(\widehat{\varepsilon}_i - \widehat{Z}_i'r \le y\big) - I\big(-\widehat{\varepsilon}_i + \widehat{Z}_i'r < y\big) \Big\}, \qquad (2.6)$$

$$\widehat{M}(r) := \int \|\widehat{S}(y,r)\|^2 \, dH(y),$$

and define $\widehat{\rho}$ as

$$\widehat{M}(\widehat{\rho}) = \inf_r \widehat{M}(r).$$

This section is organized as follows. Sections 2.1.1 and 2.1.2 investigate the asymptotic distributions of m.d. estimators of both regression and AR parameters. Findings of a finite sample simulations are described in Subsection 2.1.3. This section also includes a direct application of the m.d. estimation method to some real time series data - one hundred yearly data of global temperature. Proofs in this section are postponed to Section 2.3.

2.1.1 Asymptotic distribution of $\widehat{\beta}$

In this subsection we derive the asymptotic distribution of $\widehat{\beta}$ under the current setup. The basic method of the proof is similar to that of sections 5.4, 5.5 of Koul (2002). This method amounts to showing that $T(\beta + Au)$ is asymptotically uniformly quadratic in u belonging to a bounded set and $\|A^{-1}(\widehat{\beta} - \beta)\| = O_p(1)$. To achieve these goals we need the following assumptions, where $d_{ik} = x_i'a_k$ and which in turn have roots in section 5.5 of Koul (2002).

(A.1) The matrix $X'X$ is nonsingular and, with $A = (X'X)^{-1/2}$, satisfies

$$\limsup_{n\to\infty} n \max_{1\le i\le n} \|Ax_i\|^2 < \infty.$$

(A.2) The integrating measure H is $\sigma-$finite and symmetric around 0, and

$$\int_0^\infty (1-F)^{1/2} dH < \infty.$$

(A.3) For any real sequences $\{a_n\}$, $\{b_n\}$, $b_n - a_n \to 0$,

$$\limsup_{n\to\infty} \int_{a_n}^{b_n} \int f(y+x) dH(y) dx = 0.$$

(A.4) For $a \in \mathbb{R}$, define $a^+ := \max(a, 0)$, $a^- := a^+ - a$. Let $\theta_i := \|Ax_i\|$. For all $u \in \mathbb{R}^{p+1}$, $\|u\| \le b$, for all $\delta > 0$, and for all $1 \le k \le p+1$,

$$\limsup_{n\to\infty} \int \Big[\sum_{i=1}^n d_{ik}^{\pm} \big\{ F(y + u'Ax_i + \delta\theta_i) - F(y + u'Ax_i - \delta\theta_i) \big\} \Big]^2 dH(y) \le c\delta^2,$$

where c is a constant not depending on u and δ.

(A.5) Let $e_k \in \mathbb{R}^{p+1}$ be an elementary vector whose kth coordinate is 1. For each $u \in \mathbb{R}^{p+1}$ and all $1 \le k \le p+1$,

$$\int \Big[\sum_{i=1}^n d_{ik} \big\{ F(y + u'Ax_i) - F(y) \big\} - u'e_k\, f(y) \Big]^2 dH(y) = o(1).$$

(A.6) F has a continuous density f with respect to the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$.

(A.7) $0 < \int_0^\infty f^r dH < \infty$, $r = 1/2, 1, 2$.

Remark 2.1.1. Note that (A.1) implies (2.2) and (A.2) implies $\int_0^\infty (1-F) dH < \infty$. From Corollary 5.6.3 of Koul (2002), we note that in the case of i.i.d. errors, the asymptotic normality of $\widehat{\beta}$ was established under the weaker conditions (2.2) and $\int_0^\infty (1-F) dH <$

∞. The autoregressive dependence of the errors now forces us to assume the two stronger conditions (A.1) and (A.2).

Remark 2.1.2. Here we discuss examples of H and F that satisfy (A.2). Clearly it is satisfied by any finite measure H. Next consider the σ-finite measure H given by $dH \equiv \{F(1 - F)\}^{-1}dF$, F a continuous d.f. symmetric around zero. Then $F(0) = 1/2$ and

$$\int_0^\infty (1-F)^{1/2} dH = \int_0^\infty \frac{(1-F)^{1/2}}{F(1-F)} dF \leq 2\int_{1/2}^1 (1-u)^{-1/2} du < \infty.$$

Another useful example of a σ-finite measure H is given by $H(y) \equiv y$. For this measure, (A.2) is satisfied by many symmetric error d.f.s including normal, logistic, and Laplace. For example, for normal d.f., we do not have a closed form of the integral, but by using the well celebrated tail bound for normal distribution, see e.g., Theorem 1.4 of Durrett (2005), we obtain

$$\int_0^\infty \{1 - F(y)\}^{1/2} dy \leq (2\pi)^{-1/2} \int_0^\infty y^{-1/2} \exp(-y^2/4) dy = (2/\pi)^{1/2}\, \Gamma(1/4).$$

Recall from Koul (2002) that the $\widehat{\beta}$ and $\widehat{\rho}$ corresponding to $H(y) \equiv y$ are the extensions of the one sample Hodges-Lehmann estimator of the location parameter to the above regression and autoregressive models.

Remark 2.1.3. Consider condition (A.7). If f is bounded then $\int f^{1/2} dH < \infty$ implies the other two conditions in (A.7) for any σ-finite measure H. For $H(y) \equiv y$, $\int f^{1/2}(y)dy < \infty$ when f is normal, logistic or Laplace density. In particular, when $dH = \{F(1-F)\}^{-1}dF$ and F is logistic d.f., so that $dH(y) \equiv dy$, this condition is also satisfied.

To proceed further, we need an assumption on the time series $\varepsilon_i, i \in \mathbb{Z} = \{0, \pm 1, \pm 2\}$. Let $\mathcal{F}_m^l$ be the σ-field generated by $\varepsilon_m, \varepsilon_{m+1}, ..., \varepsilon_l$, $m \le l$. The sequence $\{\varepsilon_j,\ j \in \mathbb{Z}\}$ is said to satisfy the strongly mixing condition if for every $l \ge 1$,

$$\alpha_l(k) := \sup \Big\{ |P(A \cap B) - P(A)P(B)| :\ A \in \mathcal{F}_{-\infty}^l,\ B \in \mathcal{F}_{l+k}^{\infty} \Big\} \to 0,$$

as $k \to \infty$. We recall the following theorem from Athreya and Pantula (1986), which will be used later.

Theorem 2.1.1. *Consider the qth order autoregressive process* $\{\varepsilon_n\}$ *given by* $\varepsilon_n = \rho_1 \varepsilon_{n-1} + \rho_2 \varepsilon_{n-2} + \cdots + \rho_q \varepsilon_{n-q} + \xi_n$. *If*

(A.8.1) $E\big[\{\log |\xi_1|\}^+\big] < \infty$,

(A.8.2) *the distribution of* ξ_1 *has non-trivial absolutely continuous components,*

(A.8.3) $e_0 = (\varepsilon_0, \varepsilon_{-1}, ..., \varepsilon_{1-q})$ *is independent of* $\{\xi_j,\ j = 1, 2, ...\}$,

(A.8.4) $\{\xi_j\}$ *are i.i.d. r.v.'s, and*

(A.8.5) *the roots of the characteristic equation* $z^q - \rho_1 z^{q-1} - \cdots - \rho_{q-1} z - \rho_q = 0$ *are less than one in modulus,*

then $\{\varepsilon_n\}$ *is strongly mixing and stationary.*

In addition to the assumptions of Theorem 2.1.1, we assume the following. The sequence $\{\varepsilon_n\}$ is strongly mixing with mixing number α_l satisfying

$$\sum_{k=1}^{\infty} k^2 \alpha_l(k) < \infty, \quad \forall\ l \ge 1 \qquad \textbf{(A.8.6)}$$

The above assumptions (A.8.1)-(A.8.6) will be referred to as assumption (A.8). Next, define for $t \in \mathbb{R}^{p+1}$,

$$Q(t) = \int \left\| U(y, \beta) + 2A^{-1}(t - \beta) f(y) \right\|^2 dH(y).$$

We are ready to state the needed results. The first theorem establishes the needed asymptotic uniform quadraticity while the corollary shows the boundedness of a suitably standardized $\widehat{\beta}$.

Theorem 2.1.2. *Let* $\{Y_i, 1 \le i \le n\}$ *be as in the model* (1.1). *Assume that (A.1)-(A.8) hold. Then, for any* $0 < b < \infty$,

$$E \sup_{\|A^{-1}(t-\beta)\| \le b} \|T(t) - Q(t)\| = o(1). \tag{2.7}$$

Proof. See Section 2.3.

Corollary 2.1.1. *Suppose that the assumptions of Theorem 2.1.2 hold. Then for any* $\epsilon > 0$, $0 < M < \infty$ *there exists an* N_ε, *and* $0 < b_\epsilon < \infty$ *such that*

$$P\left(\inf_{\|A^{-1}(t-\beta)\| \ge b_\epsilon} T(t) \ge M \right) \ge 1 - \epsilon, \qquad \forall n \ge N_\epsilon. \tag{2.8}$$

Proof. See Section 2.3.

Theorem 2.1.3. *Under the assumptions of Theorem 2.1.2,*

$$A^{-1}(\widehat{\beta} - \beta) = -\{2|f|_H^2\}^{-1} \int f(y) U(y, \beta)\, dH(y) + o_p(1). \tag{2.9}$$

Proof. The proof follows from Theorem 2.1.2 and Corollary 2.1.1, as in the i.i.d. case illustrated in Koul (2002).

Next, define

$$\begin{aligned}
\psi(x) &:= \int_{-\infty}^{-x} f(y)\,dH(y) - \int_{-\infty}^{x} f(y)\,dH(y), \qquad (2.10)\\
s_{nik} &:= d_{ik}\psi(\varepsilon_i), \quad z_{njk} := \sum_{i=1}^{j} s_{nik}, \quad \boldsymbol{z}_{nj} := (z_{nj1},...,z_{nj,p+1})',\\
\kappa_{nk} &:= Var(z_{nnk})\\
&= Var(\psi(\varepsilon_0)) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} d_{ik}d_{jk}\,E\psi(\varepsilon_i)\psi(\varepsilon_j),\\
K(F,H) &:= \int_0^\infty\int_0^\infty (1-F(x\vee y))f(x)f(y)\,dH(x)dH(y).
\end{aligned}$$

Symmetry of the F around 0 yields $E\psi(\varepsilon_0)=0$. Note that $Var(\psi(\varepsilon_0)) = 8K(F,H) < \infty$ by (A.7). Moreover,

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} |d_{ik}d_{jk}\,E\psi(\varepsilon_i)\psi(\varepsilon_j)| \le n\max_i d_{ik}^2\;\; n^{-1}\sum_{i=1}^{n-1}\sum_{m=1}^{n-i}\alpha_l^{1/2}(m)\|\psi(\varepsilon_0)\|_2 < \infty,$$

by (2.38), Lemma 2.3.1 below, and (A.1). Let $\boldsymbol{\Sigma}_n$ denote covariance matrix of $\boldsymbol{z}_{nn}$. Write $\boldsymbol{\Sigma}_n = ((\boldsymbol{\Sigma}_{nkh}))$, $1\le k\le p+1$, $1\le h\le p+1$. Observe that

$$\boldsymbol{\Sigma}_{nkh} = Cov(z_{nnk}, z_{nnh}) = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ik}d_{jh}E\psi(\varepsilon_i)\psi(\varepsilon_j),$$

and hence, for $1\le k\le p+1$, $1\le h\le p+1$, $|\boldsymbol{\Sigma}_{nkh}|<\infty$. We recall a theorem from Mehra and Rao (1975).

Theorem 2.1.4. *For each $n\ge 1$, let $\{\xi_{nj},\ 1\le j\le n\}$ be strong mixing with mixing number α_l. Assume these r.v.'s to be uniformly bounded and $\{\alpha_l\}$ satisfying (A.8.6). Further assume*

$$\max_i c_{ni}^2/\tau_c^2 \to 0, \qquad (2.11)$$

and

$$\liminf_{n\to\infty} \sigma_n^2/\tau_c^2 > 0, \tag{2.12}$$

where $\sigma_n^2 = Var(\sum_{i=1}^n c_{ni}\xi_{ni})$ *and* $\tau_c^2 = \sum_{i=1}^n c_{ni}^2$. *Then*

$$\sigma_n^{-1}\sum_{i=1}^n c_{ni}\xi_{ni} \to_D N(0,1). \tag{2.13}$$

As a consequence of Theorem 2.1.4, we obtain the following results.

Corollary 2.1.2. *Assume* $\mathbf{\Sigma}_n$ *is positive definite for all* $n \geq p+1$. *In addition, assume that*

$$\sup_{u\in\mathbb{R}^{p+1},\|u\|=1} u'\mathbf{\Sigma}_n^{-1}u = O(1). \tag{2.14}$$

Then

$$\mathbf{\Sigma}_n^{-1/2}\boldsymbol{z}_{nn} \to_D N(0, I_{(p+1)\times(p+1)}), \tag{2.15}$$

where $I_{(p+1)\times(p+1)}$ *is the* $(p+1)\times(p+1)$ *identity matrix.*

Proof. To prove (2.15), it suffices to show that for any $\lambda \in \mathbb{R}^{p+1}$, $\lambda'\mathbf{\Sigma}_n^{-1/2}\boldsymbol{z}_{nn}$ is asymptotically normally distributed. But $\lambda'\mathbf{\Sigma}_n^{-1/2}\boldsymbol{z}_{nn} = \sum_{i=1}^n \lambda'\mathbf{\Sigma}_n^{-1/2}Ax_i\psi(\varepsilon_i)$, which is the sum as in Theorem 2.1.4 with $c_{ni} = \lambda'\mathbf{\Sigma}_n^{-1/2}Ax_i$ and $\xi_{ni} = \psi(\varepsilon_i)$. Note that

$$\tau_c^2 = \sum_{i=1}^n(\lambda'\mathbf{\Sigma}_n^{-1/2}Ax_i)^2 = \lambda'\mathbf{\Sigma}_n^{-1}\lambda, \qquad \sigma_n^2 = E\Big\{\sum_{i=1}^n(\lambda'\mathbf{\Sigma}_n^{-1/2}Ax_i)\psi(\varepsilon_i)\Big\}^2 = \|\lambda\|^2.$$

Also, we have

$$\max_{1\leq i\leq n} c_{ni}^2/\tau_c^2 \leq \max_{1\leq i\leq n}\frac{\|\lambda'\mathbf{\Sigma}_n^{-1/2}\|^2\|Ax_i\|^2}{\lambda'\mathbf{\Sigma}_n^{-1}\lambda} = \max_{1\leq i\leq n}\|Ax_i\|^2 \to 0,$$

by assumption (A.1). Finally, we obtain

$$\liminf_{n\to\infty} \sigma_n^2/\tau_c^2 \geq \|\lambda\|^2/(\limsup \lambda' \mathbf{\Sigma}_n^{-1}\lambda) > 0,$$

by (2.14). Hence, the desired result follows from Theorem 2.1.4.

Corollary 2.1.3. *In addition to the assumptions of Theorem 2.1.2, assume* (2.14). *Then*

$$\mathbf{\Sigma}_n^{-1/2} A^{-1}(\widehat{\beta} - \beta) \Rightarrow_{\mathcal{D}} \{2|f|_H^2\}^{-1} N(0, I_{(p+1)\times(p+1)}). \tag{2.16}$$

Proof. Claim (2.16) follows from Corollary 2.1.2 upon noting that

$$\int f(y)U(y,\beta)\, dH(y) = \boldsymbol{z}_{nn}.$$

2.1.2 Asymptotic distribution of $\widehat{\rho}$

In this section we investigate asymptotic distribution of $\widehat{\rho}$. Recall $M(r)$ and $\widehat{M}(r)$ from (2.5) and (2.6). Asymptotic uniform quadraticity of $M(\rho + v/\sqrt{n})$ in v is illustrated in the Section 7.4 of Koul (2002). See also Theorem 2.1.5 below. Since errors in M are, however, unobservable, M is no longer fit for the purpose of estimating ρ. To that end, we use $\widehat{M}$ instead. We first prove that the supremum over v in compact sets of the absolute difference between $M(\rho + v/\sqrt{n})$ and $\widehat{M}(\rho + v/\sqrt{n})$ is $o_p(1)$. As a consequence, we accomplish the desired asymptotic uniform quadraticity of $\widehat{M}(\rho+v/\sqrt{n})$ in v. Through this section, similar method of the proof to that used in the Section 7.4 of Koul (2002) will be employed. However, the proofs will be a little bit more complicated since they contain additional terms including ζ_{ni} and d_{ni} - see (2.18) below - as a consequence of using residuals instead of unobservable errors.

To begin with we state the following lemma.

Lemma 2.1.1. *Let $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ be positive definite (or semi-positive definite) $n \times n$ matrices such that*

$$\boldsymbol{M}_1 + \boldsymbol{M}_2 = I_{n\times n}.$$

Let λ_1 and λ_2 be eigenvalues of $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$. Then $0 \le \lambda_1,\ \lambda_2 \le 1$.

Proof. Since $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are positive definite (semi-positive definite), λ_1 and λ_2 are non-negative. Note that

$$0 \;=\; det(\boldsymbol{M}_2 - \lambda_2 I) = (-1)^n det(\boldsymbol{M}_1 - (1-\lambda_2)I),$$

and hence $(1-\lambda_2)$ is eigenvalue of $\boldsymbol{M}_1$. Therefore, $(1-\lambda_2) \ge 0$ since $\boldsymbol{M}_1$ is positive definite, which in turn implies that $\lambda_2 \le 1$. Similar fact holds for λ_1, thereby completing the proof of the lemma.

Next, we need to recall that for any $m_1 \times m_2$ real matrix $\boldsymbol{M}$, its spectral norm is $\|\boldsymbol{M}\|_2 := \sup_{u\in\mathbb{R}^{m_2},\|u\|=1} \|\boldsymbol{M}u\|_2$. Recall that

$$\|\boldsymbol{M}\|_2 = \big(\max\{\lambda :\ \lambda \text{ is an eigenvalue of } \boldsymbol{M}'\boldsymbol{M}\}\big)^{1/2}. \tag{2.17}$$

Partition $D = XA$ into three blocks

$$D' = [D'_{-i}\ D'_i\ D'_{+i}]$$

where D_{-i}, D_i, and D_{+i} are $(i-q-1)\times(p+1)$, $q\times(p+1)$, $(n-i+1)\times(p+1)$ matrices respectively. Note that $D'_{-i}D_{-i} + D'_iD_i + D'_{+i}D_{+i} = I_{(p+1)\times(p+1)}$. Lemma 2.1.1

implies $D_i' D_i$ has non-negative eigenvalues, which are less than or equal to one. By (2.17), $\|D_i\|_2 \leq 1$, and hence, for any $u \in \mathbb{R}^{p+1}$

$$\|D_i' u\| \leq \|D_i\|_2 \cdot \|u\| \leq \|u\|, \quad \forall\, i \geq 1.$$

Moreover, for $u \in \mathbb{R}^{p+1}$, we also have, with $c_i = Ax_i$,

$$\max_{1 \leq i \leq n} \|D_i u\|^2 \leq q\|u\|^2 \max_{1 \leq i \leq n} \|c_i\|^2.$$

Now, let $u = A^{-1}(\widehat{\beta} - \beta) \in \mathbb{R}^{p+1}$ and $v = n^{1/2}(r - \rho) \in \mathbb{R}^q$ where $r = (r_1, ..., r_q)' \in \mathbb{R}^q$. Define

$$\begin{aligned} \eta_{ni}(u) &:= \varepsilon_i - \widehat{\varepsilon}_i = (Y_i - x_i'\beta) - (Y_i - x_i'\widehat{\beta}) = u'c_i, \quad \vartheta_n := \max_{1 \leq i \leq n} \|c_i\|, \qquad (2.18) \\ \zeta_{ni}(u, v) &:= \eta_{ni}(u) - \sum_{k=1}^{q} r_k \eta_{n,i-k} = u'c_i - \rho' D_i u - v' D_i u / \sqrt{n}, \\ d_{ni}(\delta, u, v) &:= \delta\big\{(1 + q\|\rho\|)\vartheta_n + n^{-1/2}\|Z_i\| + n^{-1/2}\|u\| + n^{-1/2}\|v\| + n^{-1/2}\delta\big\}. \end{aligned}$$

To begin with we state a set of assumptions. The following assumptions are found similar to ones in the Section 7.4 of Koul (2002). In order to define randomly weighted residual empirical process, he used a measurable function g of r.v.'s as a random weight. By choosing various g's, he demonstrated m.d. estimators turn out to be well celebrated estimators such as Hodges-Lehmann estimator, the least absolute deviation estimator, etc. Here we only consider either $g(x) \equiv x$ or $g \equiv 1$. Let g_{i-k} denote $g(\varepsilon_{i-k})$. Hence, g_{i-k} assumes only the values of ε_{i-k} and 1 in the subsequent assumptions.

(B.1) (a) $E\varepsilon_0^4 < \infty$ and $E\xi^2 < \infty$.

(b) The d.f. F_ξ of the innovation ξ has a continuous density f_ξ with respect to the Lebesgue measure on $(\mathbb{R}, \mathcal{R})$.

(c) $0 < \int_0^\infty f_\xi^r dH < \infty$, $r = 0.5, 1, 2$.

(d) $0 < \int_0^\infty (1 - F_\xi) dH < \infty$.

(B.2) For all $\|u\| \le b$, $\|v\| \le b$, $\delta \in \mathbb{R}$, and for $k = 1, 2, ..., q$

$$E \int n^{-1} \sum_{i=1}^n g_{i-k}^2 |F_\xi(y + n^{-1/2} v' Z_i + \zeta_{ni}(u, v) + d_{ni}(\delta, u, v)) - F_\xi(y)| \, dH(y) = o(1).$$

(B.3) There exists a constant $0 < c < \infty$ such that for all $\delta > 0$, $\|u\| \le b$, $\|v\| \le b$, and for all $k = 1, 2, ..., q$

$$\liminf_n P\Bigg(\int n^{-1} \Big[\sum_{i=1}^n g_{i-k}^\pm \{ F_\xi(y + n^{-1/2} v' Z_i + \zeta_{ni}(u, v) + d_{ni}(\delta, u, v)) - F_\xi(y + n^{-1/2} v' Z_i + \zeta_{ni}(u, v) - d_{ni}(\delta, u, v)) \} \Big]^2 dH(y) \le c\delta^2 \Bigg) = 1,$$

where $g^\pm$ is as in the assumption (A.4).

(B.4) For every $\|u\| \le b$, $\|v\| \le b$, and for $k = 1, 2, ..., q$

$$\int n^{-1} \Big[\sum_{i=1}^n g_{i-k} \big[F_\xi(y + n^{-1/2} v' Z_i + \zeta_{ni}(u, v)) - F_\xi(y) - f_\xi(y) \{ n^{-1/2} v' Z_i + \zeta_{ni}(u, v) \} \big] \Big]^2 dH(y) = o_p(1).$$

(B.5)

$$\lim_{s \to 0} \int E\Big[\|Z_1\| \, f_\xi(y + s\|Z_1\|) \Big]^\ell dH(y) = E\|Z_1\|^\ell \int f_\xi^\ell dH, \;\; \ell = 1, 2.$$

Recall $S(y,r)$ in (2.5) and $\widehat{S}(y,r)$ in (2.6). Let $S_j(y,r)$ and $\widehat{S}_j(y,r)$ denote their jth coordinate, respectively. Note that

$$S_j(y,r) = n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-j}\Big\{\, I\big(\varepsilon_i - Z_i'r \le y\big) - I\big(-\epsilon_i + Z_i'r < y\big)\Big\}.$$

Next, let

$$\mathcal{Q}(r) := \sum_{j=1}^{q}\int \Big[S_j(y,\rho) + 2\sqrt{n}(r-\rho)'n^{-1}\sum_{i=1}^{n}\varepsilon_{i-j}Z_i\, f_\xi(y)\Big]^2\, dH(y).$$

The first theorem states the asymptotic uniform quadraticity of M, which is proved in Koul (2002, p.328). The second theorem states that the difference between M and $\widehat{M}$ is small enough to establish the asymptotic uniform quadraticity of $\widehat{M}$ while the corollary shows the boundedness of $\widehat{\rho}$.

Theorem 2.1.5. *Suppose that the autoregressive model* (2.1) *holds. Assume that (B.1)-(B.5) hold. Then, for any* $0 < b < \infty$*,*

$$\sup_{\sqrt{n}\|(r-\rho)\|\le b} |M(r) - \mathcal{Q}(r)| = o_p(1).$$

Theorem 2.1.6. *Suppose the assumptions of Theorem 2.1.5 hold. Assume that* $\widehat{\beta}$ *used for obtaining* $\widehat{\varepsilon}_i$ *in (2.1.6) satisfies* $A^{-1}\|\widehat{\beta}-\beta\| = O_p(1)$*. Then, for any* $0 < b < \infty$*,*

$$\sup_{\sqrt{n}\|r-\rho\|\le b} |\widehat{M}(r) - M(r)| = o_p(1).$$

Proof. See Section 2.3.

Corollary 2.1.4. *Suppose that the assumptions of Theorem 2.1.5. For any* $\epsilon > 0$, $0 < B < \infty$ *there exists an* N_ϵ, *and* $0 < b_\epsilon < \infty$ *such that*

$$P\left(\inf_{\sqrt{n}\|r-\rho\|\geq b_\epsilon} \widehat{M}(r) \geq B\right) \geq 1-\epsilon, \qquad \forall n \geq N_\epsilon.$$

Proof. The proof is similar to that of Lemma 5.5.4 of Koul (2002). The role of Γ_n in Lemma 5.5.4 is played by $n^{-1}\sum_i Z_i Z_i'$. Then the rest of the proof is exactly the same with the proviso that we use new $\Gamma_n = n^{-1}\sum_i Z_i Z_i'$.

Note that by the stationarity of $Z_i, i \in \mathbb{Z}$ and the Ergodic Theorem, $n^{-1}\sum_i Z_i Z_i' \to E(Z_1 Z_1')$, almost surely. Moreover, the covariance matrix $E(Z_1 Z_1')$ is positive definite. Thus for all sufficiently large n, $n^{-1}\sum_i Z_i Z_i'$ is almost surely positive definite. This fact is used throughout the proofs here without further mention.

Theorem 2.1.7. *Under the assumptions of Theorem 2.1.5,*

$$\sqrt{n}(\widehat{\rho}-\rho) = -\{2|f_\xi|_H^2 n^{-1}\sum_{i=1}^n Z_i Z_i'\}^{-1} \int S(y,\rho) f_\xi(y) dH(y) + o_p(1).$$

Consequently,

$$\sqrt{n}(\widehat{\rho}-\rho) \to_D N(0, (EZ_1 Z_1')^{-1}\tilde{\tau}^2)$$

where $\tilde{\tau}^2 := Var(\psi(\xi_1))/4(\int f_\xi^2 dH)^2$.

Proof. The proof follows from Theorem 2.1.6, Corollary 2.1.4, and Theorem 7.4.5 of Koul (2002).

Remark 2.1.4. Consider an estimator of $(EZ_1 Z_1')^{-1}\tilde{\tau}^2$. Note that $Var(\psi(\xi_1)) = 8K(F_\xi, H)$ where $K(F_\xi, H)$ is as in (2.10). When $H(y) \equiv y$, $K(F_\xi, H) = 1/3$. In order to estimate asymptotic variance of $\widehat{\rho}$, we need to estimate EZ_1Z_1' and $\int f_\xi^2(x)dx$. A valid estimator of EZ_1Z_1' is $n^{-1}\sum_i \widehat{Z}_i\widehat{Z}_i'$, and hence, it remains to estimate

$$B(f_\xi) := \int f_\xi^2(x)dx = \int f_\xi(x)dF_\xi(x). \tag{2.19}$$

We replace f_ξ and F_ξ with a kernel type density estimator and the empirical d.f. based on the estimated residuals. Let G be a probability density on $\mathbb{R}$, h_n be a sequence of positive numbers, $h_n \to 0$. Define, for $x \in \mathbb{R}$,

$$\begin{aligned}
\widehat{\xi}_i &:= \widehat{\varepsilon}_i - \widehat{Z}_i'\widehat{\rho}, \quad \widehat{F}_n(x) := n^{-1}\sum_{i=1}^n I(\widehat{\xi}_i \le x), \quad F_n(x) := n^{-1}\sum_{i=1}^n I(\xi_i \le x) \\
\widehat{f}_n(x) &:= (nh_n)^{-1}\sum_{i=1}^n G\big((x-\widehat{\xi}_i)/h_n\big), \quad f_n(x) := (nh_n)^{-1}\sum_{i=1}^n G\big((x-\xi_i)/h_n\big) \\
\widehat{B}_n &:= \int \widehat{f}_n(x)d\widehat{F}_n(x)
\end{aligned}$$

Lemma 2.1.2. *Assume that F_ξ has uniformly continuous density f_ξ w.r.t. the Lebesgue measure λ and $f_\xi > 0$ a.e. λ. Then, for any fixed u, v,*

$$\sup_{x\in\mathbb{R}}\left|\sqrt{n}\{\widehat{F}_n(x) - F_n(x)\} - n^{-1/2}\sum_{i=1}^n \big(Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big)\, f_\xi(x)\right| = o_p(1).$$

Proof. Note that

$$\widehat{\xi}_i = \xi_i - Z_i'v/\sqrt{n} - \zeta_{ni}(u,v).$$

Therefore,

$$\left|\sqrt{n}\{\widehat{F}_n(x) - F_n(x)\} - n^{-1/2}\sum_{i=1}^n \big(Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big)\, f_\xi(x)\right| \le W_{n1}(x) + W_{n2}(x),$$

where

$$
\begin{aligned}
W_{n1}(x) \quad &:= \quad \Big| n^{-1/2} \sum_{i=1}^{n} \big\{ I(\xi_i \leq x + Z_i' v/\sqrt{n} + \zeta_{ni}(u,v)) - F_\xi(x + Z_i' v/\sqrt{n} + \zeta_{ni}(u,v)) \big\} \\
&\qquad - n^{-1/2} \sum_{i=1}^{n} \big\{ I(\xi_i \leq x) - F_\xi(x) \big\} \Big|, \\
W_{n2}(x) \quad &:= \quad \Big| n^{-1/2} \sum_{i=1}^{n} \big\{ F_\xi(x + Z_i' v/\sqrt{n} + \zeta_{ni}(u,v)) - F_\xi(x) - \big(Z_i' v/\sqrt{n} + \zeta_{ni}(u,v) \big) f_\xi(x) \big\} \Big|.
\end{aligned}
$$

Thus, to prove the lemma, it suffices to show that

$$
\sup_{x \in \mathbb{R}} |W_{ni}(x)| = o_p(1), \qquad i = 1, 2.
$$

Together with the assumption of the uniform continuity of f_ξ and $\zeta(u,v)) = O(n^{-1/2})$, we obtain $\sup_x |W_{n1}(x)| = o_p(1)$ after applying the Theorem 2.2.3 of Koul (2002). Now consider W_{n2}. Let $z_{ni} := Z_i' v/\sqrt{n} + \zeta_{ni}(u,v)$ and $z_n := \max_i z_{ni}$. Note that $z_n = O_p(n^{-1/2})$. Therefore,

$$
\begin{aligned}
\sup_x |W_{n2}(x)| \quad &\leq \quad n^{-1/2} \sum_{i=1}^{n} \int_x^{x+z_{ni}} \sup_{|y-x| \leq z_{ni}} |f_\xi(y) - f_\xi(x)| dy \\
&\leq \quad \sqrt{n}\, z_n \sup_{|y-x| \leq z_n} |f_\xi(y) - f_\xi(x)| \rightarrow_p 0,
\end{aligned}
$$

where the convergence to zero in probability follows from the uniform continuity of f_ξ.

Theorem 2.1.8. *In addition to the assumptions of Lemma 2.1.2, assume that the following conditions hold.*

(i) $h_n > 0,\ h_n \rightarrow 0,\ n h_n^{1/2} \rightarrow \infty.$

(ii) *G is absolutely continuous with its a.e. derivative $\dot{G}$ satisfying $\int |\dot{G}(z)| dz < \infty$.*

Then

$$|\widehat{B}_n - B(f_\xi)| = o_p(1).$$

Proof. The proof is very similar to that of Theorem 4.5.3 of Koul (2002). Only difference arises in the part of showing $\|\widehat{f}_n - f_n\|_\infty = o_p(1)$ which requires the above Lemma 2.1.2. The rest of the proof is the same.

2.1.3 Simulation & application to real data

2.1.3.1 Simulation study

In this section, we present a simulation study corresponding to four symmetric innovations: normal, Laplace, logistic, and mixture of the two normals (MTN). In each case, we estimate β and ρ by m.d., generalized least squares (GLS), ordinary least squares (OLS), and Gaussian maximum likelihood (GML) methods. We report empirical bias, standard error (SE), and mean squared error (MSE) of these estimators. To obtain GLS estimators, we use `R` package `orcutt` which implements Cochrane-Orcutt (CO) iterative estimation procedure. To obtain m.d. estimators, we employ "two-stage" m.d. estimation method which is proposed by Kim (2016) to deal with the linear regression model with autocorrelated errors. To implement the two-stage m.d. estimation method, we use `R` package `KoulMde`. The package is available from Comprehensive `R` Archive Network (CRAN) at `https://cran.r-project.org/web/packages/KoulMde/index.html`.

In this simulation studies, $p = 3$ and $q = 1$, and $H(y) \equiv y$. We set the true $\beta = (-2, 3, 1.5, -4.3)'$ and true $\rho = 0.4$. For each $k = 1, 2, 3$, we obtain $\{x_{ik}\}_{i=1}^n$ in (1.1) as a random sample from the uniform distribution on $[0, 50]$. The innovation has a Laplace distribution if its density function is

$$f_{La}(x) := (2s_1)^{-1} \exp(-|x - \mu_1|/s_1)$$

while the density function of Logistic innovation is given by

$$f_{Lo}(x) := s_2^{-1} \exp(-|x - \mu_2|/s_2)/(1 + \exp(-|x - \mu_2|/s_2))^2.$$

When we generate $\{\xi\}_{i=1}^n$, we set mean of normal, Laplace, and logistic innovations at 0 (i.e., $\mu_1 = \mu_2 = 0$) since we assumed the innovation is symmetric. We set the standard deviation of normal innovation at 2 while both s_1 and s_2 are set at 5 for Laplace and logistic innovation, respectively. For MTN, we consider $(1-\delta)N(0, 2^2) + \delta N(0, 10^2)$ where $\delta = 0.1$. Then we generate $\{\varepsilon_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ subsequently using models (1.1) and (2.1). Table 1, 2, and 3 report biases, SE's and MSE's of estimators for the sample sizes 50, 100, and 200, each repeated 1,000 times. Author used High Performance Computing Center (HPCC) to accelerate the simulations. All of the simulations were done in the R-3.2.2.

First, we consider the normal innovation. When innovation is normal, and n is 50, the GML estimators show the smallest biases among all estimators of β_j's, $j = 0, 1, 2, 3$. However, as n increases, this fact does not hold any more; when n is 200, the GML estimators of β_0, β_1, and β_2 display the largest biases. When we consider the SE, the GLS estimators display the smallest SE regardless of n; following GLS, both the GML and the m.d. estimators show similar SE's. When we evaluate the performance of the estimators in terms of the MSE, we conclude, not surprisingly, that the GLS is the best, the GML and the m.d. are similar and the second best, and the OLS is the worst.

For non-Gaussian innovations, we come up with a different conclusion: the m.d. estimators outperform all other estimators. Note that the OLS method exhibits very poor performance which is not surprising. Also, note that the GML estimation method shows the loss of competitiveness compared to GLS and m.d. estimators since it assumes the normality of the innovation. Therefore, we leave the OLS and GML out of discussion for the non-Gaussian innovations. Note that weighing the merits of the GLS and the m.d. estimators in terms of bias is hard. For example, for the Laplace innovation when $n = 200$, the m.d. estimators of β_1 and ρ display smaller biases than the GLS estimators while the opposite is true for β_j, $j = 0, 2, 3$. When we consider logistic innovation, the result is similar. The m.d. estimators of β_1 and ρ again show smaller biases than the GLS estimators while the opposite is true for β_j, $j = 0, 3$; both the m.d. and the GLS estimators of β_1 show the same bias. However, the m.d. estimators display the smallest SE's regardless of n and innovations. As a result, the m.d. estimators display the smallest MSE's for all non-Gaussian innovations. One remarkable fact is the MSE's corresponding to the m.d. estimators are approximately 60% of those corresponding to the GLS estimators when the innovation is MTN with $n = 100$ or 200. Therefore we conclude that the m.d. estimators outperform all other estimators at these chosen non-Gaussian innovation distributions, especially when innovation distribution is MTN.

Tables 2.4 and 2.5 summarize the averages of the estimated SE's of m.d. estimators of β_j, $j = 0, 1, 2, 3$ corresponding to four innovations when n is 100 and 200. When the simulation is iterated 1000 times, we estimate SE of m.d. estimators of β_j, $j = 0, 1, 2, 3$ each time as done in Remark 2.1.4. Next, we calculate the average of 1000 SE's for each estimator of β_j and compare it with SE which is reported in Tables 2.2 and 2.3. In order to distinguish between two SE's, we refer to the average of SE's and SE from Tables 2.2 and 2.3 as "estimated

		OLS			GLS			GML			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_0	-0.1873	2.6333	6.9694	-0.1008	2.4235	5.8837	-0.0165	2.6462	7.0027	-0.0404	2.5159	6.3312
	β_1	0.0022	0.056	0.0031	0.0011	0.0498	0.0025	6e-04	0.0505	0.0025	7e-04	0.0521	0.0027
	β_2	0.0029	0.0549	0.003	0.001	0.0494	0.0024	3e-04	0.0497	0.0025	5e-04	0.0521	0.0027
	β_3	0.0022	0.0533	0.0028	0.0015	0.0475	0.0023	8e-04	0.0483	0.0023	9e-04	0.0497	0.0025
	ρ	-0.0707	0.1412	0.0249	-0.0334	0.1521	0.0242	-0.0405	0.1528	0.025	-0.039	0.1568	0.0261
La	β_0	0.0707	1.9605	3.8484	0.053	1.7817	3.1773	0.3357	1.9909	4.0764	0.0638	1.4843	2.2071
	β_1	-9e-04	0.038	0.0014	-8e-04	0.0322	0.001	-0.0036	0.0332	0.0011	-7e-04	0.0268	7e-04
	β_2	-0.001	0.0396	0.0016	2e-04	0.0343	0.0012	-0.0026	0.0353	0.0013	-3e-04	0.0289	8e-04
	β_3	1e-04	0.0387	0.0015	-4e-04	0.0333	0.0011	-0.0032	0.0343	0.0012	-6e-04	0.0273	7e-04
	ρ	-0.0164	0.0659	0.0046	-0.0065	0.0662	0.0044	-0.0085	0.0666	0.0045	-0.0045	0.0565	0.0032
Lo	β_0	0.1716	4.6629	21.7719	0.1422	4.2534	18.1119	0.5701	4.7242	22.6427	0.1358	4.1827	17.5131
	β_1	-0.0036	0.1041	0.0108	-0.0028	0.0896	0.008	-0.0074	0.0922	0.0085	-0.0032	0.0895	0.008
	β_2	-0.0031	0.0936	0.0088	-0.0024	0.0847	0.0072	-0.0069	0.0855	0.0074	-0.002	0.0838	0.007
	β_3	4e-04	0.0995	0.0099	5e-04	0.0866	0.0075	-0.0038	0.0893	0.008	0.0011	0.0865	0.0075
	ρ	-0.0679	0.1318	0.022	-0.0299	0.1416	0.0209	-0.0378	0.1422	0.0217	-0.0341	0.1382	0.0202
M	β_0	0.1204	4.1002	16.826	0.024	3.6986	13.6803	0.4541	4.1075	17.0777	0.0633	3.6242	13.1387
	β_1	-0.0034	0.0834	0.007	-0.0022	0.0738	0.0054	-0.0064	0.0755	0.0057	-0.0022	0.0726	0.0053
	β_2	-0.0041	0.0859	0.0074	-0.0021	0.0759	0.0058	-0.0062	0.0775	0.0061	-0.0033	0.0739	0.0055
	β_3	0.0016	0.0871	0.0076	0.0012	0.0767	0.0059	-0.0024	0.0783	0.0061	0.0019	0.0749	0.0056
	ρ	-0.0755	0.1332	0.0234	-0.0398	0.1417	0.0217	-0.0484	0.1429	0.0228	-0.0443	0.139	0.0213

† N, La, Lo, and M denote normal, Laplace, logistic, and MTN innovations, respectively.

Table 2.1: Bias, SE, and MSE when $n = 50$

		OLS			GLS			GML			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_0	-0.0495	1.8032	3.2539	-0.0222	1.5935	2.5397	0.112	1.8219	3.3318	-0.0097	1.6683	2.7834
	β_1	0	0.0373	0.0014	-0.001	0.0324	0.0011	-0.0021	0.0333	0.0011	-0.001	0.0331	0.0011
	β_2	0.0014	0.0381	0.0015	0.001	0.034	0.0012	-4e-04	0.035	0.0012	9e-04	0.0352	0.0012
	β_3	2e-04	0.0395	0.0016	2e-04	0.0336	0.0011	-0.0011	0.0348	0.0012	-4e-04	0.0346	0.0012
	ρ	-0.0333	0.0943	0.01	-0.0136	0.0962	0.0094	-0.018	0.0967	0.0097	-0.0162	0.0984	0.0099
La	β_0	-0.0399	2.584	6.6788	-0.1123	2.3927	5.7378	0.2255	2.5842	6.7289	-0.1034	2.0057	4.0335
	β_1	4e-04	0.0547	0.003	0.0022	0.0474	0.0023	-0.0012	0.0487	0.0024	0.0016	0.041	0.0017
	β_2	-2e-04	0.0533	0.0028	9e-04	0.0464	0.0022	-0.0029	0.0471	0.0022	0.001	0.0399	0.0016
	β_3	-1e-04	0.0529	0.0028	-3e-04	0.0468	0.0022	-0.0031	0.0475	0.0023	2e-04	0.0399	0.0016
	ρ	-0.0305	0.0888	0.0088	-0.0106	0.091	0.0084	-0.0157	0.0914	0.0086	-0.0097	0.0801	0.0065
Lo	β_0	0.0891	3.2276	10.4251	0.0699	2.9627	8.7823	0.4833	3.257	10.8414	0.122	2.8969	8.4069
	β_1	-0.0049	0.0684	0.0047	-0.0035	0.0594	0.0035	-0.0078	0.0616	0.0039	-0.0038	0.0581	0.0034
	β_2	-0.0011	0.0703	0.0049	-8e-04	0.06	0.0036	-0.005	0.0619	0.0039	-0.001	0.0589	0.0035
	β_3	0.0021	0.0674	0.0046	0.0013	0.0608	0.0037	-0.0027	0.0613	0.0038	0.0011	0.059	0.0035
	ρ	-0.0385	0.0928	0.0101	-0.019	0.0955	0.0095	-0.0227	0.0962	0.0098	-0.0191	0.0937	0.0091
M	β_0	0.0243	2.9132	8.4871	0.0281	2.6472	7.0084	0.3583	2.9276	8.6994	0.0292	2.0247	4.1003
	β_1	2e-04	0.0611	0.0037	-5e-04	0.054	0.0029	-0.0039	0.0553	0.0031	-5e-04	0.0416	0.0017
	β_2	-8e-04	0.0595	0.0035	0	0.0523	0.0027	-0.0035	0.0533	0.0029	-3e-04	0.0394	0.0016
	β_3	0.001	0.0595	0.0035	7e-04	0.0526	0.0028	-0.0025	0.0541	0.0029	0	0.0394	0.0016
	ρ	-0.0264	0.0849	0.0079	-0.0071	0.0879	0.0078	-0.0112	0.0889	0.008	-0.0062	0.0723	0.0053

Table 2.2: Bias, SE, and MSE when $n = 100$

SE" and "simulated SE," respectively. As shown in the Tables 2.4 and 2.5, the right choice of bandwidth h_n results in the small difference between estimated SE and simulated SE of m.d. estimators of β_j's except β_0. For example, the difference between two SE's of each m.d. estimator of β_j-except β_0-ranges from 0 to 0.0021. Furthermore, the optimal bandwidth

		OLS			GLS			GML			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_0	-0.0304	1.3333	1.7785	-0.0147	1.1642	1.3556	0.1174	1.3477	1.83	-0.0229	1.1944	1.4271
	β_1	-4e-04	0.0281	8e-04	-1e-04	0.0235	6e-04	-0.0015	0.0245	6e-04	-3e-04	0.024	6e-04
	β_2	-3e-04	0.0271	7e-04	-6e-04	0.0233	5e-04	-0.0018	0.0242	6e-04	-6e-04	0.0241	6e-04
	β_3	0.0014	0.0265	7e-04	7e-04	0.0227	5e-04	-6e-04	0.0235	6e-04	0.0011	0.0235	6e-04
	ρ	-0.0188	0.0656	0.0047	-0.008	0.0665	0.0045	-0.0103	0.0668	0.0046	-0.0084	0.0678	0.0047
La	β_0	0.0707	1.9605	3.8484	0.053	1.7817	3.1773	0.3357	1.9909	4.0764	0.0638	1.4843	2.2071
	β_1	-9e-04	0.038	0.0014	-8e-04	0.0322	0.001	-0.0036	0.0332	0.0011	-7e-04	0.0268	7e-04
	β_2	-0.001	0.0396	0.0016	2e-04	0.0343	0.0012	-0.0026	0.0353	0.0013	-3e-04	0.0289	8e-04
	β_3	1e-04	0.0387	0.0015	-4e-04	0.0333	0.0011	-0.0032	0.0343	0.0012	-6e-04	0.0273	7e-04
	ρ	-0.0164	0.0659	0.0046	-0.0065	0.0662	0.0044	-0.0085	0.0666	0.0045	-0.0045	0.0565	0.0032
Lo	β_0	0.0904	2.3621	5.5876	0.0612	2.074	4.3051	0.4828	2.4096	6.0394	0.0698	2.0112	4.0497
	β_1	0.0014	0.0471	0.0022	0.0014	0.0406	0.0017	-0.0025	0.042	0.0018	0.0012	0.0398	0.0016
	β_2	-0.0024	0.0495	0.0025	-0.0022	0.0411	0.0017	-0.0063	0.0436	0.0019	-0.0022	0.04	0.0016
	β_3	-0.0011	0.0487	0.0024	-3e-04	0.0421	0.0018	-0.0045	0.0433	0.0019	-4e-04	0.0408	0.0017
	ρ	-0.0154	0.067	0.0047	-0.0054	0.0678	0.0046	-0.0076	0.0678	0.0047	-0.0051	0.0655	0.0043
M	β_0	-0.0195	2.1131	4.4655	-0.0365	1.8834	3.5485	0.2563	2.1421	4.6544	-0.0377	1.3869	1.9248
	β_1	0.0013	0.0434	0.0019	0.0014	0.0368	0.0014	-0.0031	0.0376	0.0014	1e-04	0.0266	7e-04
	β_2	-0.0011	0.0429	0.0018	-7e-04	0.0357	0.0013	-0.0029	0.0386	0.0015	1e-04	0.0278	8e-04
	β_3	0.0012	0.0413	0.0017	0.0013	0.0355	0.0013	-0.0032	0.0385	0.0015	-2e-04	0.0277	8e-04
	ρ	-0.0151	0.0633	0.0042	-0.0053	0.0639	0.0041	-0.0096	0.0638	0.0042	-0.0057	0.0497	0.0025

Table 2.3: Bias, SE, and MSE when $n = 200$

h_n-which yields the estimated SE closest to the simulated SE-is shown to be proportional to n^{-a} where $a > 0$ for all innovations. For normal and Laplace innovations, $h_n \propto n^{-1.32}$; for other two innovations, $h_n \propto n^{-1}$.

		N				La			
		Table 2.2	h_n=0.8	0.9	1	Table 2.2	h_n=0.5	0.6	0.7
$n = 100$	β_0	1.6683	1.5562	1.5752	1.5900	2.0057	1.7543	1.8107	1.8548
	β_1	0.0331	0.0344	0.0349	0.0352	0.0410	0.0387	0.0399	0.0409
	β_2	0.0352	0.0345	0.0349	0.0352	0.0399	0.0387	0.0400	0.0408
	β_3	0.0346	0.0347	0.0351	0.0355	0.0399	0.0386	0.0399	0.0409
		N				La			
		Table 2.3	h_n =0.34	0.36	0.38	Table 2.3	h_n =0.22	0.24	0.26
$n = 200$	β_0	1.1944	1.0706	1.0788	1.0860	1.4843	1.2473	1.2643	1.2794
	β_1	0.0240	0.0238	0.0240	0.0242	0.0268	0.0277	0.0280	0.0284
	β_2	0.0241	0.0239	0.0241	0.0242	0.0289	0.0277	0.0281	0.0284
	β_3	0.0235	0.0238	0.0240	0.0241	0.0273	0.0277	0.0281	0.0285

Table 2.4: Averages of SE's of $\widehat{\beta}_j$'s with normal and Laplace innovations

		Lo				M			
		Table 2.2	h_n =1.3	1.4	1.5	Table 2.2	h_n =0.5	0.6	0.7
$n = 100$	β_0	2.8969	2.6129	2.6337	2.6522	2.0247	1.7396	1.7955	1.8380
	β_1	0.0581	0.0581	0.0585	0.0589	0.0416	0.0386	0.0399	0.0408
	β_2	0.0589	0.0579	0.0584	0.0588	0.0394	0.0387	0.0399	0.0409
	β_3	0.0590	0.0580	0.0584	0.0588	0.0394	0.0387	0.0399	0.0409
		Lo				M			
		Table 2.3	h_n =0.66	0.68	0.7	Table 2.3	h_n =0.28	0.3	0.32
$n = 200$	β_0	2.0112	1.7955	1.8081	1.8198	1.3869	1.2470	1.2647	1.2803
	β_1	0.0398	0.0397	0.0400	0.0403	0.0266	0.0275	0.0279	0.0282
	β_2	0.0400	0.0397	0.0399	0.0402	0.0278	0.0274	0.0278	0.0282
	β_3	0.0408	0.0399	0.0402	0.0404	0.0277	0.0275	0.0279	0.0282

Table 2.5: Averages of SE's of $\widehat{\beta}_j$'s with logistic and MTN innovations

2.1.3.2 Global temperature

In this section, we apply m.d. estimation method to real data: one hundred yearly data of global temperature. The data can be obtained at International Panel on Climate Change (IPCC) website: `http://www.ipcc.ch/`. The IPCC website provides with freely accessible database. The left panel of Figure 2.1 shows the series of global temperature over the past

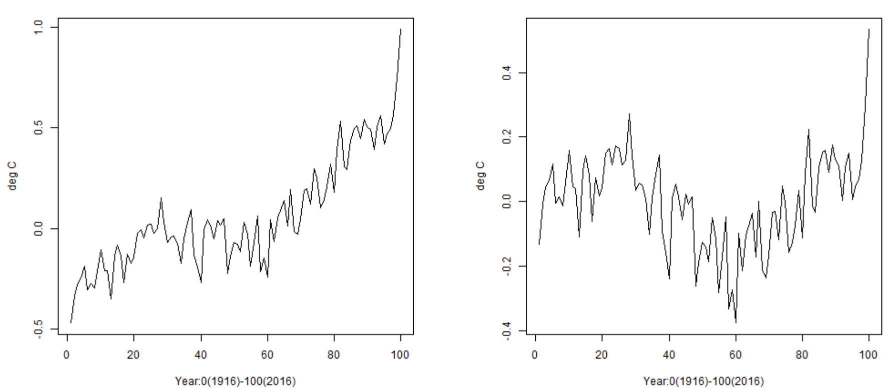

Figure 2.1: The IPCC global temperature (left) and OLS-fitted errors (right) from 1916 to 2016

100 years. As shown in the figure, the upward linear time trend is observed, which may mean there is an evidence of global warming. Therefore, we consider a linear regression model with

$t_0 = 1916$

$$y_t = \beta_0 + \beta_1(t - t_0) + \varepsilon_t, \quad t = 1917, 1918, ..., 2016, \tag{2.20}$$

where y_t is global temperature in year t. The OLS estimates of the β_0 and β_1 are -0.3432 and 0.0079, respectively. Next, we define OLS-fitted error

$$\widehat{\varepsilon}_{ols,t} = y_t - (-0.03432 + 0.0079(t - t_0)).$$

The right panel of Figure 2.1 shows the series of OLS-fitted errors; it is not difficult to see there exist a correlation between the lagged errors. To carry out the correlation analysis in the sequel, we check sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the series of the interest. Figure 2.2 shows the sample ACF and PACF of the $\widehat{\varepsilon}_{ols,t}$'s of the first 15 lags. First, the ACF exhibits a mixture of the slow exponential decay and damped sinusoidal component. Second, the series of $\widehat{\varepsilon}_{ols,t}$'s displays significant PACF coefficients up to the fourth lag. These two facts suggest that true underlying error model is likely to be AR model of order 4. Bearing these facts in mind, we will fit AR model of various orders to the errors for the purpose of specifying its unknown order. Under the assumption that ε_t obeys AR model (2.1) of the unknown order, we estimate both the regression and AR parameters of model (2.20); for the orders of AR model, 1, 2, 3, and 4 are tried. Table 2.6 summarizes m.d. and GML estimation results; CO estimation is not presented since `orcutt` works for AR errors of order 1 only. Both the m.d. and the GML estimation methods give similar estimates of β and ρ except when AR model of order 2 is fitted. As summarized in the table, the linear trend of the m.d. estimation method becomes steeper as AR model of higher order is fitted; the slopes corresponding to the AR order 1, 2, 3, and 4 are 0.081, 0.084, 0.087, and 0.096, respectively. The GML estimation method

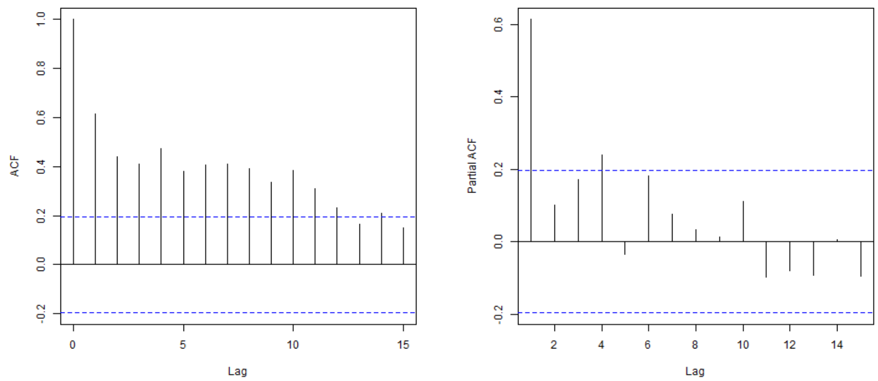

Figure 2.2: Lag vs. sample ACF and PACF of OLS fitted errors

exhibits the similar result.

It was observed above that when innovations ($n = 100$) are normal, Laplace, logistic, and MTN, the estimated SE which is the closest to the simulated SE can be obtained with bandwidths of 0.9, 0.6, 1.4, and 0.6, respectively; see, e.g., Tables 2.4 and 2.5. With the assumption that the error obeys AR models of orders 1, 2, 3, and 4, the estimated SE of the m.d. estimator of β_1 and 95% confidence interval (CI) of β_1 corresponding to these bandwidths are reported in Table 2.7. The values in the table are expressed in thousandths: e.g., SE corresponding to $h_n = 0.6$ and AR order of 1 is 1.4559×10^{-3}. When AR model of the same order is considered, the SE's corresponding to the different bandwidths differ only by millionths. As a result, 95% CI's for β_1 are almost the same. But as AR model of the different order is considered, quite different SE's are reported; so are different CI's. All CI's are well-separated from zero, and hence we can conclude $\beta_1 \neq 0$, which implies that there is a strong evidence of the global warming.

Next, we proceed to check whether the assumption of independence of the innovations is met in order to conclude AR model of which order provides the best fit. To that end, we

Order of AR error	m.d.		GML	
	β	ρ	β	ρ
1	(-0.3207, 0.0081)'	0.6842	(-0.3540, 0.0085)'	0.7146
2	(-0.3390, 0.0084)'	(0.6248, 0.0946)'	(-0.4430, 0.0099)'	(0.6500, 0.1378)'
3	(-0.3540, 0.0087)'	(0.5992, 0.0252, 0.1682)'	(-0.3574, 0.0089)'	(0.6152, 0.0216, 0.1897)'
4	(-0.4025, 0.0096)'	(0.5298, -0.0389, 0.0420, 0.3103)'	(-0.3268, 0.0091)'	(0.5584, 0.0076, 0.0571, 0.2757)'

Table 2.6: M.d. and GML estimates of β and ρ

Order of AR error	$h_n = 0.6$		$h_n = 0.9$		$h_n = 1.4$	
	SE	CI	SE	CI	SE	CI
1	1.4559	(5.2728, 10.9800)	1.4556	(5.2734, 10.9794)	1.4554	(5.2738, 10.9790)
2	1.5202	(5.4452, 11.4045)	1.5188	(5.4479, 11.4018)	1.5188	(5.4478, 11.4019)
3	1.6190	(5.6124, 11.9589)	1.6176	(5.6150, 11.9562)	1.6176	(5.6150, 11.9563)
4	1.9522	(5.8703, 13.5232)	1.9485	(5.8776, 13.5159)	1.9485	(5.8776, 13.5159)

Table 2.7: SE of m.d. estimator of β_1 and 95% CI

obtain fitted innovations of AR model by solving

$$\widehat{\xi}_{md,t} = \widehat{\varepsilon}_{md,t} - \widehat{Z}_{md,t}\ \widehat{\rho}_{md}.$$

where $\widehat{\varepsilon}_{md,t}$'s are the fitted errors after the two-stage m.d. estimation, $\widehat{Z}_{md,t}$ is a vector of the lagged $\widehat{\varepsilon}_{md,t}$'s, and $\widehat{\rho}_{md}$ is the two-stage m.d. estimator of the AR parameter. Figure 2.3 shows the sample ACF's of the fitted innovations corresponding to the AR models of order 1 (top left), 2 (top right), 3 (bottom left), and 4 (bottom right). It can be seen that the independence assumption becomes viable as the order increases; the dependence seems to disappear when AR order is 4 as expected from the ACF and the PACF of OLS-fitted errors. When AR model of order 1 is fitted, there are two spikes out of the approximate 95% confidence band [-0.2, 0.2]. However, there is no spike out of the band if AR model of order 4 is fitted to the errors. Therefore, we conclude that AR model of order 4 provides better fit than AR models of other orders: Figure 2.4 - the sample PACF's of the fitted innovations - confirms the conclusion.

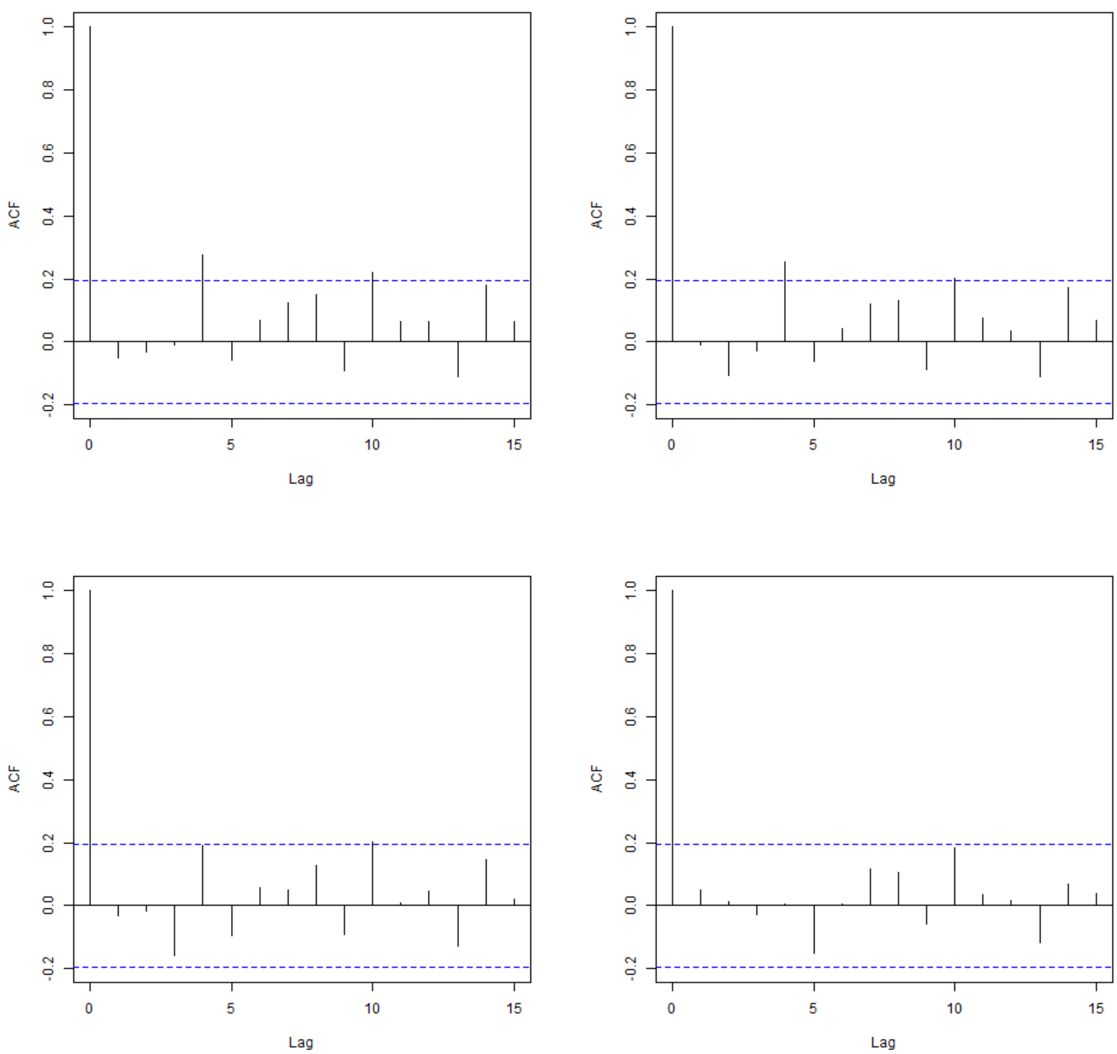

Figure 2.3: Lag vs. sample ACF with AR models of order 1, 2, 3, and 4.

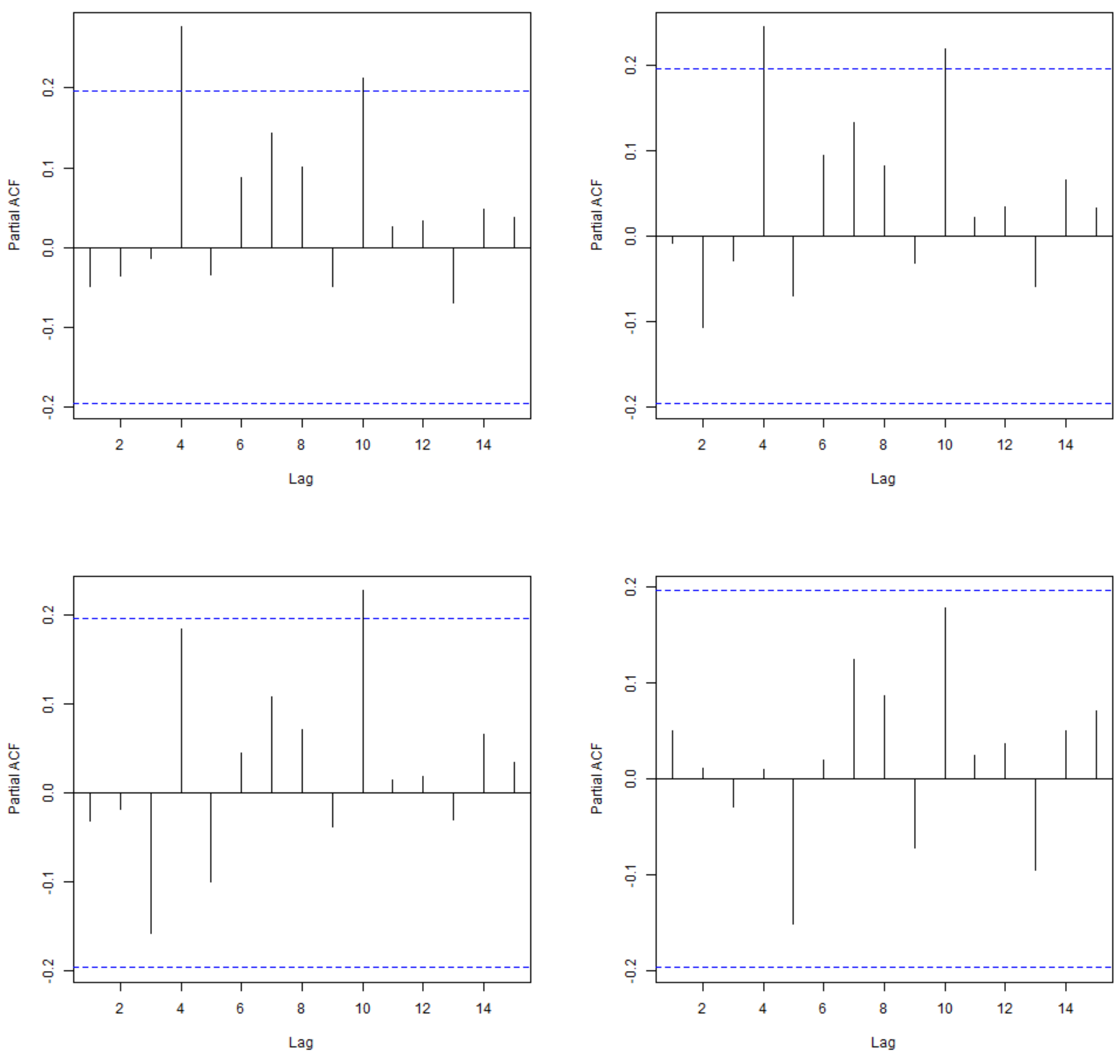

Figure 2.4: Lag vs. sample PACF with AR models of order 1, 2, 3, and 4.

2.2 Panel regression model

In this section, we apply the m.d. estimation method to panel regression model and investigate the asymptotic properties of m.d. estimators of the regression parameters. We discuss only the case where the number of cross section units is relatively larger than the length of time periods, which is called *short panel data*. Section 2.2.2 will discuss the asymptotic properties when the number of cross section units goes to infinity while time periods are kept finite .

2.2.1 Random effect model

Recall the panel regression model (1.2). Define

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix}_{T\times 1}, \quad \mathbf{X}_i = \begin{bmatrix} x_{i1}^1 & x_{i1}^2 & \cdots & x_{i1}^p \\ x_{i2}^1 & x_{i2}^2 & \cdots & x_{i2}^p \\ \vdots & \vdots & \ddots & \vdots \\ x_{iT}^1 & x_{iT}^2 & \cdots & x_{iT}^p \end{bmatrix}_{T\times p}, \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}_{T\times 1},$$

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_n \end{bmatrix}_{nT\times 1}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{bmatrix}_{nT\times p}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}_{nT\times 1}.$$

Then standard panel regression model can be expressed as

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon}$$

where $\beta = (\beta_1, \beta_2, ..., \beta_p)' \in \mathbb{R}^p$. As seen in (1.2), the error can be rewritten as sum of time-invariant individual effect and remainder disturbance: $\varepsilon_{it} = \gamma_i + \nu_{it}$. The panel regression model which assumes that time-invariant individual effect γ_i is uncorrelated with $\boldsymbol{x}_{it}$ is called *random effect model*. Another classical assumption on the error in the random effect model is that correlation over time does not change across the panel, i.e., $E(\varepsilon_{it}\varepsilon_{is}) = \sigma_\nu^2 + \sigma_\gamma^2$. However, errors over cross-sections are independent, i.e., $E(\varepsilon_{it}\varepsilon_{js}) = 0$ for all $i \neq j$. Assume that

$$
\begin{aligned}
E(\boldsymbol{\nu}) &= 0, \qquad E(\boldsymbol{\nu}\boldsymbol{\nu}') = \sigma_\nu^2 \boldsymbol{I}_{nT \times nT}, \\
E(\gamma_i\gamma_j) &= 0, \qquad E(\gamma_i\gamma_i) = \sigma_\gamma^2, \\
E(\gamma_i\nu_{jt}) &= 0, \qquad E(\gamma_i) = 0, \qquad 1 \leq i, j \leq n,\ 1 \leq t \leq T
\end{aligned}
$$

where $\boldsymbol{\nu} = (\nu_{11}, ..., \nu_{nT})' \in \mathbb{R}^{nT}$. Let $\boldsymbol{\Sigma}$ denote $E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')$. Then, we have

$$
\boldsymbol{\Sigma} = \sigma_\nu^2 \boldsymbol{I}_{T \times T} + \sigma_\gamma^2 \boldsymbol{i}\boldsymbol{i}' = \begin{bmatrix} \sigma_\nu^2 + \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\nu^2 + \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\nu^2 + \sigma_\gamma^2 \end{bmatrix}.
$$

where $\boldsymbol{i}$ is a $T \times 1$ vector of ones. Let $\boldsymbol{\Omega}$ be the covariance of the error for all the observations, i.e., $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$. Then,

$$
\boldsymbol{\Omega} = \boldsymbol{I}_{n \times n} \otimes \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma} & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Sigma} \end{bmatrix}
$$

Before we investigate the asymptotic behavior of the m.d. estimators, we need assumptions on the structure of $\mathbf{X}$, similar to those in the Section 2.1. Hence, we replace X and x_i with $\mathbf{X}$ and $\boldsymbol{x}_{it}$ in (2.2). Similarly, let $\mathbf{A} := (\mathbf{X}'\mathbf{X})^{-1/2}$ and $\mathbf{D} := \mathbf{XA}$. Note that $\mathbf{D}$ can be partitioned into $n \times 1$ blocks as follows:

$$\mathbf{D} = \begin{bmatrix} \mathbf{X}_1\mathbf{A} \\ \vdots \\ \mathbf{X}_i\mathbf{A} \\ \vdots \\ \mathbf{X}_n\mathbf{A} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_i \\ \vdots \\ \mathbf{D}_n \end{bmatrix}$$

where

$$\mathbf{D}_i := \mathbf{X}_i\mathbf{A} = \begin{bmatrix} d_{i11} & \cdots & d_{i1k} & \cdots & d_{i1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{it1} & \ddots & d_{itk} & \ddots & d_{itp} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{iT1} & \cdots & d_{iTk} & \cdots & d_{iTp} \end{bmatrix}$$

for $1 \leq i \leq n$, $1 \leq t \leq T$, and $1 \leq k \leq p$. Note that for all $1 \leq k \leq p$

$$\sum_{i=1}^{n} \sum_{t=1}^{T} d_{itk}^2 = 1. \tag{2.21}$$

Similar to what was done in the Section 2.1.1 for asymptotic uniform quadraticity, we define the distance function

$$\begin{aligned} U_k(y,\boldsymbol{b}) &:= \sum_{i=1}^{n}\sum_{t=1}^{T} d_{itk}\Big\{ I\big(Y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}\le y\big) - I\big(-Y_{it}+\boldsymbol{x}_{it}'\boldsymbol{b}< y\big)\Big\}, \\ \mathbf{U}(y,\boldsymbol{b}) &:= (U_1(y,\boldsymbol{b}),...,U_p(y,\boldsymbol{b}))', \quad y\in\mathbb{R}, \\ L(\boldsymbol{b}) &:= \int \|\mathbf{U}(y,\boldsymbol{b})\|^2\, dH(y), \quad \boldsymbol{b}\in\mathbb{R}^p, \end{aligned}$$

where H is a $\sigma-$finite measure on $\mathbb{R}$ and symmetric around 0, i.e., $dH(-x) = -dH(x)$, $x \in \mathbb{R}$. Subsequently, define $\widehat{\boldsymbol{\beta}}$ as

$$L(\widehat{\boldsymbol{\beta}}) := \inf_{\boldsymbol{b}} L(\boldsymbol{b}). \tag{2.22}$$

For the analogue of Q in the Theorem 2.1.2, define for $\boldsymbol{b}\in\mathbb{R}^p$

$$\mathcal{L}(\boldsymbol{b}) := \int \|\mathbf{U}(y,\boldsymbol{\beta}) + 2\mathbf{A}^{-1}(\boldsymbol{b}-\boldsymbol{\beta})f(y)\|^2\, dH(y).$$

2.2.2 Asymptotic distribution of $\widehat{\boldsymbol{\beta}}$

In this section we derive the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ under the current setup. The basic method of the proof is similar to that of the Section 2.1.1. We first state the necessary assumptions. Let F denote the d.f. of ε.

(C.1) The matrix $\mathbf{X}'\mathbf{X}$ is nonsingular and, with $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1/2}$, satisfies

$$\limsup_{n\to\infty} n \max_{1\le i\le n}\max_{1\le t\le T} \|\mathbf{A}\boldsymbol{x}_{it}\|^2 < \infty.$$

(C.2) The integrating measure H is $\sigma-$finite and symmetric around 0, and

$$\int_0^\infty (1-F)dH < \infty.$$

(C.3) For any real sequences $\{a_n\}$, $\{b_n\}$, $b_n - a_n \to 0$,

$$\limsup_{n\to\infty} \int_{a_n}^{b_n} \int f(y+x)dH(y)dx = 0.$$

(C.4) For $a \in \mathbb{R}$, define $a^+ := \max(a, 0)$, $a^- := a^+ - a$. Let $\theta_{it} := \|\mathbf{A}\boldsymbol{x}_{it}\|$. For all $\mathbf{u} \in \mathbb{R}^p$, $\|\mathbf{u}\| \le b$, for all $\delta > 0$, and for all $1 \le k \le p$,

$$\limsup_{n\to\infty} \int \Big[\sum_{i=1}^{n}\sum_{t=1}^{T} d_{itk}^{\pm}\{F(y+\mathbf{u}'\mathbf{A}\boldsymbol{x}_{it}+\delta\theta_{it}) - F(y+\mathbf{u}'\mathbf{A}\boldsymbol{x}_{it}-\delta\theta_{it})\}\Big]^2 dH(y) \le c\delta^2,$$

where c does not depend on u and δ.

(C.5) Let $\boldsymbol{e}_k \in \mathbb{R}^p$ be elementary vector whose kth coordinate is 1. For each $\mathbf{u} \in \mathbb{R}^p$ and all $1 \le k \le p+1$,

$$\int \left[\sum_{i=1}^{n}\sum_{t=1}^{T} d_{itk}\{F(y+\mathbf{u}'\mathbf{A}\boldsymbol{x}_{it}) - F(y)\} - \mathbf{u}'\boldsymbol{e}_k\, f(y)\right]^2 dH(y) = o(1).$$

(C.6) F has a continuous density f with respect to the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$.

(C.7) $0 < \int_0^\infty f^r dH < \infty$, $r = 1, 2$.

Now, we are ready to state the result.

Theorem 2.2.1. *Assume $\{Y_{it},\ 1 \le i \le n,\ 1 \le t \le T\}$ be in the model (1.2). Assume that*

(C.1)-(C.7) hold. Then for any $0 < c < \infty$

$$E \sup_{\|\mathbf{A}^{-1}(\boldsymbol{b}-\boldsymbol{\beta})\| \le c} \|L(\boldsymbol{b}) - \mathcal{L}(\boldsymbol{b})\| = o(1). \tag{2.23}$$

Corollary 2.2.1. *Suppose that the assumptions of Theorem 2.2.1 hold. Then for any* $\varepsilon > 0$, $0 < z < \infty$ *there exists an* N_ε*, and* $0 < c < \infty$ *(depending on* ε *and* M*) such that*

$$P\left(\inf_{\|\mathbf{A}^{-1}(\boldsymbol{b}-\boldsymbol{\beta})\| \ge c} L(\boldsymbol{b}) \ge z \right) \ge 1 - \varepsilon, \quad \forall n \ge N_\varepsilon. \tag{2.24}$$

Proofs of Theorem 2.2.1 and Corollary 2.2.1 are very similar to those of Theorem 2.1.2 and Corollary 2.1.1; however, the proofs are much simpler since the correlation of errors between different cross section units is zero. Therefore, they are not presented here.

Theorem 2.2.2. *Under the assumptions of Theorem 2.2.1,*

$$\mathbf{A}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\{2|f|_H^2\}^{-1} \int f(y)\mathbf{U}(y, \boldsymbol{\beta})\, dH(y) + o_p(1). \tag{2.25}$$

Proof. By Theorem 2.2.1 and Corollary 2.2.1, $L(\boldsymbol{b})$ is uniformly locally asymptotically quadratic and

$$\|\mathbf{A}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\| = O_p(1).$$

Therefore, the proof is straightforward, as in i.i.d. case illustrated in Koul (2002).

Recall ψ in the (2.10) and define $\boldsymbol{z}_n := (z_{n1}, ..., z_{np})' \in \mathbb{R}^p$ where

$$z_{nk} := \sum_{i=1}^{n} \sum_{t=1}^{T} d_{itk} \psi(\varepsilon_{it}), \qquad k = 1, 2, ...p.$$

Let $\boldsymbol{\Sigma}_n$ denote covariance matrix of $\boldsymbol{z}_n$ and $\boldsymbol{\Sigma}_{nkh}$ denote its (k, h) entry. Note that

$$\begin{aligned}
\boldsymbol{\Sigma}_{nkh} &= E\left[\sum_{i=1}^{n}\sum_{t=1}^{T} d_{itk}\psi(\varepsilon_{it})\sum_{j=1}^{n}\sum_{s=1}^{T} d_{jsh}\psi(\varepsilon_{js})\right] \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} d_{itk}d_{isk}E\psi(\varepsilon_{it})\psi(\varepsilon_{is}) \\
&\leq n \max_{1\leq i\leq n}\max_{1\leq t\leq T} d_{itk}^2\ T^2\|\psi(\varepsilon_{11})\|_2^2 \\
&< \infty
\end{aligned}$$

The second equality follows from the independence of ε_{it} and ε_{js} for $i \neq j$; the first inequality follows from (A.1) and (A.7). Let $\mathbf{a}_k$ denote kth column vector of $\mathbf{A}$. Rewrite

$$\boldsymbol{\Sigma}_{nkh} = \mathbf{a}_k'\left[\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'E\psi(\varepsilon_{it})\psi(\varepsilon_{is})\right\}\right]\mathbf{a}_h. \tag{2.26}$$

Consider $\lambda \in \mathbb{R}^p$. Let $\zeta_i := \sum_{t=1}^{T}\lambda'\boldsymbol{\Sigma}_n^{-1/2}\mathbf{A}\boldsymbol{x}_{it}\psi(\varepsilon_{it})$. By symmetry of errors, $E\zeta_i = 0$. Also note that

$$\begin{aligned}
\sum_{i=1}^{n} E\zeta_i^2 &= \sum_{i=1}^{n}\left[E\sum_{t=1}^{T}\sum_{s=1}^{T}\{\lambda'\boldsymbol{\Sigma}_n^{-1/2}\mathbf{A}\boldsymbol{x}_{it}\psi(\varepsilon_{it})\}\{\lambda'\boldsymbol{\Sigma}_n^{-1/2}\mathbf{A}\boldsymbol{x}_{is}\psi(\varepsilon_{is})\}\right] \\
&= \lambda'\boldsymbol{\Sigma}_n^{-1/2}\mathbf{A}\left[\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}\sum_{s=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{is}'E\psi(\varepsilon_{it})\psi(\varepsilon_{is})\right\}\right]\mathbf{A}\boldsymbol{\Sigma}_n^{-1/2}\lambda \\
&= \|\lambda\|^2
\end{aligned}$$

where the last equality follows from (2.26). Recall Lyapunov's Theorem.

Theorem 2.2.3. *(**Lyapunov's Theorem**) Suppose $\{X_1, X_2, ...\}$ is a sequence of independent random variables, each with finite expected value μ_i and variance σ_i^2. Define*

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2.$$

If there is some $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} E|X_i - \mu_i|^{2+\delta} = 0, \tag{2.27}$$

then

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \to_D N(0,1).$$

As a consequence of Theorem 2.2.3, we obtain the following corollary.

Corollary 2.2.2. *Assume $\boldsymbol{\Sigma}_n$ is positive definite for all $n \geq p$. In addition, assume that*

$$\sup_{u\in\mathbb{R}^p, \|u\|=1} u'\boldsymbol{\Sigma}_n^{-1} u = O(1). \tag{2.28}$$

Then

$$\boldsymbol{\Sigma}_n^{-1/2} \boldsymbol{z}_n \to_D N(0, \boldsymbol{I}_{p\times p}), \tag{2.29}$$

where $\boldsymbol{I}_{p\times p}$ is the $p \times p$ identity matrix.

Proof. To prove (2.29), it suffices to show that for any $\lambda \in \mathbb{R}^p$, $\lambda'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{z}_n$ is asymptotically normally distributed. Note that $\lambda'\boldsymbol{\Sigma}_n^{-1/2}\boldsymbol{z}_n = \sum_{i=1}^{n} \zeta_i$ where $\{\zeta_1, \zeta_2, ..., \zeta_n\}$ is a sequence of independent random variables. Observe that (C.1), (C.7), and (2.28) imply

$$\begin{aligned} |\zeta_i| &\leq \left\| \lambda' \boldsymbol{\Sigma}_n^{-1/2} \right\| \left\| \sum_{t=1}^{T} \mathbf{A} \boldsymbol{x}_{it} \psi(\varepsilon_{it}) \right\| \\ &\leq (\lambda' \boldsymbol{\Sigma}_n \lambda)^{1/2} \, T \max_{1 \leq i \leq n} \max_{1 \leq t \leq T} \| \mathbf{A} \boldsymbol{x}_{it} \| \, |\psi(\varepsilon_{it})| \\ &= O(n^{-1/2}). \end{aligned}$$

Consequently, for any $\delta > 0$, we have

$$\sum_{i=1}^{n} E|\zeta_i|^{2+\delta} = O(n^{-\delta/2}) \sum_{i=1}^{n} E\zeta_i^2 \to 0.$$

Therefore (2.29) follows after the direct application of Theorem 2.2.3.

Corollary 2.2.3. *Assume that (C.1)-(C.7) hold. Then*

$$\boldsymbol{\Sigma}_n^{-1/2} \mathbf{A}^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_D \{2|f|_H^2\}^{-1} N(0, \boldsymbol{I}_{p \times p}) \tag{2.30}$$

Proof. Note that

$$\boldsymbol{z}_n = \int f(y) \mathbf{U}(y, \boldsymbol{\beta}) \, dH(y),$$

Hence (2.30) directly follows from Corollary 2.2.2.

2.2.3 Other panel data estimators

In this section, we introduce other estimators commonly used in the literature of econometrics. All are variations of OLS estimators. Recall the model (1.2). When we average response variables, design variables, and errors over all time periods and subtract the averages from the original variables, we then obtain *between model*

$$\bar{Y}_i = \bar{\boldsymbol{x}}_i' \beta + \bar{\varepsilon}_i, \qquad i = 1, 2, ..., n,$$

where $\bar{Y}_i = T^{-1} \sum_{t=1}^{T} Y_{it}$, $\bar{\boldsymbol{x}}_i = T^{-1} \sum_{t=1}^{T} \boldsymbol{x}_{it}$, and $\bar{\varepsilon}_i = T^{-1} \sum_{t=1}^{T} \varepsilon_{it}$. The *between estimator* is nothing but OLS estimator of regression parameter of the between model. The panel data has a special feature-that the variables vary with time-which the merits of panel data originates in. As seen above, the between model loses this feature by averaging the original variables over all time periods. To redress this issue, another panel regression model which is called *within model* is proposed:

$$Y_{it} - \bar{Y}_i = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i), \qquad i = 1, 2, ..., n;\ t = 1, 2, ..., T.$$

The *within estimator* is the OLS estimator obtained from the within model. Note that the time-invariant individual effect γ_i does not exist in the within model after the average of the error is subtracted from the original error. Another well-celebrated panel data estimator is *random effect estimator*. The random effect estimator is feasible GLS estimator of the model (1.2); it can be obtained by applying OLS estimation to the following model

$$Y_{it} - \widehat{\lambda} \bar{Y}_i = (\boldsymbol{x}_{it} - \widehat{\lambda} \bar{\boldsymbol{x}}_i)' \beta + (\varepsilon_{it} - \widehat{\lambda} \bar{\varepsilon}_i),$$

where $\widehat{\lambda}$ is consistent for $\lambda := 1 - \sigma_\nu^2 / \sqrt{\sigma_\nu^2 + T\sigma_\gamma^2}$. Note that OLS and within estimator are special case of the random effect estimators corresponding to $\widehat{\lambda} = 0$ and $\widehat{\lambda} = 1$, respectively.

In order to obtain the m.d. estimators in the next section, we apply the m.d. estimation method to the within model instead of the model (1.2). For the case of the regression model with i.i.d. non-Gaussian errors, Kim (2016) empirically showed the m.d. estimators outperform OLS estimators. Note that the extent of the dependence of errors in the panel regression model is between those of the regression with i.i.d. and AR errors. Since the superiority of the m.d. estimators to the OLS family is demonstrated, they are expected to still remain superior to the panel data estimators.

2.2.4 Simulation

In this section we present a simulation study corresponding to sixteen pairs of symmetric individual effect and remainder disturbance. Both individual effect and remainder disturbance are generated from normal, Laplace, logistic, and MTN. For example, Table 2.8 reports the findings corresponding to the normal individual effect with normal, Laplace, logistic, and MTN remainder disturbances when $n = 10$ and $T = 5$. The simulation setup is very similar to one in the Section 2.1.3. The bias, SE, and MSE of the estimators of regression parameters are reported in the tables; for easy comparison purpose, we analyze the findings and evaluate the performance of estimators in terms of MSE since those which display the least MSE also approximately display the least bias and SE.

In case of the normal individual effect, the within estimators show the best performance regardless of remainder disturbance; the random effect estimators display almost the same performance as the within estimators. The m.d. estimators follows the within and the random effect estimators, and, not surprisingly, OLS estimators are the worst. For the non-Gaussian individual effects, the m.d. estimators outperform other panel data estimators as expected. The simulation study in this section exhibits the similar pattern to one in the Section 2.1.3; the superiority of the m.d. estimators to other estimators is salient especially when the model contains the MTN individual effect.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
	β_1	0.0006	0.0534	0.0028	0.0019	0.0388	0.0015	0.0018	0.0388	0.0015	0.0018	0.0398	0.0016
N	β_2	-0.0017	0.0550	0.0030	0.0002	0.0389	0.0015	0.0002	0.0395	0.0016	0.0005	0.0400	0.0016
	β_3	0.0031	0.0504	0.0026	0.0012	0.0383	0.0015	0.0016	0.0385	0.0015	0.0014	0.0395	0.0016
	β_1	-0.0034	0.0657	0.0043	7e-04	0.0361	0.0013	4e-04	0.0361	0.0013	8e-04	0.0371	0.0014
La	β_2	0.0043	0.072	0.0052	0.0014	0.0388	0.0015	0.0014	0.0387	0.0015	0.0013	0.0398	0.0016
	β_3	-0.004	0.0662	0.0044	-4e-04	0.0373	0.0014	-7e-04	0.0373	0.0014	-9e-04	0.0384	0.0015
	β_1	0.001	0.0887	0.0079	3e-04	0.0379	0.0014	6e-04	0.0381	0.0015	4e-04	0.039	0.0015
Lo	β_2	0	0.0873	0.0076	0.0013	0.0393	0.0015	0.0013	0.0395	0.0016	0.0011	0.0402	0.0016
	β_3	0.0015	0.0874	0.0076	0	0.0386	0.0015	1e-04	0.0387	0.0015	-3e-04	0.0402	0.0016
	β_1	0.0015	0.0756	0.0057	-9e-04	0.0383	0.0015	-6e-04	0.0384	0.0015	-9e-04	0.0395	0.0016
M	β_2	0.0021	0.077	0.0059	-3e-04	0.0376	0.0014	0	0.0374	0.0014	-4e-04	0.0384	0.0015
	β_3	8e-04	0.0758	0.0058	0.0018	0.0381	0.0015	0.0013	0.0383	0.0015	0.0018	0.0391	0.0015

† Within and RE denote the within and random effect estimators, respectively.

Table 2.8: Bias, SE, and MSE of estimators: γ is normal with $n = 10$ and $T = 5$.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
	β_1	6e-04	0.0671	0.0045	0.0017	0.0673	0.0045	0.0012	0.0674	0.0045	0.0016	0.0669	0.0045
N	β_2	0.0025	0.0663	0.0044	4e-04	0.0708	0.005	0.0015	0.0689	0.0048	7e-04	0.0699	0.0049
	β_3	-0.0012	0.0661	0.0044	-0.0026	0.0664	0.0044	-0.002	0.068	0.0046	-0.0028	0.0655	0.0043
	β_1	-0.0138	0.1405	0.0199	-0.0148	0.1386	0.0194	-0.0149	0.1376	0.0192	-0.0153	0.1383	0.0194
La	β_2	0.0487	0.3609	0.1326	0.0461	0.359	0.131	0.047	0.3588	0.131	0.0468	0.3588	0.1309
	β_3	-0.0339	0.2484	0.0629	-0.0297	0.2454	0.0611	-0.0308	0.2457	0.0613	-0.0308	0.2452	0.0611
	β_1	0.0055	0.0906	0.0082	0.002	0.066	0.0044	0.0023	0.0669	0.0045	0.0013	0.0659	0.0043
Lo	β_2	-9e-04	0.0927	0.0086	0.0017	0.0668	0.0045	0.0013	0.0672	0.0045	0.0013	0.0665	0.0044
	β_3	-0.0061	0.0917	0.0085	-9e-04	0.066	0.0044	-0.0012	0.0665	0.0044	-4e-04	0.0666	0.0044
	β_1	-0.0032	0.089	0.0079	-5e-04	0.0701	0.0049	-0.0011	0.0707	0.005	-8e-04	0.0689	0.0047
M	β_2	9e-04	0.0964	0.0093	0.0039	0.0681	0.0047	0.0042	0.0691	0.0048	0.0039	0.068	0.0046
	β_3	-0.0012	0.0909	0.0083	-5e-04	0.0709	0.005	-5e-04	0.0709	0.005	0	0.0712	0.0051

Table 2.9: Bias, SE, and MSE of estimators: γ is logistic with $n = 10$ and $T = 5$.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
	β_1	0.0017	0.0592	0.0035	-5e-04	0.0541	0.0029	-6e-04	0.0542	0.0029	4e-04	0.0506	0.0026
N	β_2	5e-04	0.059	0.0035	-1e-04	0.051	0.0026	-1e-04	0.0504	0.0025	-4e-04	0.0488	0.0024
	β_3	0	0.0585	0.0034	-8e-04	0.0518	0.0027	-6e-04	0.0526	0.0028	-0.0011	0.0488	0.0024
	β_1	-0.0109	0.1236	0.0154	-0.0106	0.1137	0.0131	-0.0099	0.1138	0.0131	-0.0103	0.1119	0.0126
La	β_2	0.0324	0.3074	0.0955	0.0302	0.3033	0.0929	0.0305	0.3032	0.0928	0.0311	0.3027	0.0926
	β_3	-0.0174	0.2126	0.0455	-0.02	0.2056	0.0427	-0.0189	0.2062	0.0429	-0.0196	0.2049	0.0424
	β_1	-0.0138	0.1405	0.0199	-0.0148	0.1386	0.0194	-0.0149	0.1376	0.0192	-0.0153	0.1383	0.0194
Lo	β_2	0.0487	0.3609	0.1326	0.0461	0.359	0.131	0.047	0.3588	0.131	0.0468	0.3588	0.1309
	β_3	-0.0339	0.2484	0.0629	-0.0297	0.2454	0.0611	-0.0308	0.2457	0.0613	-0.0308	0.2452	0.0611
	β_1	-0.005	0.0808	0.0066	-0.0025	0.0528	0.0028	-0.0029	0.0519	0.0027	-0.0015	0.0487	0.0024
M	β_2	0.0039	0.0825	0.0068	0.0019	0.0561	0.0032	0.0021	0.0561	0.0031	0.0019	0.0522	0.0027
	β_3	0	0.0851	0.0073	-0.002	0.0541	0.0029	-0.0014	0.0544	0.003	-0.0019	0.0501	0.0025

Table 2.10: Bias, SE, and MSE of estimators: γ is Laplace with $n = 10$ and $T = 5$.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_1	0.0038	0.0749	0.0056	0.0011	0.0823	0.0068	0.0021	0.0832	0.0069	0.0013	0.0587	0.0034
	β_2	-9e-04	0.077	0.0059	2e-04	0.0771	0.0059	1e-04	0.0779	0.0061	-8e-04	0.0545	0.003
	β_3	-9e-04	0.0736	0.0054	-0.0035	0.0818	0.0067	-0.0031	0.08	0.0064	-3e-04	0.0572	0.0033
La	β_1	-0.0069	0.0972	0.0095	-0.0036	0.0909	0.0083	-0.0041	0.0895	0.008	-0.0027	0.0713	0.0051
	β_2	0.0067	0.1601	0.0257	0.0057	0.1561	0.0244	0.0067	0.1563	0.0245	0.0046	0.1456	0.0212
	β_3	-0.0031	0.1234	0.0152	-0.0031	0.1179	0.0139	-0.0028	0.1173	0.0138	-0.0056	0.1049	0.011
Lo	β_1	-0.0141	0.1369	0.0189	-0.0105	0.1256	0.0159	-0.0096	0.1253	0.0158	-0.0098	0.1118	0.0126
	β_2	0.0292	0.3004	0.0911	0.0303	0.2951	0.088	0.0298	0.2952	0.088	0.0292	0.2895	0.0847
	β_3	-0.021	0.2136	0.0461	-0.0149	0.206	0.0427	-0.0149	0.2063	0.0428	-0.0165	0.1974	0.0392
M	β_1	0.0023	0.0966	0.0093	-0.0037	0.0811	0.0066	-0.0017	0.0801	0.0064	-0.002	0.0575	0.0033
	β_2	1e-04	0.0947	0.009	2e-04	0.0808	0.0065	0	0.0791	0.0063	3e-04	0.0578	0.0033
	β_3	-0.0011	0.0913	0.0083	0.0025	0.0793	0.0063	0.0017	0.0784	0.0061	4e-04	0.0571	0.0033

Table 2.11: Bias, SE, and MSE of estimators: γ is MTN with $n = 10$ and $T = 5$.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_1	-3e-04	0.0271	7e-04	-2e-04	0.0174	3e-04	-2e-04	0.0174	3e-04	-3e-04	0.0176	3e-04
	β_2	7e-04	0.0283	8e-04	-6e-04	0.0175	3e-04	-5e-04	0.0174	3e-04	-6e-04	0.0177	3e-04
	β_3	-0.0011	0.0262	7e-04	-1e-04	0.0174	3e-04	-2e-04	0.0174	3e-04	0	0.018	3e-04
La	β_1	0.0028	0.0375	0.0014	2e-04	0.0169	3e-04	2e-04	0.017	3e-04	3e-04	0.0174	3e-04
	β_2	0.0015	0.0362	0.0013	0.001	0.0169	3e-04	0.001	0.0169	3e-04	0.0011	0.0172	3e-04
	β_3	-6e-04	0.0379	0.0014	-0.0013	0.0174	3e-04	-0.0014	0.0174	3e-04	-0.0014	0.0179	3e-04
Lo	β_1	-4e-04	0.0446	0.002	-1e-04	0.0174	3e-04	-1e-04	0.0174	3e-04	-3e-04	0.0178	3e-04
	β_2	-8e-04	0.0469	0.0022	0	0.0171	3e-04	0	0.0171	3e-04	0	0.0177	3e-04
	β_3	0.0015	0.0449	0.002	-0.0011	0.0175	3e-04	-0.0011	0.0176	3e-04	-0.001	0.0179	3e-04
M	β_1	-0.0101	0.1218	0.0149	-0.0126	0.1148	0.0133	-0.0126	0.1148	0.0133	-0.0125	0.1148	0.0133
	β_2	0.0411	0.3424	0.1189	0.0389	0.3406	0.1175	0.0389	0.3406	0.1175	0.0388	0.3406	0.1175
	β_3	-0.027	0.2305	0.0539	-0.0257	0.2275	0.0524	-0.0258	0.2275	0.0524	-0.0259	0.2275	0.0524

Table 2.12: Bias, SE, and MSE of estimators: γ is normal with $n = 20$ and $T = 10$.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_1	0.001	0.0349	0.0012	-0.0011	0.0308	0.001	-0.001	0.0306	9e-04	-8e-04	0.0301	9e-04
	β_2	-0.0024	0.036	0.0013	-0.002	0.0332	0.0011	-0.0022	0.033	0.0011	-0.0019	0.0322	0.001
	β_3	0.0012	0.0331	0.0011	1e-04	0.0325	0.0011	2e-04	0.0322	0.001	1e-04	0.0312	0.001
La	β_1	-0.0065	0.0881	0.0078	-0.0067	0.0835	0.007	-0.0068	0.0834	0.007	-0.0066	0.0833	0.007
	β_2	0.0175	0.2357	0.0559	0.0181	0.2339	0.0551	0.0181	0.2339	0.0551	0.0176	0.2339	0.055
	β_3	-0.014	0.1602	0.0259	-0.013	0.1577	0.025	-0.0131	0.1577	0.025	-0.0137	0.1576	0.025
Lo	β_1	-0.002	0.0742	0.0055	-0.0025	0.0632	0.004	-0.0025	0.0632	0.004	-0.0026	0.0628	0.004
	β_2	0.01	0.172	0.0297	0.0079	0.1673	0.0281	0.008	0.1673	0.0281	0.0076	0.1672	0.028
	β_3	-0.0063	0.1201	0.0145	-0.0061	0.1137	0.013	-0.0062	0.1137	0.013	-0.0066	0.1135	0.0129
M	β_1	2e-04	0.0468	0.0022	4e-04	0.0306	9e-04	4e-04	0.0304	9e-04	7e-04	0.0298	9e-04
	β_2	-0.0034	0.0463	0.0022	-0.0033	0.0324	0.0011	-0.0033	0.0322	0.001	-0.0031	0.0318	0.001
	β_3	0.0031	0.0468	0.0022	6e-04	0.0325	0.0011	7e-04	0.0324	0.0011	8e-04	0.0319	0.001

Table 2.13: Bias, SE, and MSE of estimators: γ is logistic with $n = 20$ and $T = 10$.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_1	-0.0017	0.031	0.001	-6e-04	0.0247	6e-04	-6e-04	0.0247	6e-04	-1e-04	0.0216	5e-04
	β_2	4e-04	0.0306	9e-04	0.0012	0.0251	6e-04	0.0012	0.0251	6e-04	9e-04	0.0222	5e-04
	β_3	9e-04	0.0302	9e-04	0.0012	0.0242	6e-04	0.0013	0.0242	6e-04	7e-04	0.0214	5e-04
La	β_1	-0.0065	0.0881	0.0078	-0.0067	0.0835	0.007	-0.0068	0.0834	0.007	-0.0066	0.0833	0.007
	β_2	0.0175	0.2357	0.0559	0.0181	0.2339	0.0551	0.0181	0.2339	0.0551	0.0176	0.2339	0.055
	β_3	-0.014	0.1602	0.0259	-0.013	0.1577	0.025	-0.0131	0.1577	0.025	-0.0137	0.1576	0.025
Lo	β_1	0.0013	0.0891	0.0079	-1e-04	0.0824	0.0068	0.0017	0.0837	0.007	-0.0013	0.0593	0.0035
	β_2	0.0017	0.0872	0.0076	3e-04	0.0821	0.0067	3e-04	0.0808	0.0065	5e-04	0.0582	0.0034
	β_3	-7e-04	0.0824	0.0068	0.0016	0.0852	0.0073	0.0027	0.0832	0.0069	0.0012	0.0592	0.0035
M	β_1	-0.0023	0.0405	0.0016	-4e-04	0.0235	6e-04	-5e-04	0.0235	6e-04	-6e-04	0.0208	4e-04
	β_2	-5e-04	0.044	0.0019	3e-04	0.0238	6e-04	4e-04	0.0237	6e-04	0	0.0208	4e-04
	β_3	-0.0012	0.044	0.0019	-8e-04	0.025	6e-04	-8e-04	0.0248	6e-04	-0.001	0.022	5e-04

Table 2.14: Bias, SE, and MSE of estimators: γ is Laplace with $n = 20$ and $T = 10$.

		OLS			Within			RE			m.d.		
		bias	SE	MSE	bias	SE	MSE	bias	SE	MSE	bias	SE	MSE
N	β_1	0.0018	0.0377	0.0014	0.0012	0.0374	0.0014	0.0011	0.0371	0.0014	7e-04	0.0233	5e-04
	β_2	-0.0011	0.0363	0.0013	-9e-04	0.0351	0.0012	-0.0013	0.035	0.0012	0	0.0217	5e-04
	β_3	5e-04	0.0381	0.0015	-9e-04	0.0355	0.0013	-8e-04	0.0354	0.0013	-8e-04	0.0222	5e-04
La	β_1	-0.0165	0.137	0.019	-0.0167	0.1345	0.0184	-0.0168	0.1344	0.0184	-0.0171	0.1313	0.0175
	β_2	0.0519	0.3904	0.1551	0.0499	0.3901	0.1546	0.05	0.39	0.1546	0.0495	0.3891	0.1538
	β_3	-0.0334	0.2625	0.07	-0.0327	0.2615	0.0695	-0.0328	0.2615	0.0694	-0.0337	0.2599	0.0687
Lo	β_1	-0.0019	0.0696	0.0048	-0.0032	0.0579	0.0034	-0.003	0.0578	0.0033	-0.0026	0.0502	0.0025
	β_2	0.0079	0.1443	0.0209	0.0053	0.1391	0.0194	0.0055	0.139	0.0194	0.0058	0.1361	0.0186
	β_3	-0.0039	0.103	0.0106	-0.0037	0.0965	0.0093	-0.0036	0.0965	0.0093	-0.0042	0.0923	0.0085
M	β_1	4e-04	0.0493	0.0024	0.0013	0.0356	0.0013	0.0013	0.0356	0.0013	2e-04	0.0219	5e-04
	β_2	-7e-04	0.0489	0.0024	-4e-04	0.0366	0.0013	-5e-04	0.0365	0.0013	0	0.0225	5e-04
	β_3	-4e-04	0.0497	0.0025	1e-04	0.0387	0.0015	1e-04	0.0386	0.0015	-1e-04	0.0238	6e-04

Table 2.15: Bias, SE, and MSE of estimators: γ is MTN with $n = 20$ and $T = 10$.

2.3 Proofs for Chapter 2

Proof of Theorem 2.1.2. Section 5.5 of Koul (2002) illustrates the claim holds for independent errors. Proof of the theorem, therefore, will be similar to the one of Theorem 5.5.1 in that section. Define for $k = 1, 2, ..., p+1$, $u \in \mathbb{R}^{p+1}, y \in \mathbb{R}$,

$$\begin{aligned} J_k(y,u) &:= \sum_{i=1}^{n} d_{ik}\, F(y + u'Ax_i), \quad Y_k(y,u) := \sum_{i=1}^{n} d_{ik}\, I\big(\varepsilon_i \le y + u'Ax_i\big), \quad (2.31) \\ W_k(y,u) &:= Y_k(y,u) - J_k(y,u), \end{aligned}$$

where $x_i \in \mathbb{R}^{p+1}$, $i = 1, 2, ..., n$ are in the model (1.1), and $d_{ik} = a_k' x_i$. Note that $A \sum_{i=1}^{n} d_{ik} x_i = e_k$, where $e_k \in \mathbb{R}^{p+1}$ is elementary vector whose kth entry is 1. Therefore, rewrite

$$\begin{aligned} T(\beta + Au) = \sum_{k=1}^{p+1} \int & \Big[\{W_k(y,u) - W_k(y,0)\} + \{W_k(-y,u) - W_k(-y,0)\} \quad (2.32) \\ & + \{(J_k(y,u) - J_k(y,0)) - u'e_k\, f(y)\} \\ & + \{(J_k(-y,u) - J_k(-y,0)) - u'e_k\, f(y)\} \\ & + \{U_k(y,\beta) + 2u'e_k\, f(y)\}\Big]^2 \, dH(y). \end{aligned}$$

Note that the last term of the integrand is the kth coordinate of $U(y,\beta) + 2A^{-1}(t-\beta)f(y)$ vector in $Q(t)$. If we can show that suprema of L_H^2 norms of the first four terms of the integrand are $o_p(1)$, then applying Cauchy-Schwarz (C-S) inequality on the cross product terms in (2.32) will complete the proof. Therefore to prove theorem it suffices to show that for all $k = 1, 2, ..., p+1$,

$$E \sup \int \left| W_k(\pm y, u) - W_k(\pm y, 0) \right|^2 dH(y) = o(1), \tag{2.33}$$

$$\sup \int \left| (J_k(\pm y, u) - J_k(\pm y, 0)) - u' e_k \, f(y) \right|^2 dH(y) = o(1), \tag{2.34}$$

$$E \sup \int \left| U_k(y, \beta) + 2u' e_k \, f(y) \right|^2 dH(y) = O(1). \tag{2.35}$$

where sup is taken over $\|u\| \le b$.

First consider (2.35). By the symmetry of F and H and Fubini's theorem, we obtain

$$E \int \left\{ I\big(\varepsilon_i \le y\big) - I\big(- \varepsilon_i < y\big) \right\}^2 dH(y) = 4 \int_0^\infty (1 - F) \, dH.$$

In addition, Lemma 2.3.1 yields, for $j > i$,

$$\begin{aligned} & E \int \left\{ I\big(\varepsilon_i \le y\big) - I\big(- \varepsilon_i < y\big) \right\} \left\{ I\big(\varepsilon_j \le y\big) - I\big(- \varepsilon_j < y\big) \right\} dH(y) \\ \le \;\; & 20\sqrt{2} \; \alpha_l^{1/2}(j - i) \int_0^\infty (1 - F)^{1/2} dH. \end{aligned}$$

Together with the fact that $(1 - F) \le (1 - F)^{1/2}$, by (A.1), (A.2), (A.7), and Lemma 2.3.2, we obtain

$$\begin{aligned} E \int U_k(y, \beta)^2 dH(y) \;\; & \le \;\; 4n \max_{1 \le i \le n} d_{ik}^2 \int_0^\infty (1 - F) \, dH + 40\sqrt{2} \; n \max_{1 \le i \le n} d_{ik}^2 \\ & \qquad \times n^{-1} \sum_{i=1}^n \sum_{j=i+1}^n \alpha_l^{1/2}(j - i) \int_0^\infty (1 - F)^{1/2} \, dH \\ & < \;\; \infty. \end{aligned} \tag{2.36}$$

Therefore, (2.35) immediately follows.

The proof of (2.34) does not involve the dependence of errors, and hence, it is the same as the proof of (5.5.11) of Koul (2002). Thus, we shall prove (2.33). Here we consider the proof for the case $+y$ only. The similar facts will hold for the case $-y$.

To begin with let $J_{ku}^{\pm}(\cdot)$, $Y_{ku}^{\pm}(\cdot)$, and $W_{ku}^{\pm}(\cdot)$ denote $J_k(\cdot,u)$, $Y_k(\cdot,u)$ and $W_k(\cdot,u)$ in (2.31) when d_{ik} is replaced with $d_{ik}^{\pm}$ so that $J_k = J_k^+ - J_k^-$, $Y_k = Y_k^+ - Y_k^-$, and $W_k = W_k^+ - W_k^-$. Define for $x \in \mathbb{R}^{p+1}$, $u \in \mathbb{R}^{p+1}$, $y \in \mathbb{R}$,

$$\begin{aligned} p(y,u;x) &:= F(y+u'Ax) - F(y), \\ B_{ni} &:= I\big(\varepsilon_i \le y + u'Ax_i\big) - I\big(\varepsilon_i \le y\big) - p(y,u;x_i). \end{aligned}$$

Rewrite

$$W_{ku}^{\pm} - W_{k0}^{\pm} = \sum_{i=1}^{n} d_{ik}^{\pm}\left\{I\big(\varepsilon_i \le y + u'Ax_i\big) - I\big(\varepsilon_i \le y\big) - p(y,u;x_i)\right\}.$$

Note that

$$E\,B_{ni}^2 \le F(y+\theta_i\|u\|) - F(y) \tag{2.37}$$

Recall a lemma from Deo(1973).

Lemma 2.3.1. *Suppose for each $n \ge 1$, $\{\xi_{nj},\, 1 \le j \le n\}$ are strongly mixing random variables with mixing number α_l. Suppose X and Y are two random variables respectively measurable with respect to $\sigma\{\xi_{n1},...,\xi_{nk}\}$ and $\sigma\{\xi_{nk+m},...,\xi_{nn}\}$, $1 \le m$, $m+k \le n$. Assume p, q and r are such that $p^{-1}+q^{-1}+r^{-1}$, and $\|X\|_p \le \infty$ and $\|Y\|_q \le \infty$. Then for each $1 \le m$, $k+m \le n$*

$$|E(XY) - E(X)E(Y)| \le 10 \cdot \alpha_l^{1/r}(m)\|X\|_p\|Y\|_q.$$

Consequently if $\|X\|_\infty = B < \infty$ *then for* $q > 1$ *and each* $1 \le m$, $k + m \le n$

$$|E(XY) - E(X)E(Y)| \le 10 \cdot \alpha_l^{1-1/q}(m)\|Y\|_q.$$

In addition, consider following lemma.

Lemma 2.3.2. *For* $1 < r < 3$,

$$n^{-1}\sum_{i=1}^{n-1}\sum_{k=1}^{n-i}\alpha_l^{1/r}(k) = O(1).$$

Proof. Fix an $1 < r < 3$. Let p such that $1/r + 1/p = 1$. Note that

$$\frac{p}{r} = \frac{1}{r-1} > \frac{1}{2}.$$

Therefore, by Hölder's inequality, we have

$$n^{-1}\sum_{i=1}^{n-1}\sum_{k=1}^{n-i}\alpha_l^{1/r}(k) \;\le\; \left(\sum_{k=1}^{n-1}\frac{(n-k)^p}{n^p}\cdot\frac{1}{k^{2p/r}}\right)^{1/p}\left(\sum_{k=1}^{n-1}k^2\alpha_l(k)\right)^{1/r} < \infty. \quad (2.38)$$

The last inequality follows from the assumption (A.8), thereby completing the proof of the lemma.

Now, we consider the cross product terms of $E\big|W_{ku} - W_{k0}\big|_H^2$. Then

$$\begin{aligned}
&\Big|E\int \sum_{i=1}^{n}\sum_{j=i+1}^{n}\big[\,d_{ik}d_{jk}\{I\big(\varepsilon_i \le y + u'Ax_i\big) - I\big(\varepsilon_i \le y\big) - p(y,u;x_i)\}\\
&\qquad \times\{I\big(\varepsilon_j \le y + u'Ax_j\big) - I\big(\varepsilon_j \le y\big) - p(y,u;x_j)\}\big]\ dH(y)\Big|\\
\le\ & \sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ik}d_{jk}\int \big|E\,B_{ni}\,B_{nj}\big|\ dH\\
\le\ & 10b^{1/2}\{n\max_i d_{ik}^2\}\cdot\{\max_i \theta_i\}^{1/2}\cdot n^{-1}\sum_{i=1}^{n-1}\sum_{m=1}^{n-i}\alpha_l^{1/2}(m)\int f^{1/2}\,dH \to 0.
\end{aligned}$$

The second inequality follows from Lemma 2.3.1, and the convergence to zero follows from the Lemma 2.3.2 with $r = 2$, (A.1), and (A.7). Consequently, by Fubini's Theorem together with (A.3), we obtain, for every fixed $\|u\| \le b$,

$$\limsup_{n\to\infty} E\,|W_{ku} - W_{k0}|_H^2 \ \le\ \limsup_{n\to\infty}\int_{-a_n}^{a_n}\int f(y+s)dH(y)ds = 0,$$

where $a_n = b\max_i\theta_i \to 0$.

To complete the proof of (2.33), it suffices to show that for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $v \in \mathbb{R}^{p+1}$, $\|u - v\| \le \delta$,

$$\limsup_{n\to\infty} E \sup_{\|u-v\|\le\delta} |\mathcal{K}_{ku} - \mathcal{K}_{kv}| \le \epsilon, \tag{2.39}$$

where

$$\mathcal{K}_{ku} := |W_{ku} - W_{k0}|_H^2, \quad u \in \mathbb{R}^{p+1}, \quad 1 \le k \le p+1.$$

(2.39) follows from (5.5.5) of Koul (2002), thereby completing the proof of theorem.

Proof of Corollary 2.1.1. The proof of the claim for independent errors can again be found in the section 5.5 of Koul (2002). The difference between the proof in the section 5.5 and one here arises only in the part which involves the dependence of the error. Thus, we present only the proof of an analogue of (5.5.27) in Koul (2002). Let

$$\begin{aligned} L_k &:= \int \Big[\sum_{i=1}^{n} d_{ik}\, \big\{I\big(\varepsilon_i \le y\big) - I\big(-\varepsilon_i < y\big)\big\}\Big] f(y)\, dH(y), \\ L &:= (L_1, ..., L_{p+1}). \end{aligned}$$

Note that $L_k = \int U_k(y, \beta) f(y) dH(y)$. Therefore, using C-S inequality together with (2.36), we obtain, for some $0 < C < \infty$

$$E\|L\|^2 \le C(p+1)|f|_H^2 \int_0^\infty (1-F)^{1/2}\, dH.$$

Using $E\, L_k = 0$ for $k = 1, ..., p+1$ and Chebyshev inequality, for all $\epsilon > 0$ there exists N_1 and c_ϵ such that

$$P\big(\|L\| \le c_\epsilon\big) \ge 1 - \frac{C(p+1)|f|_H^2 \int (1-F)^{1/2} dH}{c_\epsilon} \ge 1 - \epsilon/2, \quad n \ge N_1.$$

The rest of the proof is the same as that of Lemma 5.5.4 of Koul (2002).

Proofs of Section 2.1.2

Proof of Theorem 2.1.6. Note that

$$\begin{aligned}
\widehat{S}_k(y,r) &= S_k(y,r) + n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\Big[\Big\{ I\big(\xi_i \le y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big) - I\big(\xi_i \le y + Z_i'v/\sqrt{n}\big)\Big\} \\
&\quad -\Big\{ I\big(-\xi_i \le y - Z_i'v/\sqrt{n} - \zeta_{ni}(u,v)\big) - I\big(-\xi_i < y - Z_i'v/\sqrt{n}\big)\Big\}\Big] \\
&\quad - n^{-1/2}\sum_{i=1}^{n}\eta_{n,i-k}(u)\Big\{ I\big(\xi_i \le y + Z_i'v/\sqrt{n}\big) - I\big(-\xi_i < y - Z_i'v/\sqrt{n}\big)\Big\} \\
&\quad - n^{-1/2}\sum_{i=1}^{n}\eta_{n,i-k}(u)\Big[\Big\{ I\big(\xi_i \le y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big) - I\big(\xi_i \le y + Z_i'v/\sqrt{n}\big)\Big\} \\
&\quad -\Big\{ I\big(-\xi_i \le y - Z_i'v/\sqrt{n} - \zeta_{ni}(u,v)\big) - I\big(-\xi_i < y - Z_i'v/\sqrt{n}\big)\Big\}\Big] \\
&= S_r + R1_{uv}^k - R2_{uv}^k - R3_{uv}^k, \quad (say).
\end{aligned}$$

In what follows, the index k runs from 1 to $p+1$. In order to conserve the space, we write η_u^{ik}, ζ_{iuv}, etc. for $\eta_{n,i-k}(u)$, $\zeta_{ni}(u,v)$, etc. Similar to the way as that used in Theorem 2.1.2, in order to prove Theorem 2.1.6, it suffices to show that for $j = 1,2,3$

$$\sup_{\|u\|\le b, \|v\|\le b} \big|Rj_{uv}^k\big|_H^2 = o_p(1).$$

Consider $R1_{uv}^k$. First, define for $1 \le k \le q$

$$\begin{aligned}
w_k(y,u,v) &:= n^{-1}\sum_{i=1}^{n}\varepsilon_{i-k} I\big(\xi_i \le y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big), \\
\nu_k(y,u,v) &:= n^{-1}\sum_{i=1}^{n}\varepsilon_{i-k} F_\xi(y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)), \\
\mathcal{W}_k(y,u,v) &:= \sqrt{n}\big[w_k(y,u,v) - \nu_k(y,u,v)\big].
\end{aligned}$$

Also consider the following processes:

$$\begin{aligned} \mathcal{T}_k(y;u,v,\pm\delta) &:= n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}I\big(\xi_i \le y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v) \pm d_{ni}(\delta,u,v)\big), \\ m_k(y;u,v,\pm\delta) &:= n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}F_\xi(y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v) \pm d_{ni}(\delta,u,v)), \\ \mathcal{Z}_k(y;u,v,\pm\delta) &:= \mathcal{T}_k(y;u,v,\pm\delta) - m_k(y;u,v,\pm\delta). \end{aligned}$$

Note that $d_{ni}(-\delta,u,v) \neq -d_{ni}(\delta,u,v)$. However, for the purpose of conserving the space, let

$$\mathcal{T}_k(y;u,v,-\delta) := n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}I\big(\xi_i \le y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v) - d_{ni}(\delta,u,v)\big).$$

Similarly, we define $m_k(y;u,v,-\delta)$ and $\mathcal{Z}_k(y;u,v,-\delta)$ by subtracting term $d_{ni}(\delta,u,v)$. Let $\varepsilon_{i-k}^{+} = 0 \vee \varepsilon_{i-k}$ and $\varepsilon_{i-k}^{-} = \varepsilon_{i-k}^{+} - \varepsilon_{i-k}$. Let any process with $\pm$ superscript denote the one with ε_{i-k} replaced by $\varepsilon_{i-k}^{\pm}$. Also drop the subscript k, and let w_{uv}, ν_{uv}, $\mathcal{W}_{uv}$, etc. stand for $w_k(\cdot,u,v)$, $\nu_k(\cdot,u,v)$, $\mathcal{W}_k(\cdot,u,v)$, etc.

Now, consider $s \in \mathbb{R}^{p+1}, t \in \mathbb{R}^q$ where $\|s-u\| \le \delta$ and $\|t-v\| \le \delta$. Then we obtain

$$Z_i'v/\sqrt{n} + \zeta_{ni}(u,v) - d_{ni}(\delta,u,v) \le Z_i't/\sqrt{n} + \zeta_{ni}(s,t) \le Z_i'v/\sqrt{n} + \zeta_{ni}(u,v) + d_{ni}(\delta,u,v). \tag{2.40}$$

Note that $\mathcal{T}(y;u,v,0) = \sqrt{n}\,w_{uv}$ and $m(y;u,v,0) = \sqrt{n}\,\nu_{uv}$. Monotonicity of indicator function together with (2.40) in turn implies

$$\mathcal{T}^{\pm}(y;u,v,-\delta) - \mathcal{T}^{\pm}(y;u,v,0) \le \sqrt{n}\,[w_{st}^{\pm} - w_{uv}^{\pm}] \le \mathcal{T}^{\pm}(y;u,v,\delta) - \mathcal{T}^{\pm}(y;u,v,0)$$

for all $y \in \mathbb{R}$, $\|s\| \le b$, $\|t\| \le b$, $\|s-u\| \le \delta$ and $\|t-v\| \le \delta$. Using the fact that $a_1 \le a_2 \le a_3$ implies $|a_2| \le |a_1| + |a_3|$, appropriate centering of $\mathcal{T}^{\pm}$ yields

$$\begin{aligned}
& \sqrt{n}\Big|w^{\pm}_{st} - w^{\pm}_{uv}\Big| \qquad (2.41) \\
\le\ & \Big|\mathcal{T}^{\pm}(y; u, v, -\delta) - \mathcal{T}^{\pm}(y; u, v, 0)\Big| + \Big|\mathcal{T}^{\pm}(y; u, v, \delta) - \mathcal{T}^{\pm}(y; u, v, 0)\Big| \\
\le\ & \Big|\mathcal{Z}^{\pm}(y; u, v, -\delta) - \mathcal{Z}^{\pm}(y; u, v, 0)\Big| + \Big|m^{\pm}(y; u, v, -\delta) - m^{\pm}(y; u, v, 0)\Big| \\
& + \Big|\mathcal{Z}^{\pm}(y; u, v, \delta) - \mathcal{Z}^{\pm}(y; u, v, 0)\Big| + \Big|m^{\pm}(y; u, v, \delta) - m^{\pm}(y; u, v, 0)\Big|.
\end{aligned}$$

Lemma 2.3.3. *Assumption (B.2) implies*

$$E \int \big[\mathcal{Z}^{\pm}(y; u, v, \delta) - \mathcal{Z}^{\pm}(y; u, v, 0)\big]^2 dH(y) = o(1). \qquad (2.42)$$

Also, assumption (B.3) implies with c and δ in (B.3)

$$\liminf_{n\to\infty} P\Big(\sup \int n\big\{\nu^{\pm}(y, s, t) - \nu^{\pm}(y, u, v)\big\}^2 dH(y) \le c\delta^2\Big) = 1 \qquad (2.43)$$

where supremum is taken over $\|s-u\| \le \delta$ and $\|t-v\| \le \delta$. Moreover, assumptions (B.1)-(B.5) imply

$$\begin{aligned}
\sup_{\|u\|\le b, \|v\|\le b} \int n^{-1}\Big[\sum_{i=1}^{n} \varepsilon_{i-k}\big\{F_{\xi}(y + n^{-1/2}v'Z_i + \zeta_{ni}(u, v)) - F_{\xi}(y) \qquad (2.44) \\
- f_{\xi}(y)\{n^{-1/2}v'Z_i + \zeta_{ni}(u, v)\}\big\}\Big]^2 dH(y) = o_p(1).
\end{aligned}$$

Proof. Let $\mathcal{F}_{i-1} := \sigma\{\xi_{i-1}, \xi_{i-2}, ...\}$. Define

$$\tilde{p}(y, u, v; z) := F_{\xi}(y + n^{-1/2}v'Z_i + \zeta_{ni}(u, v) + d_{ni}(\delta, u, v)) - F_{\xi}(y + n^{-1/2}v'Z_i + \zeta_{ni}(u, v)).$$

Observe that ith summand is conditionally centered r.v.'s, given $\mathcal{F}_{i-1}$, and hence, covariance of two summands is zero. Similar to (2.37), the conditional variance of ith summand, given $\mathcal{F}_{i-1}$, is bounded by

$$\begin{aligned} E(\varepsilon_{i-k}^{\pm})^2|\tilde{p}(y,u,v;Z_i)| \leq{}& E(\varepsilon_{i-k}^{\pm})^2|F_\xi(y+n^{-1/2}v'Z_i+\zeta_{ni}(u,v)+d_{ni}(\delta,u,v))-F_\xi(y)| \\ &+E(\varepsilon_{i-k}^{\pm})^2|F_\xi(y+n^{-1/2}v'Z_i+\zeta_{ni}(u,v))-F_\xi(y)|. \end{aligned}$$

Therefore, (2.42) follows from Fubini's Theorem and repeated application of (B.2) with δ and $\delta = 0$. To prove (2.43), observe that the monotonicity of F_ξ and (2.40) imply

$$\begin{aligned} m^{\pm}(y;u,v,-\delta) \leq m^{\pm}(y;s,t,0) \leq m^{\pm}(y;u,v,\delta), \\ m^{\pm}(y;u,v,-\delta) \leq m^{\pm}(y;u,v,0) \leq m^{\pm}(y;u,v,\delta), \end{aligned}$$

and hence,

$$m^{\pm}(y;u,v,-\delta) - m^{\pm}(y;u,v,0) \leq n^{1/2}[\nu_{st}^{\pm} - \nu_{uv}^{\pm}] \leq m^{\pm}(y;u,v,\delta) - m^{\pm}(y;u,v,0).$$

Using the fact that $a_1 \leq a_2 \leq a_3$ implies $|a_2| \leq |a_1| + |a_3|$, we obtain

$$\begin{aligned} n^{1/2}|\nu_{st}^{\pm} - \nu_{uv}^{\pm}| \leq{}& n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}^{\pm}\Big\{F_\xi(y+Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)+d_{ni}(\delta,u,v)) \\ &-F_\xi(y+Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)-d_{ni}(\delta,u,v))\Big\}. \end{aligned} \quad (2.45)$$

As a consequence of (2.45) and the assumption (B.3), we finally obtain

$$\begin{aligned}
&\liminf_{n\to\infty} P\Big(\sup\big|\sqrt{n}\{\nu^{\pm}(y,s,t)-\nu^{\pm}(y,u,v)\}\big|_H^2 \le c\delta^2\Big) \\
\ge\ &\liminf_{n\to\infty} P\Bigg(\sup\Big|n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}^{\pm}\Big\{F_{\xi}(y+Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)+d_{ni}(\delta,u,v)) \\
&\qquad -F_{\xi}(y+Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)-d_{ni}(\delta,u,v))\Big\}\Big|_H^2 \le c\delta^2\Bigg) \\
=\ &1,
\end{aligned}$$

where the supremums are taken over $\|s-u\|\le\delta$ and $\|t-v\|\le\delta$, and hence, (2.43) follows. Next, write $m_{uv}^{\pm}$ and $m^{\pm}$ for $m^{\pm}(y;u,v,0)$ and $m^{\pm}(y;0,0,0)$. Let

$$\mathcal{L}_{uv}=\Big|m_{uv}^{\pm}-m^{\pm}-n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}^{\pm}f_{\xi}\{n^{-1/2}v'Z_i+\zeta_{ni}(u,v)\}\Big|_H^2.$$

In order to prove (2.44), by using (B.4) and compactness of $\|u\|\le b$ and $\|v\|\le b$ (see Theorem 9.1.1 of Koul (2002)), it suffices to show that for all $\epsilon>0$ there exists a $\delta>0$ such that

$$\limsup_{n\to\infty} P\Big(\sup\big|\mathcal{L}_{st}-\mathcal{L}_{uv}\big|>\epsilon\Big)=0, \tag{2.46}$$

where sup is taken over $\|s\|,\|t\|\le b$, $\|s-u\|\le\delta$, and $||t-v||\le\delta$. Using monotonicity of F_{ξ}, (2.40), triangle inequality, $(a_1\pm a_2)^2\le 2(a_1^2+a_2^2)$, and (B.3), we obtain for $\|s\|,\|t\|\le b$ with $\|s-u\|\le\delta$ and $\|t-v\|\le\delta$

$$\limsup_{n\to\infty} P\big(\sup|m_{st}-m_{uv}|_H^2\ge 4c\delta^2\big)=0 \tag{2.47}$$

where the equality follows from (2.43). Next, observe that

$$
\begin{aligned}
& \big|\,\mathcal{L}_{st}-\mathcal{L}_{uv}\big| \\
\le\ & \int \Big[(m_{st}-m_{uv}) - n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\big\{\big(Z_i't/\sqrt{n}+\zeta_{ni}(s,t)\big)-\big(Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)\big)\big\}f_{\xi}(y)\Big] \\
& \times\Big[(m_{st}-m_{uv}) + 2\big\{m_{uv}-m-n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\big(Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)\big)f_{\xi}(y)\big\} \\
& -n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\big\{\big(Z_i't/\sqrt{n}+\zeta_{ni}(s,t)\big)-\big(Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)\big)\big\}f_{\xi}(y)\Big]dH(y)
\end{aligned}
$$

$$
\begin{aligned}
\le\ & 2\big|m_{st}-m_{uv}\big|_H^2 + 2\big|n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}d_{ni}(\delta,u,v)f_{\xi}\big|_H^2 + 2\big|m_{st}-m_{uv}\big|_H \\
& \times\big|n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}d_{ni}(\delta,u,v)\big|_H + 2\big|m_{uv}-m-n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\big(Z_i'v/\sqrt{n}+\zeta_{ni}(u,v)\big)f_{\xi}\big|_H \\
& \times\Big\{\big|m_{st}-m_{uv}\big|_H + \big|n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}d_{ni}(\delta,u,v)f_{\xi}\big|_H\Big\}.
\end{aligned}
$$

In the first inequality, we use the fact that $(a_1^2-a_2^2)=(a_1-a_2)\cdot(a_1+a_2)$; (2.47) implies that supremum of the first term of the last equation is $\delta o_p(1)$; the assumption (B.1) and the fact that $d_{ni}(\delta,u,v)=O_p(n^{-1/2}\delta\|Z_i\|)$ imply that supremum of the second term is $\delta O_p(1)$, and hence, supremum of the third term is $\delta o_p(1)$; the assumption (B.4) ensures that supremum of the last term is $\delta o_p(1)$. Therefore, by making δ as small as desired, (2.46) follows, thereby completing the proof of (2.44).

Lemma 2.3.4.

$$
\sup_{\|u\|\le b,\|v\|\le b}\int\Big[n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\zeta_{ni}(u,v)f_{\xi}(y)\Big]^2 dH(y)=o_p(1).
$$

Proof. Note that $\sup_{\|u\|\le b,\|v\|\le b}\zeta_{ni}(u,v) = O(n^{-1/2})$. Together with this fact, Fubini's theorem yields

$$\begin{aligned}
&\limsup_{n\to\infty} E \sup_{\|u\|\le b,\|v\|\le b} \int n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_{i-k}\varepsilon_{j-k}\zeta_{ni}(u,v)\zeta_{nj}(u,v)f_\xi^2(y)\,dH(y)\\
\le\ &\limsup_{n\to\infty} C\,|f_\xi|_H^2\cdot n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_l^{1/2}(j-i)\|\varepsilon_1\|_4^2 = 0,
\end{aligned}$$

where $0 < C < \infty$. The inequality follows from Lemma 2.3.1; the last equality follows from (B.1.a) and Lemma 2.3.2. Consequently,

$$E\sup\left|n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\zeta_{ni}(u,v)f_\xi\right|_H^2 \to 0,$$

where the supremum is taken over $\|u\| \le b$ and $\|v\| \le b$, thereby completing the proof of lemma.

Corollary 2.3.1.

$$\sup_{\|u\|\le b,\|v\|\le b}\left|m_{uv} - m_{0v}\right|_H^2 = o_p(1).$$

Proof. Note that

$$\begin{aligned}
&m(y,u,v,0) - m(y,0,v,0)\\
=\ &\Big\{m(y,u,v,0) - m(y,0,0,0) - n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\big(Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big)f_\xi(y)\Big\}\\
&-\Big\{m(y,0,v,0) - m(y,0,0,0) - v'n^{-1}\sum_{i=1}^{n}\varepsilon_{i-k}Z_i f_\xi(y)\Big\} + n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\zeta_{ni}(u,v)f_\xi(y).
\end{aligned}$$

Hence the proof of corollary will be completed by simply applying C-S inequality on the cross product terms after we show that

$$\begin{aligned}
&\sup \left|m_{uv} - m - n^{-1/2}\sum_{i=1}^{n} \varepsilon_{i-k}\big(Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big) f_\xi\right|_H^2 = o_p(1),\\
&\sup \left|m_{0v} - m - v'n^{-1}\sum_{i=1}^{n} \varepsilon_{i-k} Z_i f_\xi\right|_H^2 = o_p(1),\\
&\sup \left|n^{-1/2}\sum_{i=1}^{n} \varepsilon_{i-k}\zeta_{ni}(u,v) f_\xi\right|_H^2 = o_p(1),
\end{aligned}$$

where the supremums are taken over $\|u\| \le b$ and $\|v\| \le b$. The first and the second follow from (2.44) of Lemma 2.3.3; the last follows from Lemma 2.3.4.

Lemma 2.3.5.

$$\sup_{\|u\|\le b, \|v\|\le b} \int \Big\{\mathcal{W}^{\pm}(y,u,v) - \mathcal{W}^{\pm}(y,0,0)\Big\}^2 dH(y) = o_p(1).$$

Proof. Similar to the square root of the integrand in (2.42), $\sqrt{n}\big[\mathcal{W}^{\pm}(y,u,v) - \mathcal{W}^{\pm}(y,0,0)\big]$ is sum of conditionally centered r.v.'s. Thus, by repeated application of (B.2) with δ and $\delta = 0$, we easily obtain for fixed $\|u\|$, $\|v\| \le b$,

$$\big|\mathcal{W}^{\pm}_{uv} - \mathcal{W}^{\pm}\big|_H^2 = o_p(1). \tag{2.48}$$

Let $\mathcal{C}_{uv} := \big|\mathcal{W}^{\pm}_{uv} - \mathcal{W}^{\pm}\big|_H^2$. Similar to the proof of (2.46), in order to complete the proof, it suffices to show that for every ϵ there exists δ such that for every $\|u\|$, $\|v\| \le b$,

$$\limsup_{n\to\infty} P\Big(\sup \big|\mathcal{C}_{st} - \mathcal{C}_{uv}\big| > \epsilon\Big) = 0, \tag{2.49}$$

where sup is taken over $\|s\|, \|t\| \le b$, $\|s-u\| \le \delta$, and $\|t-v\| \le \delta$. Note that

$$\big|\,\mathcal{C}_{st} - \mathcal{C}_{uv}\big| \le |\,\mathcal{W}^{\pm}_{st} - \mathcal{W}^{\pm}_{uv}\,|^2_H + 2|\,\mathcal{W}^{\pm}_{st} - \mathcal{W}^{\pm}_{uv}|_H \cdot |\,\mathcal{W}^{\pm}_{uv} - \mathcal{W}^{\pm}\,|_H.$$

By the monotonicity of F_ξ together with (2.40) we obtain that

$$|m^{\pm}(y;u,v,\pm\delta) - m^{\pm}(y;u,v,0)| \le m^{\pm}(y;u,v,\delta) - m^{\pm}(y;u,v,-\delta).$$

Recall definition of $\mathcal{W}^{\pm}$. Then triangle inequality yields that

$$\begin{aligned} |\,\mathcal{W}_{st} - \mathcal{W}_{uv}\,| &\le |\,\mathcal{W}^{+}_{st} - \mathcal{W}^{+}_{uv}\,| + |\,\mathcal{W}^{-}_{st} - \mathcal{W}^{-}_{uv}\,|, \\ |\,\mathcal{W}^{\pm}_{st} - \mathcal{W}^{\pm}_{uv}\,| &\le \sqrt{n}\Big\{\,|w^{\pm}_{st} - w^{\pm}_{uv}| + |\nu^{\pm}_{st} - \nu^{\pm}_{uv}|\Big\}. \end{aligned}$$

Therefore, using $(a_1+a_2)^2 \le 2(a_1^2+a_2^2)$ repeatedly and (2.41), we have

$$\begin{aligned} |\,\mathcal{W}^{\pm}_{st} - \mathcal{W}^{\pm}_{uv}\,|^2_H \le\ & 16\Bigg\{\int [\mathcal{Z}^{\pm}(y;u,v,\delta) - \mathcal{Z}^{\pm}(y;u,v,0)]^2\, dH(y) \\ & + \int [\mathcal{Z}^{\pm}(y;u,v,-\delta) - \mathcal{Z}^{\pm}(y;u,v,0)]^2\, dH(y) \\ & + \int [m^{\pm}(y;u,v,\delta) - m^{\pm}(y;u,v,-\delta)]^2\, dH(y) \\ & + \big|\sqrt{n}(\nu^{\pm}_{st} - \nu^{\pm}_{uv})\big|^2_H \Bigg\} \end{aligned}$$

for all $\|s\|$, $\|t\| \le$, $\|s-u\| \le \delta$, $\|t-v\| \le \delta$. Finally, assumption (B.3), (2.42), (2.43), and

(2.48) prove (2.49), thereby completing the proof of the lemma. Define

$$
\begin{aligned}
SS_k(y,u,v) \;\; &:= \;\; n^{-1/2}\sum_{i=1}^{n}\varepsilon_{i-k}\Big\{ I\big(\xi_i \le y + Z_i'v/\sqrt{n} + \zeta_{ni}(u,v)\big) \\
&\qquad -I\big(-\xi_i < y - Z_i'v/\sqrt{n} - \zeta_i(u,v)\big)\Big\}.
\end{aligned}
$$

Rewrite

$$
\begin{aligned}
R1_{uv}^k \;\; &= \;\; \Big[\mathcal{W}(y,u,v) - \mathcal{W}(y,0,0)\Big] - \Big[\mathcal{W}(y,0,v) - \mathcal{W}(y,0,0)\Big] \\
&\quad + \Big[\mathcal{W}(-y,u,v) - \mathcal{W}(-y,0,0)\Big] - \Big[\mathcal{W}(-y,0,v) - \mathcal{W}(-y,0,0)\Big] \\
&\quad + \Big[m(y,u,v,0) - m(y,0,v,0)\Big] - \Big[m(-y,u,v,0) - m(-y,0,v,0)\Big].
\end{aligned}
$$

Using quadratic expansion and applying C-S inequality involving L_H^2 norm on the cross product, the claim that $\sup_{\|u\|,\|v\|} R1_{uv}^k = o_p(1)$ will be true if we can show that

$$
\begin{aligned}
&\text{(a)} \sup_{\|u\|\le b,\|v\|\le b} |\mathcal{W}_{uv} - \mathcal{W}|_H^2 = o_p(1), \quad \text{(b)} \sup_{\|v\|\le b} |\mathcal{W}_{0v} - \mathcal{W}|_H^2 = o_p(1), \\
&\text{(c)} \sup_{\|u\|\le b,\|v\|\le b} |m_{uv} - m_{0v}|_H^2 = o_p(1).
\end{aligned}
$$

(a) and (b) follow from Lemma 2.3.5; (c) follows from Corollary 2.3.1.

Next, consider $R2_{uv}^k$. To begin with, note that by C-S inequality and Fubini's theorem together, we obtain for $v \in \mathbb{R}^q$, and $0 < b < \infty$

$$
\begin{aligned}
& E\int\Big\{F_\xi(y + v'Z_1/\sqrt{n}) - F_\xi(y - v'Z_1/\sqrt{n})\Big\}^2 dH(y) \qquad (2.50)\\
\le \;\; & 4b^2 n^{-1}\big(2b/\sqrt{n}\big)^{-1}\int_{-b/\sqrt{n}}^{b/\sqrt{n}}\int E\big[\|Z_1\| f_\xi(y + s\|Z_1\|)\big]^2 dH(y)\, ds.
\end{aligned}
$$

Moreover, by the assumptions (B.1.a) and (B.1.c) and bounded convergence theorem, we have

$$\begin{aligned} & \int \Big[E\big|F_\xi(y+v'Z_1/\sqrt{n}) - F_\xi(y-v'Z_1/\sqrt{n})\big|\Big]^{1/2} dH(y) \qquad (2.51) \\ \leq \; & \sqrt{2b}\, n^{-1/4} \int \Big[(2b/\sqrt{n})^{-1} \int_{-b/\sqrt{n}}^{b/\sqrt{n}} E\{\|Z_1\| f_\xi(y+s\|Z_1\|)\} ds\Big]^{1/2} dH(y) = O(n^{-1/4}). \end{aligned}$$

Finally, consider $v_1,\ v_2 \in \mathbb{R}^q$ and let $\|v_1\| \leq \|v_2\|$. Using C-S inequality and Fubini's theorem again, we have

$$\begin{aligned} & E\int \Big\{F_\xi(y+v_2'Z_1/\sqrt{n}) - F_\xi(y+v_1'Z_1/\sqrt{n})\Big\}^2 dH(y) \qquad (2.52) \\ \leq \; & n^{-1}(\|v_2\|+\|v_1\|)^2 \big\{(\|v_2\|+\|v_1\|)/\sqrt{n}\big\}^{-1} \int_{-\|v_1\|/\sqrt{n}}^{\|v_2\|/\sqrt{n}} \int E\big[\|Z_1\| f_\xi(y+s\|Z_1\|)\big]^2 dH(y)\, ds. \end{aligned}$$

Define

$$\boldsymbol{q}(y,v;z) := F_\xi(y+n^{-1/2}v'z) - F_\xi(y-n^{-1/2}v'z).$$

Next, rewrite

$$\begin{aligned} R2^k_{uv} &= n^{-1/2}\sum_{i=1}^n \eta_u^{ik}\Big\{\, I\big(\xi_i \leq y+n^{-1/2}v'Z_i\big) - I\big(-\xi_i < y - n^{-1/2}v'Z_i\big) - \boldsymbol{q}(y,v;Z_i)\,\Big\} \\ & \quad + n^{-1/2}\sum_{i=1}^n \eta_u^{ik}\, \boldsymbol{q}(y,v;Z_i) \\ &= R21^k_{uv} + R22^k_{uv}, \qquad \textit{say}. \end{aligned}$$

Due to the monotonicity of F_ξ, we obtain

$$\sup_{\|v\|\leq b} |\boldsymbol{q}(y,v;Z_i)| \leq F_\xi(y+n^{-1/2}b\|Z_i\|) - F_\xi(y-n^{-1/2}b\|Z_i\|). \qquad (2.53)$$

Therefore, by C-S inequality, (2.53), $\vartheta_n = o(1)$, and (2.50), we have

$$E \sup_{\|u\|,\|v\|\le b} |R22^k_{uv}|^2_H \le 4b^4\,\vartheta_n^2\,\big(2b/\sqrt{n}\big)^{-1}\int_{-b/\sqrt{n}}^{b/\sqrt{n}}\int E\big[\|Z_1\| f_\xi(y+s\|Z_1\|)\big]^2\,dH(y)\,ds \to 0.$$

Next, consider the cross product of $R21^k_{uv}$. For fixed u, v

$$\begin{aligned}
& E\int\frac{1}{n}\sum_{i=1}^n\sum_{j=i+1}^n\Big[\eta_u^{ik}\eta_u^{jk}\Big\{I\big(\xi_i\le y+n^{-1/2}v'Z_i\big)-I\big(-\xi_i<y-n^{-1/2}v'Z_i\big)\\
&\quad -\boldsymbol{q}(y,v;Z_i)\Big\}\times\Big\{I\big(\xi_j\le y+n^{-1/2}v'Z_j\big)-I\big(-\xi_j<y-n^{-1/2}v'Z_j\big)\\
&\qquad -\boldsymbol{q}(y,v;Z_j)\Big\}\Big]\,dH(y)\\
\le\;& b^2\vartheta_n^2 n^{-1}\sum_{i=1}^n\sum_{j=i+1}^n\int\Big|E\Big\{I\big(\xi_i\le y+n^{-1/2}v'Z_i\big)-I\big(-\xi_i<y-n^{-1/2}v'Z_i\big)\\
&\quad -\boldsymbol{q}(y,v;Z_i)\Big\}\times\Big\{I\big(\xi_j\le y+n^{-1/2}v'Z_j\big)-I\big(-\xi_j<y-n^{-1/2}v'Z_j\big)\\
&\qquad -\boldsymbol{q}(y,v;Z_j)\Big\}\Big|dH(y)\\
\le\;& 10b^2\,\vartheta_n^2\,n^{-1}\sum_{i=1}^n\sum_{j=i+1}^n\alpha_l^{1/2}(j-i)\\
&\qquad\times\int\Big[E\big|F_\xi(y+v'Z_1/\sqrt{n})-F_\xi(y-v'Z_1/\sqrt{n})\big|\Big]^{1/2}dH(y)\\
\to\;& 0.
\end{aligned}$$

The first inequality follows from Fubini's theorem; the second inequality follows from Lemma 2.3.1; the convergence to zero follows from $\vartheta_n = 0(1)$, (2.38) with $r = 2$, and (2.51). Therefore, the sum of the covariance of two summands in $R21^k_{uv}$ will tend to zero. Consequently, for fixed u and v, by Fubini's theorem, stationarity of ε, and (B.5), we obtain

$$
\begin{aligned}
& \limsup_{n\to\infty} E|R21^k_{uv}|^2_H && (2.54)\\
\le\ & \limsup_{n\to\infty} 2b^3\,\vartheta_n^2\, n^{-1/2}\,(2b/n^{1/2})^{-1}\int_{-b/n^{1/2}}^{b/n^{1/2}}\int E\|Z_1\| f_\xi(y+s\cdot\|Z_1\|)dH(y)ds\\
=\ & 0.
\end{aligned}
$$

Finally, we shall show that $E\sup|R21^k_{uv}|^2_H = o(1)$. Let $u_h = x_h\, e_u$ and $v_l = w_l\, e_v$ for $0 \le l,\, h \le r$ where $x_h = \|u_h\|$ and $w_l = \|v_l\|$, and e_u, e_v are unit vectors in $\mathbb{R}^{p+1}$ and $\mathbb{R}^q$. Let

$$
-b = x_0 \le x_1 \le \cdots \le x_r = b, \quad -b = w_0 \le w_1 \le \cdots \le w_r = b
$$

so that $\max_h(x_h - x_{h-1}) \to 0$ and $\max_l(w_l - w_{l-1}) \to 0$ as $r \to \infty$. Note that $\eta^{ik}_u = u'c_{i-k}$ and rewrite

$$
\begin{aligned}
R21^k_{uv} &= n^{-1/2}\sum_{i=1}^n u'c_{i-k}\Big\{ I\big(\xi_i \le y + n^{-1/2}v'Z_i\big) - I\big(-\xi_i < y - n^{-1/2}v'Z_i\big)\Big\}\\
&\quad -n^{-1/2}\sum_{i=1}^n u'c_{i-k}\,\boldsymbol{q}(y,v;Z_i)\\
&= T1_{uv} - T2_{uv}, \quad say.
\end{aligned}
$$

Define $\eta^{ik+}_u := \eta^{ik}_u I(\eta^{ik}_u Z_i'v > 0) = u'c_{i-k}I(u'c_{i-k}Z_i'v > 0)$, $\eta^{ik-}_u = \eta^{ik}_u - \eta^{ik+}_u$. Let $R21^{\pm}_{h,l}$, $T1^{\pm}_{h,l}$, $T2^{\pm}_{h,l}$ stand for the $R21^k_{u_hv_l}$, $T1_{u_hv_l}$, $T2_{u_hv_l}$ with η^{ik}_u being replaced by $\eta^{ik\pm}_u$. Observe that

$$
|R21^k_{uv}|^2_H \le 2\Big\{|T1^+_{uv} - T2^+_{uv}|^2_H + |T1^-_{uv} - T2^-_{uv}|^2_H\Big\}.
$$

Note that $R21^+_{uv}\,(R21^-_{uv})$ is a difference of two nondecreasing (nonincreasing) function of

x, w, and hence, for all $u = x\,e_u \in \mathbb{R}^{p+1}$ and $v = w\,e_v \in \mathbb{R}^q$ where $x_{h-1} \le x \le x_h$ and $w_{l-1} \le w \le w_l$,

$$R21^+_{h-1,l-1} - (T2^+_{h,l} - T2^+_{h-1,l-1}) \le R21^+_{uv} \le R21^+_{h,l} + (T2^+_{h,l} - T2^+_{h-1,l-1}).$$

Therefore, using the fact that $a_1 \le a_2 \le a_3$ implies $|a_2| \le |a_1| + |a_3|$ and that $(a \pm b)^2 \le 2(a^2 + b^2)$ repeatedly, we obtain

$$\begin{aligned} \sup |R21^+_{uv}|^2_H \;\; &\le \;\; 8\Big[\max_{0\le h,l\le r} |R21^+_{h,l}|^2_H + 2\max_{0\le h,l\le r} |T2^+_{h,l} - T2^+_{h-1,l}|^2_H \\ &\qquad + 2\max_{0\le h,l\le r} |T2^+_{h-1,l} - T2^+_{h-1,l-1}|^2_H\Big]. \end{aligned} \tag{2.55}$$

Using $\|\eta^{ik+}_u\|^2 \le \|\eta^{ik}_u\|^2$ and ergodicity of ε,

$$\begin{aligned} &E|T2^+_{h,l-1} - T2^+_{h-1,l-1}|^2_H \\ \le \;\; &16 b^4\, \vartheta_n^2\, \big(2b/\sqrt{n}\big)^{-1} \int_{-b/\sqrt{n}}^{b/\sqrt{n}} \int E\big[\|Z_1\| f_\xi(y + s\|Z_1\|)\big]^2\, dH(y)\, ds \to 0, \end{aligned} \tag{2.56}$$

where the inequality follows from (2.50), and the convergence to zero follows from assumption (B.5) and $\vartheta_n = o(1)$. Similarly, using (2.52), we have

$$\begin{aligned} &E|T2^+_{h-1,l} - T2^+_{h-1,l-1}|^2_H \\ \le \;\; &b^2\vartheta_n^2(|w_l| + |w_{l-1}|)^2 \cdot \big\{(|w_l| + |w_{l-1}|)/\sqrt{n}\big\}^{-1} \\ &\times\Bigg\{ \int_{-|w_{l-1}|/\sqrt{n}}^{|w_l|/\sqrt{n}} \int E\big[\|Z_1\| f_\xi(y + s\|Z_1\|)\big]^2\, dH(y)\, ds \\ &\qquad + \int_{-|w_{l-1}|/\sqrt{n}}^{|w_l|/\sqrt{n}} \int E\big[\|Z_1\| f_\xi(y - s\|Z_1\|)\big]^2\, dH(y)\, ds \Bigg\} \\ \to \;\; &0. \end{aligned} \tag{2.57}$$

Consequently, (2.54), (2.55), (2.56), and (2.57) imply

$$E \sup_{u,v} |R21^{+}_{uv}|^{2}_{H} \longrightarrow 0.$$

Similar facts hold for $|R21^{-}_{uv}|^{2}_{H}$, and this completes the proof of $\sup R2^{k}_{uv} = o_p(1)$.

To show $\sup_{\|u\| \leq b, \|v\| \leq b} \left|R3^{k}_{uv}\right|^{2}_{H}$ is $o_p(1)$, note that $R3^{k}_{uv}$ is the same as $R1^{k}_{uv}$ except the fact that it has $o(1)$ as a weight instead of random error. So, the proof is similar, but much simpler. Hence we do not present here.

Chapter 3

Goodness-of-fit Testing: Khmaladze Transformation vs Empirical Likelihood

3.1 KMT & EL methods

KMT method has not gained much attention despite asymptotic distribution free (ADF) property. Koul provides a review of KMT method in the chapter 9 of Fan and Koul (2006). Let $X_1, ..., X_n$ be i.i.d. random sample from a location-scale family where F is the error d.f., having an absolutely continuous density f. Assume that $\dot{f}$ exists almost everywhere such that $0 < \int(\dot{f}/f)^2 dF < \infty$. Let μ and σ denote unknown location and scale parameters, respectively. Consider the problem to test for all $x \in \mathbb{R}$, and some $\mu \in \mathbb{R}$, $\sigma > 0$

$$H_0 : F(x) = F_0\big((x-\mu)/\sigma\big), \quad \text{vs.} \quad H_a : \ H_0 \text{ is not true.} \tag{3.1}$$

where F_0 is a known d.f. Define

$$\begin{aligned} Z_i &:= (X_i - \mu)/\sigma, \qquad \widehat{Z}_i := (X_i - \widehat{\mu}_n)/\widehat{\sigma}_n, \\ F_n(x) &:= n^{-1}\sum_{i=1}^{n} I(Z_i \le x), \quad \widehat{F}_n(x) := n^{-1}\sum_{i=1}^{n} I(\widehat{Z}_i \le x). \end{aligned} \tag{3.2}$$

where $\widehat{\mu}_n$ and $\widehat{\sigma}_n$ are consistent estimators of μ and σ under the null hypothesis.

As mentioned in the introduction, it is well known that the null distribution of classical K-S test based on F_n does not depend on F_0. However, this fact does not hold any more when the necessity of estimating μ and σ arises, i.e., a test based on $\widehat{F}_n$ is not distribution free. Durbin (1973) showed that null distribution of the test based on $\widehat{F}_n$ depends on the estimators $(\widehat{\mu}_n, \widehat{\sigma})$ as well as F_0. To obtain ADF test, we pay attention to a transformation based on $\widehat{F}_n$ which was proposed by Khmaladze (1979, 1980). Let

$$\begin{aligned} \phi_0(x) &:= -\dot{f}_0(x)/f_0(x), \quad l(x) := (1, \phi_0(x), x\phi_0(x) - 1)', \\ p_0(t) &:= f_0(F_0^{-1}(t)), \quad q_0(t) := F_0^{-1}(t) f_0(F_0^{-1}(t)), \\ \Gamma_t &= \begin{pmatrix} 1-t & p_0(t) & q_0(t) \\ p_0(t) & \int_t^1 \dot{p}_0^2(u)du & \int_t^1 \dot{p}_0(u)\dot{q}_0(u)du \\ q_0(t) & \int_t^1 \dot{p}_0(u)\dot{q}_0(u)du & \int_t^1 \dot{q}_0^2(u)du \end{pmatrix}. \end{aligned} \tag{3.3}$$

Define martingale transformed process

$$\widehat{U}_n(t) := n^{-1/2} \sum_{i=1}^n \left\{ I(\widehat{Z}_i \le z) - l(\widehat{Z}_i)' \int_{-\infty}^{z \wedge \widehat{Z}_i} \Gamma_{F_0(x)}^{-1} l(x) dF_0(x) \right\}, \ t = F_0(z), \ z \in \mathbb{R}. \tag{3.4}$$

Then, weak convergence of $\widehat{U}_n$ to Brownian motion in uniform metric follows from Khmaladze (1981): see the section 4 for the details. Hence, any test based on $T := \sup_{0 \le t \le 1} |\widehat{U}_n(t)|$ is ADF. When we test H_0 in (3.1) via KMT method, we shall use T for the test statistic.

When $Y_1, Y_2, ..., Y_n$ are i.i.d. observations from distribution K, EL is defined as

$$L(K) = \prod_{i=1}^n K(\{Y_i\}) = \prod_{i=1}^n p_i \tag{3.5}$$

where $p_i = K(\{Y_i\}) = K(Y_i) - K(Y_i-)$. It is well-known that the empirical d.f. $K_n(x) := n^{-1}\sum_{i=1}^n I(Y_i \le x)$ maximizes (3.5). Let $R(K) := L(K)/L(K_n)$ denote the EL ratio. Let K_0 be a d.f. with mean μ_0. Owen (1988, 1990) used $R(K)$ and proposed

$$\begin{aligned}\mathbb{R}(\mu_0) &= \sup\left\{R(K)\,|\int Y dK = \mu_0\right\}\\ &= \sup\left\{\prod_{i=1}^{n} np_i\,|\,p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i Y_i = \mu_0\right\},\end{aligned} \tag{3.6}$$

to construct a nonparametric confidence region and test for the mean of Y. Owen (1990) showed that for $Y \sim K_0$

$$-2\log\mathbb{R}(\mu_0) \rightarrow_D \chi^2_{(1)}, \tag{3.7}$$

which is an analog of Wilks's (1938) theorem for nonparametric likelihood. Owen (1990) referred to $\mathbb{R}(\mu_0)$ as "profile" empirical likelihood ratio (PELR) since nuisance parameters are "profiled out."

The PELR has been of interest to statisticians and extended to various settings. Consider the case where unknown K depends on d-dimensional parameter θ. Qin and Lawless (1994) assumed that there exists information about K and θ: there are $h \geq d$ "unbiased estimating functions," that is, $g_1(Y,\theta), ..., g_h(Y,\theta)$ where $E_K[g_j(Y,\theta)] = 0$ for $j = 1, 2, ..., h$. For example, let $E(Y) = \theta$ and $E(Y^2) = \kappa(\theta)$ where $\kappa(\cdot)$ is a known function. Then, g_1 and g_2 can be written as $g_1(Y,\theta) = Y - \theta$ and $g_2(Y,\theta) = Y^2 - \kappa(\theta)$, respectively. With the unbiased estimating functions, they considered an analog of (3.6), i.e., a maximization problem

$$\max\prod_{i=1}^{n} np_i \quad \text{subject to } \ p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i g_j(Y_i,\theta) = 0,\ \ j = 1, 2, ..., h.$$

Hence, they linked PELR with finitely many constraints. By the method of Lagrange multipliers, they obtained the maximum and defined profile empirical log-likelihood ratio (PELLR)

$$l_E(\theta) = \sum_{i=1}^{n}\log\left[1 + \sum_{j=1}^{h}\lambda_j g_j(Y_i,\theta)\right], \tag{3.8}$$

where λ_j is the Lagrange multiplier corresponding to the constraint $\sum_{i=1}^{n} p_i g_j(Y_i,\theta) = 0$. They showed λ_j's are determined in terms of θ and under $H_0 : \theta = \theta_0$,

$$2l_E(\theta_0) - 2l_E(\widehat{\theta}) \to_D \chi^2_{(d)}, \tag{3.9}$$

where $\widehat{\theta}$ minimizes $l_E(\theta)$.

In contrast to Qin and Lawless (1994), Peng and Schick (2013) considered PELR approach combined with infinitely many constraints (or unbiased estimating functions), i.e., the number of constraints increases as the sample size increases. Again, let Y be r.v. which comes from unknown distribution K. Consider testing $H_0 : K = K_0$ where K_0 is a known fixed d.f. Define $\varphi_h(x) := \sqrt{2}\cos(h\pi x)$, for $h = 1, 2, \ldots$. Consequently, we have for all h

$$\int_{[0,1]} \varphi_h(x)dx = 0, \quad \int_{[0,1]} \varphi_h^2(x)dx = 1.$$

Note that when $H_0 : K = K_0$ is true, $K_0(Y)$ will be a uniform r.v., and hence, $E[\varphi_h(K_0(Y))] = 0$ and $E[\varphi_h(K_0(Y))]^2 = 1$. With infinitely many unbiased estimating functions, $\varphi_h \circ K_0$, $h = 1, 2, ...,$ Peng and Schick (2013) proposed an analog of (3.6):

$$\mathbb{R}_n(K_0) = \sup\left\{\prod_{i=1}^{n} np_i : p_i \geq 0, \forall i, \sum_{i=1}^{n} p_i = 1, \sum_{j=1}^{n} p_j\varphi_h(K_0(Y_j)) = 0,\ h = 1, ..., m_n\right\}. \tag{3.10}$$

Peng and Schick (2013) showed that under H_0

$$P(-2\log\mathbb{R}_n(K_0) > \chi^2_{1-\alpha}(m_n)) \to \alpha$$

where m_n and n tend to infinity and m_n^3/n tends to 0, and $\chi^2_{1-\alpha}(m_n)$ denotes the $(1-\alpha)$-quantile of the chi-square distribution with m_n degrees of freedom. They extended the result to testing H_0 where underlying distribution K depends on an unknown $d-$dimensional parameter θ. With efficient estimator $\widehat{\theta}$ - e.g., maximum likelihood estimator (MLE) - they derived $\mathbb{R}_n(K_{\widehat{\theta}})$ - the test statistic in (3.10) with K_0 replaced by $K_{\hat{\theta}}$ - and showed that under the null hypothesis

$$P(-2\log\mathbb{R}_n(K_{\hat{\theta}}) > \chi^2_{1-\alpha}(m_n - d)) \to \alpha$$

where m_n and n tend to infinity and m_n^3/n tends to 0. Since their approach is free from the question of how many constraints should be used, we use $-2\log R_n(K_{\hat{\theta}})$ for the test statistic when we implement EL method to test H_0 in (3.1).

At this time, it should be mentioned that EL approach has one critical drawback when it is employed for hypothesis testing. Note that all the EL methods introduced in this section solve the maximization problem subject to finitely or infinitely many constraints (or unbiased estimating functions). Assume that r.v. Y comes from an unknown distribution K. Consider testing $H_0 : K = K_0$ vs $H_a : K = K_a$ via the EL method. Let θ_0 and θ_a denote parameters associated with K_0 and K_a, respectively. When we use the test statistic in (3.8) ((3.10)), constraints corresponding to H_0 and H_a are $E_{K_0}[g(Y, \theta_0)] = 0$ and $E_{K_a}[g(Y, \theta_a)] = 0$. When K_0 are similar to K_a, we will therefore obtain the similar test statistics $l_E(\theta_0)$ and $l_E(\theta_a)$ (or $\mathbb{R}_n(K_0)$ and $\mathbb{R}_n(K_a)$). As a result, it is very likely to make a type II error when true d.f. of Y is K_a, i.e., EL method will have very poor power. Examples of such K_0 and K_a are; logistic and normal; and logistic and Student's t (STT) where degrees of freedom (df) is greater than or equal to 5. The poor power of EL method for these two cases are illustrated in the next section. See, e.g., Table 3.2 and 3.3.

3.2 Simulation study

Let F_i for $i = N, L$ denote d.f. of standard normal and logistic r.v., respectively. Note that

$$F_N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-y^2/2)dy, \qquad F_L(x) = \frac{1}{1+e^{-x}}, \qquad x \in \mathbb{R}.$$

In this simulation study, we report the findings obtained from goodness-of-fit test for two location-scale distributions: $H_0 : F(x) = F_i((x - \mu)/\sigma)$, $i = N, L$. Table 3.1 reports asymptotic critical values for KMT and EL methods. The critical values of KMT test are

α	KMT	EL1				EL2			
		n=50	100	200	500	n=50	100	200	500
0.05	2.24	5.99	7.81	9.49	12.59	7.81	9.49	11.07	14.07
0.01	2.81	9.21	11.34	13.28	16.81	11.34	13.28	15.09	18.48

Table 3.1: Critical value for KMT and EL

available at homepages.ecs.vuw.ac.nz/ ray/Brownian which is made by Dr. R. Brownrigg. For those of EL method, we consider two different df's: $m_{n1} = \lfloor n^{1/3} \rfloor - 2$ and $m_{n2} = \lfloor n^{1/3} \rfloor - 1$ where $\lfloor x \rfloor$ is the largest integer not greater than x. Let EL1 and EL2 denote EL methods with df of m_{n1} and one of m_{n2}, respectively. When we generate $n = 50, 100, 200$, and 500 samples from each null distribution, we use 2 and 5 for location and scale parameters. We repeat random sample generation 1,000 times and obtain KMT and EL test statistics. In order to obtain two different EL test statistics, we use $\lfloor n^{1/3} \rfloor$ and $\lfloor n^{1/3} \rfloor + 1$ unbiased estimating functions for the constraints in (3.10). Empirical levels and powers are then calculated from dividing the number of rejection of null hypothesis by 1,000. We use MLE for $\widehat{\mu}$ and $\widehat{\sigma}$ in the subsequent sections.

3.2.1 Testing for normal distribution

In this section, we compare KMT and EL methods to test for normality. Let F denote standard normal d.f. and f be its density. $l(x)$ and $\Gamma_{F(x)}$ in (3.3) turn out

$$
\begin{aligned}
&l(x) = (1, x, x^2 - 1)', \\
&\Gamma_{F(x)} = \begin{pmatrix} 1 - F(x) & f(x) & xf(x) \\ f(x) & xf(x) + (1 - F(x)) & (1 + x^2)f(x) \\ xf(x) & (1 + x^2)f(x) & (x^3 + x)f(x) + 2(1 - F(x)) \end{pmatrix}.
\end{aligned}
$$

Let

$$
\begin{aligned}
A_{11}(x) &:= 1 - F(x),\ A_{12}(x) := (f(x), xf(x)),\ A_{21}(x) := (f(x), xf(x))', \\
A_{22}(x) &:= \begin{pmatrix} xf(x) + (1 - F(x)) & (1 + x^2)f(x) \\ (1 + x^2)f(x) & (x + x^3)f(x) + 2(1 - F(x)) \end{pmatrix}.
\end{aligned}
$$

Then the inverse of $\Gamma_{F(x)}$ also can be expressed in partitioned form, i.e.,

$$
\Gamma^{-1}_{F(x)} = \begin{pmatrix} B_{11}(x) & B_{12}(x) \\ B_{21}(x) & B_{22}(x) \end{pmatrix}
$$

where

$$
\begin{aligned}
B_{11}(x) &= (A_{11}(x) - A_{12}(x)A_{22}(x)^{-1}A_{21}(x))^{-1}, \qquad (3.11) \\
B_{12}(x) &= -B_{11}(x)A_{12}(x)A_{22}(x)^{-1}, \quad B_{21}(x) = B_{12}(x)' \\
B_{22}(x) &= A_{22}(x)^{-1} + A_{22}(x)^{-1}A_{21}(x)B_{11}(x)A_{12}(x)A_{22}(x)^{-1}.
\end{aligned}
$$

Let $\tilde{F}(x) := 1 - F(x)$, $c_1(x) := \{-(x^2+1)f^2(x) + (x^3+3x)f(x)\tilde{F}(x) + 2\tilde{F}^2(x)\}^{-1}$ and $c_2(x) := 2\tilde{F}^3(x) + (x^3+3x)f(x)\tilde{F}^2(x) - (2x^2+3)f^2(x)\tilde{F}(x) + xf^3(x)$. Note that with $r(x) := (x, (x^2-1))'$, we have

$$\Gamma^{-1}_{F(x)} l(x) f(x) = \begin{bmatrix} f(x)B_{11}(x) + f(x)B_{12}(x)\, r(x) \\ f(x)B_{21}(x) + f(x)B_{22}(x)\, r(x) \end{bmatrix}.$$

Finally, with $\widehat{Z}_i$ in (3.2), we have

$$\begin{aligned}
& l(\widehat{Z}_i)'\Gamma^{-1}_{F(x)} l(x) f(x) \qquad (3.12)\\
= \ & 2f(x)c_2(x)\Big[\tilde{F}^2(x) + xf(x)\tilde{F}(x) - f^2(x)\}\Big] \\
& +\widehat{Z}_i f(x)c_1(x)c_2(x)\Big[4x\tilde{F}^4(x) + (2x^4+8x^2-2)f(x)\tilde{F}^3(x) + (x^5-7x)f^2(x)\tilde{F}^2(x) \\
& \qquad -(2x^4+3x^2-1)f^3(x)\tilde{F}(x) + (x^3+x)f^4(x)\Big] \\
& +(\widehat{Z}_i^2-1)f(x)c_1(x)c_2(x)\Big[2(x^2-1)\tilde{F}^4(x) + (x^5+2x^3-9x)f(x)\tilde{F}^3(x) \\
& -(4x^4+9x^2-5)f^2(x)\tilde{F}^2(x) + (5x^3+9x)f^3(x)\tilde{F}(x) - 2(x^2+1)f^4(x)\Big].
\end{aligned}$$

We then replace $l(\widehat{Z}_i)'\Gamma^{-1}_{F(x)} l(x) f(x)$ in (3.4) by (3.12).

Table 3.2 reports empirical levels and powers of KMT and EL methods. The first (second) columns of KMT and EL's represent those when α is 0.05 (0.01). The value corresponding to normal F represents the empirical level while those corresponding to others-logistic, STT with df of 5, mixture of two normal distributions (MTN), Cauchy, and Laplace-represent the powers. For the MTN, we use $0.9N(2,5^2) + 0.1N(2,15^2)$; for logistic, Cauchy, and Laplace distributions, we use 2 and 5 for location and scale parameters. $(*)$ implies the corresponding method shows the closest empirical level to the α or highest power; e.g., KMT has the closest empirical level (0.01) and the highest power (0.152) when $n = 500$ with $\alpha = 0.01$ and $n = 50$ with $\alpha = 0.05$, respectively. It is hard to judge the superiority of two methods by the empirical level since neither KMT nor EL demonstrates better performance consistently.

F	n	KMT	EL1	EL2	KMT	EL1	EL2
normal	50	0.026	0.088	0.058 (∗)	0.012 (∗)	0.020	0.022
	100	0.032	0.060 (∗)	0.082	0.022	0.016 (∗)	0.018
	200	0.032	0.044	0.050 (∗)	0.016	0.006	0.008 (∗)
	500	0.024	0.062	0.060 (∗)	0.010 (∗)	0.018	0.016
logistic	50	0.152 (∗)	0.148	0.100	0.092 (∗)	0.036	0.032
	100	0.200 (∗)	0.148	0.168	0.120 (∗)	0.056	0.032
	200	0.320 (∗)	0.176	0.168	0.244 (∗)	0.052	0.056
	500	0.632 (∗)	0.372	0.384	0.532 (∗)	0.160	0.160
STT	50	0.244	0.276 (∗)	0.224	0.208 (∗)	0.116	0.092
	100	0.400 (∗)	0.304	0.264	0.292 (∗)	0.132	0.136
	200	0.556 (∗)	0.500	0.468	0.480 (∗)	0.252	0.680
	500	0.884 (∗)	0.868	0.848	0.804 (∗)	0.680	0.660
MTN	50	0.382	0.410 (∗)	0.346	0.352 (∗)	0.224	0.164
	100	0.562 (∗)	0.528	0.462	0.516 (∗)	0.296	0.258
	200	0.812 (∗)	0.756	0.706	0.716 (∗)	0.566	0.532
	500	0.982 (∗)	0.966	0.958	0.966 (∗)	0.908	0.886
Cauchy	50	0.832	0.996 (∗)	0.980	0.744	0.984 (∗)	0.968
	100	0.988 (∗)	0.984	0.968	0.948	0.984 (∗)	0.968
	200	1.000 (∗)	0.976	0.968	1.000 (∗)	0.980	0.964
	500	1.000 (∗)	0.990	0.964	1.000 (∗)	0.970	0.960
Laplace	50	0.276	0.468 (∗)	0.424	0.186	0.188 (∗)	0.188 (∗)
	100	0.480	0.788 (∗)	0.676	0.380	0.516 (∗)	0.444
	200	0.768	0.932 (∗)	0.920	0.632	0.796 (∗)	0.756
	500	0.980	1.000 (∗)	1.000 (∗)	0.964	1.000 (∗)	1.000 (∗)

Table 3.2: Empirical level and power obtained from testing $H_0 : F = F_N$

However, it is not hard to tell which of the two methods is better in terms of the power. For all distributions except Laplace, KMT outperforms EL: KMT shows better power than EL in most n's. When F is logistic, EL displays poor power (less than 0.4) as stated in the previous section. The power of KMT is more than or equal to almost twice that of EL when n is 200 or 500. For STT, MTN, and Cauchy, KMT still maintains superiority over EL even though that is weakened. As shown in the table, the performance gap between KMT and

EL is widened when α is 0.01.

In contrast, Laplace F shows the opposite result: EL outperforms KMT for all n's and α's. But the superiority of EL over KMT is not as strong as that of KMT over EL in the logistic case; when n reaches 500, the difference of powers between EL and KMT is less than 0.05 while counterpart of the logistic case is more than 0.2.

3.2.2 Testing for logistic distribution

Consider logistic d.f. $F(x) = 1/(1+e^{-x})$, and its density $f(x) = e^x/(1+e^x)^2$. Note that $\phi(x) = -\dot{f}(x)/f(x) = (e^x - 1)/(e^x + 1)$. With $t = F(x)$, we have

$$\dot{p}(t) \quad = \quad \frac{1-e^x}{1+e^x}, \quad \dot{q}(t) = -1 + \frac{x(-1+e^x)}{1+e^x}.$$

It is easy to see that

$$\begin{aligned}
\int_{F(x)}^{1} \dot{p}(u)^2 du &= \frac{3e^{2x}+1}{3(e^x+1)^3}, \\
\int_{F(x)}^{1} \dot{p}(u)\dot{q}(u) du &= \frac{1}{3}\ln(1+e^x) - \frac{e^x\{x(3+e^2x)+(1+e^x)\}}{3(1+e^x)^3}, \\
\int_{F(x)}^{1} \dot{q}(u)^2 du &= \frac{1}{(1+e^x)} + \left(-\frac{2}{(1+e^x)} - \frac{2xe^x}{(1+e^x)^2}\right) + Re(x),
\end{aligned}$$

where $Re(x) = \int_x^\infty 2s^2 e^s (1-e^s)^2/(1+e^s)^4 ds$. Note that $\int \dot{q}(u)^2 du$ does not have a closed form solution since $Re(x)$ does not have one. However, $Re(x)$ is bounded and decays to 0 fast as x goes to ∞; it converges to the finite value (2.43) as x goes to $-\infty$. Therefore, we get a numerical approximation to $Re(x)$ and use it when we calculate the inverse of Γ_F. Let

$v_1(x) := \int_{F(x)}^1 \dot{p}(u)\dot{q}(u)du$ and $v_2(x) := \int_{F(x)}^1 \dot{q}^2(u)du$. Also define

$$\begin{aligned} d(x) &:= \Big[3(1+3e^{2x})(1+e^x)^3 v_2(x) - 9(1+e^x)^6 v_1^2(x)\Big]^{-1}, \\ k_1(x) &:= 3(1+e^x)^3 v_2(x) - 3x(1+e^x)^3 v_1(x), \\ k_2(x) &:= -3(1+e^x)^3 v_1(x) + x(1+3e^{2x})^3. \end{aligned}$$

Then, using partitioned four 2×2 blocks of $\Gamma_{F(x)}$ as done in the previous section, we finally obtain

$$\begin{aligned} & l(\widehat{Z}_i)'\Gamma_{F(x)}^{-1} l(x) f(x) \\ = \;& 3e^x(1+e^x)B_{11}(x)\Big\{1 - 3d(x)e^x\big\{(1-e^x)k_1(x) + k_2(x)(1+e^x+x(1-e^x))\big\}\Big\} \\ & + \frac{3(1-e^{\widehat{Z}_i})}{(1+e^{\widehat{Z}_i})}\Big\{9e^{2x}(1+e^x)^2 d(x)B_{11}(x)k_1(x) \\ & + 3e^x d(x)\big\{3(1+e^x)^3(1-e^x)v_2(x) - 3(1+e^x)^3 v_1(x)\{(1+e^x)+x(1-e^x)\}\big\} \\ & + 27e^{3x}(1+e^x)^2 B_{11}(x)d^2(x)\big\{(1-e^x)k_1^2(x) + \{(1+e^x)+x(1-e^x)\}k_1(x)k_2(x)\big\}\Big\} \\ & + \left(1 + \frac{\widehat{Z}_i(1-e^{\widehat{Z}_i})}{(1+e^{\widehat{Z}_i})}\right)\Big\{9e^{2x}(1+e^x)^2 d(x)B_{11}(x)k_2(x) \\ & + 3e^x d(x)\big\{-3(1+e^x)^3(1-e^x)v_1(x) + (1+3e^{2x})\{(1+e^x)+x(1-e^x)\}\big\} \\ & + 27e^{3x}(1+e^x)^2 B_{11}(x)d^2(x)\big\{(1-e^x)k_1(x)k_2(x) + \{(1+e^x)+x(1-e^x)\}k_1^2(x)\big\}\Big\}, \end{aligned}$$

and hence, use above quantity for $\widehat{U}_n$ in (3.4).

Table 3.3 reports empirical levels and powers of KMT and EL's for testing $H_0 : F = F_L$; the value corresponding to logistic F stands for the empirical level, and others represent the powers. EL shows better empirical level than KMT for all n's when $\alpha = 0.05$. With $\alpha = 0.01$, KMT, however, has better empirical level than EL except $n = 50$. Therefore,

F	n	KMT	EL1	EL2	KMT	EL1	EL2
logistic	50	0.032	0.076	0.062 (∗)	0.020	0.018 (∗)	0.004
	100	0.030	0.062	0.060 (∗)	0.014 (∗)	0.016	0.018
	200	0.026	0.060	0.052 (∗)	0.010 (∗)	0.020	0.020
	500	0.054	0.052 (∗)	0.052 (∗)	0.011 (∗)	0.006	0.008
normal	50	0.004	0.076	0.080 (∗)	0.000	0.022 (∗)	0.020
	100	0.020	0.080 (∗)	0.066	0.000	0.018	0.020 (∗)
	200	0.136 (∗)	0.098	0.104	0.040 (∗)	0.034	0.028
	500	0.614 (∗)	0.176	0.204	0.342 (∗)	0.066	0.084
STT	50	0.060	0.083 (∗)	0.063	0.040 (∗)	0.025	0.019
	100	0.097 (∗)	0.070	0.072	0.070 (∗)	0.013	0.015
	200	0.132 (∗)	0.065	0.063	0.101 (∗)	0.019	0.017
	500	0.219 (∗)	0.057	0.049	0.173 (∗)	0.013	0.013
MTN	50	0.164 (∗)	0.089	0.101	0.129 (∗)	0.031	0.018
	100	0.185 (∗)	0.085	0.097	0.144 (∗)	0.029	0.011
	200	0.297 (∗)	0.076	0.087	0.234 (∗)	0.027	0.028
	500	0.583 (∗)	0.115	0.081	0.518 (∗)	0.011	0.009
Cauchy	50	0.649	0.965	0.974 (∗)	0.517	0.904	0.930 (∗)
	100	0.855	1.000 (∗)	1.000 (∗)	0.782	1.000 (∗)	1.000 (∗)
	200	1.000	1.000	1.000	1.000	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000
Laplace	50	0.137	0.146	0.208 (∗)	0.075 (∗)	0.025	0.054
	100	0.199	0.350 (∗)	0.317	0.159 (∗)	0.126	0.093
	200	0.332	0.547 (∗)	0.518	0.194	0.320 (∗)	0.296
	500	0.613	0.915 (∗)	0.911	0.399	0.734	0.754 (∗)

Table 3.3: Empirical level and power obtained from testing $H_0 : F = F_L$

KMT and EL tie in terms of the empirical level.

For MTN, normal, and STT, the fact that KMT outperforms EL in terms of the power is evident. In case of MTN, KMT overwhelms EL for all n's and α's. The difference of powers between KMT and EL increases as n increases and reaches more than 0.5 for both α's. In the normal case, the maximum power EL attains is 0.204-when $n = 500$ and $\alpha = 0.05$-while that of KMT is 0.614. When $n = 500$ and $\alpha = 0.01$, both EL's display powers less than even 0.1 while KMT shows 0.342. When F is STT, EL shows the extremely poor power with $\alpha = 0.01$; EL obtains the power of only 0.013 even though n reaches 500. Note that the counterpart of KMT is 0.173.

When F is either Cauchy or Laplace, EL shows slightly or strictly better power than KMT. For the Cauchy, EL attains the power greater than 0.9 for all α's even when $n = 50$ while KMT shows the power less than 0.8. Both KMT and EL, however, attain the powers of 1 when n reaches 200; the difference of performances between KMT and EL disappears as n increases. EL has better power than KMT when F is Laplace, as it does in the previous section. When $n = 500$ and $\alpha = 0.05$, EL attains the power of 0.9; KMT never attains the power greater than 0.8.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Ashenfelter, O. (1978), Estimating the Effect of Training Programs on Earnings with Longitudinal Data, *Review of Economics and Statistics*, **60**, 47-57.

[2] Athreya, K. B. and Pantula, S. G. (1986), Mixing Properties of Harris Chanis and Autoregressive Processes, *J. Appl. Probab.*, **23**, 880-892.

[3] Baltagi, B. D. (2001). *Econometric Analysis of Panel Data*. John Wiley and Sons, Chichester.

[4] Beran, R. J. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.*, **5** 445-463.

[5] Billingsley, P. (1999). *Convergence of Probability Measures.* John Wiley and Sons, New York.

[6] Bloomfield, P. (1992), Trends in Global Temperature, *Climatic Change*, **21**, 1-16.

[7] Deo, C. M. (1973), A Note on Empirical Processes of Strong Mixing Sequences, *Ann. Probab.*, **1**, 870-875.

[8] Donoho, D. L. and Liu, R. C. (1988 a). The "automatic" robustness of minimum distance functionals. *Ann. Statist.*, **16** 552-586.

[9] Donoho, D. L. and Liu, R. C. (1988 b). Pathologies of some minimum distance estimators. *Ann. Statist.*, **16** 587-608.

[10] Dong, L. B. and Giles, D. E. A. (2004). An empirical likelihood ratio test for normality, *Econometrics Working Paper EWP0401.*

[11] Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM, Philadelphia.

[12] Durrett, R. (2005). *Probability: Theory and Examples.*, Thomson, Brooks Cole.

[13] Fan, J. and Koul, H. (2006) *Frontiers in Statistics.* Imperial College Press, London.

[14] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, **69** 383-393.

[15] Khmaladze, E. V. (1979). The use of ω^2-tests for testing parametric hypotheses. *Theory Probab. Appl.*, **24** 283-301.

[16] Khmaladze, E. V. (1981). A martingale approach in the theory of goodness-of-fit tests. *Theory Probab. Appl.*, **26** 240-257.

[17] Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms of goodness-of-fit tests in regression models. *Ann. Statist.*, **32** 995-1034.

[18] Kim, J. (2016), `KoulMde`: An `R` Package for Koul's Minimum Distance Estimation, arXiv:1606.04182.

[19] Koul, H. L. (1977), Behavior of Robust Estimators in the Regression Model with Dependent Errors, *Ann. Statist.*, **5**, 681-699.

[20] Koul, H. L. (1985), Minimum Distance Estimation in Linear Regression with Unknown Error Distributions, *Statist. Probab. Lett.*, 3, 1-8.

[21] Koul, H. L. (1986), Minimum Distance Estimation and Goodness-of-Fit Tests in First-Order Autoregression, *Ann. Statist.*, **14**, 1194-1213.

[22] Koul, H. L. (2002), *Weighted Empirical Process in Nonlinear Dynamic Models.* Springer, Berlin, Vol. 166.

[23] Koul, H. L. and de Wet, T. (1983), Minimum Distance Estimation in a Linear Regression Model, *Ann. Statist.*, **11**, 921-932.

[24] Koul, H. and Zhu, X. (2015). Gooness-of-fit testing of error distribution in nonparametric ARCH(1) models. *J. Multivariate Anal*, **137** 141-160

[25] Mehra, K. L. and Rao, M. S. (1975), Weak Convergence of Generalized Empirical Processes Relative to d_q under Strong Mixing, *Ann. Probab.*, **3**, 979-991.

[26] Millar, P. W. (1981). Robust estimation via minimum distance methods. *Zeit fur Wahrscheinlichkeitstheorie.*, **55** 73-89.

[27] Moulton, B. R. (1986). Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics.*, **32** 385-397.

[28] Noether, G. E. (1949), On a Theorem by Wald and Wolfowitz, *Ann. Math Statist.*, **20**, 445-458.

[29] Parr, W. C. and Schucany, W. R. (1979). Minimum distance and robust estimation. *J. Amer. Statist. Assoc.*, **75** 616-624.

[30] Peng, H. and Schick, A. (2013). Empirical likelihood approach to goodness of fit testing. *Bernoulli*, **19** 954-981.

[31] Prescott, E. C. (1987), Theory ahead of Business Cycle Measurement, Carnegie-Rochester Conference on Public Policy.

[32] Qin, J. and Lawless, J. (1994). Empirical lieklihood and general estimating equations. *Ann. Statist.*, **22** 300-325.

[33] Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *J. Amer. Statist. Assoc.*, **69** 730-737.

[34] Wilks, S. S. (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. statist.*, **9** 60-2.